SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(103rd Edition)

JULY 2021

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

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

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

This issue contains the information published during **July, 2021**. 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 Bi-GRU with attention and CapsNet hybrid model for cyberbullying detection on social media, ***3.Akshi Kumar*** and ***6.Nitin Sachdeva***, CSE, DTU

2. A COMPARATIVE STUDY OF VARIOUS DEEP LEARNING TECHNIQUES FOR SPATIO-TEMPORAL SUPER-RESOLUTION RECONSTRUCTION OF FORCED ISOTROPIC TURBULENT FLOWS, ***8.T.S. Sachin Venkatesh***, ***8.Rajat Srivastava***, ***8.Pratyush Bhatt***, ***8.Prince Tyagi*** and ***3.Dr. Raj Kumar Singh***, Applied Physics and Mechanical, DTU

3. A High Gain, High-Efficiency Dual-Band Cascode Class-F PA for Ka-Band Application, ***6.Priya Puja***, ***3.Kaur Gurjit*** and ***6.Kumar Rajesh***, Electronics, DTU

4. A Novel 8.5 GHz Parallel Coupled, Half Wavelength Bandpass Filter with Dumbbell-Shaped Defected Ground Structure, ***8.Aditya Agrawal***, ***8.Adya Agarwal***, ***8.Ojas Aryam*** and ***3.Asok De***, Electronics, DTU

5. Aerodynamic mitigation by corner modification on square model under wind loads employing CFD and wind tunnel, ***6.Nikhil Gaur*** and ***3.Ritu Raj***, Civil, DTU

6. An efficient Jaya algorithm with Powell's Pattern Search for optimal power flow incorporating distributed generation, ***6.Saket Gupta***, ***3.Narendra Kumar*** and Laxmi Srivastava, Electrical, DTU

7. AN INTEGRATIVE APPROACH FOR WELLNESS: AN ASSESSMENT OF POTENTIAL OF AYURVEDA DURING COVID-19 PANDEMIC, Richa Chauhan and ***3.Nidhi Maheshwari***, DSM, DTU

8. Anomaly detection framework to prevent DDoS attack in fog empowered IoT networks, Deepak Kumar Sharma, Tarun Dhankhar, Gaurav Agrawal, Satish Kumar Singh, ***7.1.Deepak Gupta***, Jamel Nebhen and Imran Razzak, DTU

9. Application of Swarm-Based Feature Selection and Extreme Learning Machines in Lung Cancer Risk Prediction, ***8.Priya Garg*** and ***3.Deepti Aggarwal***, CSE, DTU

10. ArtiMarker: A Novel Artificially Inflated Video Marking And Characterization Method on YouTube, ***6.Deepika Varshney*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

11. CFD Modelling and Simulation of an Indirect Forced Convection Solar Dryer, ***8.Rupanshu Singh***, ***8.Pranav Salhan*** and ***3.Anil Kumar***, Mechanical & Envt, DTU

12. CFOA Based Negative Floating Fractional Order Capacitance Multiplier, ***8.Rahul Gupta***, ***8.Charvi Agrawal***, ***8.Ankit Gupta*** and ***3.Rajeshwari Pandey***, Electronics, DTU

13. CNN-based Person Recognition System for Masked Faces in a post-pandemic world, ***8.Parangat Mittal***, ***8.Kartikeya Pandey***, ***8.Peeyush Tawani*** and ***3.Prof. Rajesh Rohilla***, Electronics, DTU

14. Computational Framework of Various Semi-Active Control Strategies for Road Vehicles Thorough Bondgraphs, Ashish Gupta, Nilanjan Bharadwaj and ***3.Vikas Rastogi***, Mechanical, DTU

15. Content Based Bot Detection using Bot Language Model and BERT Embeddings, ***8.Shubham Kumar***, ***8.Shivang Garg***, ***8.Yatharth Vats*** & ***3.Anil Singh Parihar***, CSE, DTU

16. COVID 19, Pneumonia and Other Disease Classification using Chest X-Ray images, ***8.Abhishek Panwar***, ***8.Akshay Dagar***, ***8.Vishal Pal*** and ***3.Dr. Vinod Kumar***, CSE, DTU

17. COVID-19 Severity Assessment from Chest X-rays using Attention-based Weakly-Supervised Learning, ***8.Kunal Chaturvedi***, ***8.Tanmay Kansal***, ***8.Saksham Gupta***, ***3.Dinesh Kumar Vishwakarma*** and ***3.Naokant Deo***, Applied Mathematics & IT, DTU

18. Design & Analysis of Nonlinear Controller for a Standalone Photovoltaic System Using Lyapunov Stability Theory, ***3.Narendra Kumar*** and ***6.Aman Sharma***, Electrical, DTU

19. Design and Simulation of Connecting Rod using Finite Element Analysis, ***8.Suraj Kumar***, Mechanical, DTU

20. Design and Study of Dadda Multiplier by using 4:2 Compressors and Parallel Prefix Adders for VLSI Circuit Designs, ***8.Ayush Kumar***, ***8.Ankit Kumar*** and ***3.Dr. Deva Nand***, ECE, DTU

21. Design Of High Bandwidth Circularly Polarised Antipodal Vivaldi Array for 5G Applications, ***8.Anirudh Nakra***, ***8.Abhijeet Vats*** and ***3.Asok De***, Electronics, DTU

22. Detection and Classification of Power Quality Disturbances Using Discrete Wavelet Transform and Neural Network Technique, ***3.N.K. Bhagat***, ***8.Shubhechchhu Singh***, ***8.Akash Gupta*** and ***8.Ankush Tanwar***, Electrical, DTU

23. Detection of engineered surfaces using deep learning approach in AVIRIS-NG hyperspectral data, ***6.Shalini Gakhara*** and ***3.Kailash Chandra Tiwari***, IT and Geoinformatics, DTU

24. Detection of Pulsars by Classical Machine Learning Algorithms, ***8.Deepanshu Beniwal***, ***8.Abhishek Roy***, ***8.Himanshu Yadav*** and ***3.Anamika Chauhan***, IT, DTU

25. Dynamic Analysis of Steel Silo, ***6.Uzma Bashir Wani*** and ***3.Nirendra Dev***, Civil, DTU

26. Dynamic Power Consumption and Delay Analysis for Ultra-Low Power 2 to 1 Multiplexer Designs, ***8.Nishant KUMAR***, ***3.Poornima MITTAL***, ***6.Bhawna RAWAT*** and ***8.Mudit MITTAL***, Electronics, DTU

27. Effect of Rice Husk Ash, silica fume & GGBFS on compressive strength of performance based concrete, ***8.Mohit Gupta***, ***3.Dr. Ritu Raj*** and ***3.Dr. Anil Kumar Sahu***, Civil, DTU

28. Effect of Welding Parameters on Mechanical Properties of Cold Metal Transfer Welded Thin AISI 304 Stainless-Steel Sheets, ***6.Jayanta Ghosh Roy***, ***3.N. Yuvaraj*** and ***3.Vipin***, Mechanical, DTU

29. Effective Capacity Analysis Over Fisher-Snedecor F Fading Channels with MRC Reception, ***6.Manpreet Kaur*** and ***3.Rajesh Kumar Yadav***, CSE, DTU

30. Evaluation of Wind Induced Interference Effects on Shape Remodeled Tall Buildings, ***6.1.Supriya Pal*** and ***3.Ritu Raj***, Civil, DTU

31. Explainable Artificial Intelligence for Sarcasm Detection in Dialogues, ***3.Akshi Kumar***, Shubham Dikshit and Victor Hugo C. Albuquerque, CSE, DTU

32. Fabrication of activated carbon electrodes derived from peanut shell for high-performance supercapacitors, Lokesh Pandey, Subhajit Sarkar, Anil Arya, A. L. Sharma, ***3.Amrish Panwar***, R. K. Kotnala and Anurag Gaur, A Physics, DTU

33. Fishing out the Phishing Websites, ***8.Peya Mowar*** and ***3.Mini Jain***, IT, DTU

34. Forest Fires in India: A Review, ***8.Reuben Sewak***, ***8.Mehak Vashisth*** and ***3.Lovleen Gupta***, Environmental, DTU

35. FPGA Implementation of a pipelined and pseudo-randomized TDES algorithm, ***8.Arnab Roy***, ***8.Anubhav Gadhok*** and ***3.Deva Nand***, Electronics, DTU

36. FPGA Implementation of CNN based Fingertip Digit Recognition, ***8.1.MD Shahbaz Khan***, ***8.1.Niharika***, ***8.1.Priya Yadav, 8.1.Rishabh Verma*** and ***3.Indu Sreedevi***, Electronics, DTU

37. From Classical to Quantum: A Review of Recent Progress in Reinforcement Learning, ***8.Kunal Kashyap***, ***8.Lalit***, ***8.Daksh Shah*** and ***3.Lokesh Gautam***, Electronics, DTU

38. GENERATION AND PARAMETERIZATION OF FORCED ISOTROPIC TURBULENT FLOW USING AUTOENCODERS AND GENERATIVE ADVERSARIAL NETWORKS, ***8.Kanishk***, ***8.Tanishk Nandal***, ***8.Prince Tyagi*** and ***3.Raj Kumar Singh***, DTU

39. GREY WOLF OPTIMIZATION TUNED TID AND I-TD CONTROLLERS FOR TRAJECTORY TRACKING OF DYNAMICAL AERIAL SYSTEM, ***8.Karan Sehgal*** and ***8.Manik Verma***, DTU

40. Heat and water flux modeling in an earth dam, ***6.Sushant Kumar***, ***3.A. K. Sahu*** and ***3.Munendra Kumar***, Civil, DTU

41. HTGAN : An architecture for Hindi Text based Image Synthesis, ***3.Dr. Anil Singh Parihar***, ***8.Aditya Kaushik***, ***8.Aditya Vikram Choudhary*** and ***8.Amit Kumar Singh***, CSE, DTU

42. Implementation of a Shadow Filter using CFCC, ***8.Prakhar Prakash***, ***8.Yashodhan Srivastava***, ***8.Shubham Kumar*** and ***3.Alok Kumar Singh***, Electronics, DTU

43. Improved Representation Learning for Session-based Recommendation, Sai Mitheran, ***8.Abhinav Java***, Surya Kant Sahu and Arshad Shaikh, DTU

44. Improving the Productivity of the Machining Process of a Manufacturing Company: A Six Sigma Case Study, ***8.Vikram Narula***, ***3.Girish Kumar*** and MK Loganathan, Mechanical, DTU

45. IoMT and DNN-enabled Drone-Assisted Covid-19 Screening and Detection Framework for Rural Areas, Naren, Vinay Chamola, Sainath Baitragunta, Ananthakrishna Chintanpalli, ***7.1.Puneet Mishra***, Sujan Yenuganti and Mohsen Guizani, Electronics, DTU

46. KCNet: Kernel-based Canonicalization Network for entities in Recruitment Domain, Nidhi Goyal, Niharika Sachdeva, Anmol Goel, ***8.Jushaan Singh Kalra***, and Ponnurangam Kumaraguru, DTU

47. Learning image by-parts using early and late fusion of auto-encoder features, ***3.Seba Susan*** and ***7.Jatin Malhotra***, IT, DTU

48. Lower Fin Modulation Analysis for a Novel 5nm Top Bottom Gated Junctionless FinFET for improved performance, ***8.Lakshya Gangwani***, ***8.Ruchika Chakravarti*** and ***3.Rishu Chaujar***, Applied Physics, DTU

49. Machine Learning and IoT based solutions for detection of arrhythmia using ECG signals, ***3.Vinod Kumar***, ***8.Awishkar Ghimire*** and ***8.Hitesh Kumar Hoon***, CSE, DTU

50. Medical Image Fusion using Lifting based Transform Techniques, ***8.Anisha Gupta***, ***8.Priyesh Chaurasia*** and ***3.Sudipta Majumdar***, Electronics, DTU

51. Modified Sliding Singular Spectrum Analysis-Based Noncontact-Type Detection of Separated Instrument in Root Canal Therapy Using Low-Frequency Ultrasonic Pattern Transceiver Design, ***3.Rajiv Kapoor***, ***6.Deepak Sharma*** and Aarchishya Kapoor, Electronics, DTU

52. Movie genres and beyond: An analytical survey of classification techniques, ***8.Avleen Singh Mehal***, ***8.Kalpana Meena***, ***8.Rijul Bir Singh*** and ***3.Prashant Giridhar Shambharkar***, Software and Computer, DTU

53. Multi Label Toxic Comment Classification u sing Machine Learning Algorithms, ***8.Abhishek Aggarwal*** and ***8.Atul Tiwari***, Electrical, DTU

54. Multi-objective optimization of magnetic abrasive finishing by desirability function with genetic algorithm, Rajneesh Kumar Singh, Swati Gangwar, D.K. Singh and ***6.Shadab Ahmad***, Mechanical, DTU

55. Multi-objective Optimization of Thermodynamic and Economic Performances of Natural Refrigerants for Cascade Refrigeration, ***6.Kaushalendra Kumar Singh***, ***3.Rajesh Kumar*** and ***3.Anjana Gupta***, Mechanical and Mathematics, DTU

56. Multiplex Regulation System With Personalised Recommendation Using ML, Anshul Shroff, Bickey Kumar Shah, Amar Kumar Jaiswal, Pooja Sapra and Manoj Kumar and ***6.Aayush Jha***, CSE and Civil, DTU

57. Non-Lubricated Sliding Wear Performance of LM13 Alloy Foam and its Composite Foams Reinforced with ZrSiO4, Suresh Kumar, ***3.Pradeep Kumar Teotia***, O. P. Pandey and Ranvir Singh Panwar, DTU

58. Novel architecture of four quadrant analog multiplier/divider circuit employing single CFOA, ***6.Ajishek Raj***, ***3.Data Ram Bhaskar*** & ***3.Pragati Kumar***, Electrical, DTU

59. NOVEL APPROACH TO GROUNDWATER CONTAMINANT TRANSPORT MODELLING, ***6.Manoj Kumar*** and ***3.Anunay Gour***, Environmental, DTU

60. Object Tracking in a Zone using DeepSORT, YOLOv4 and TensorFlow, ***3.Dr. Rajni Jindal***, ***8.Aditya Panwar***, ***8.Nishant Sharma*** and ***8.Aman Rai***, DTU

61. Oxadiazoles: moiety to synthesis and utilize, ***6.Poonam***, ***6.Geetika Bhasin***, ***3.Richa Srivastava*** and ***3.Ram Singh***, Applied Chemistry, DTU

62. Performance enhancement of savonius turbine with the application of reorienting blade mechanism, ***8.Prashant Chandra Pujari***, ***8.Amit Jain***, ***8.Devang S. Nath*** and ***3.Naveen Kumar***, Mechanical, DTU

63. Preface—special issue "Energy Efficiency in Building using Intelligent computing for Smart Cities", Anand Paul, Anand Nayyar, ***3.Akshi Kumar*** and Jaffar Alzubi, DTU

64. Proximity effects between two plus-plan shaped high-rise buildings on mean & RMS pressure coefficients, ***6.Suresh Kumar Nagar***, ***3.Ritu Raj*** & ***3.Nirendra Dev***, Civil, DTU

65. QnA System on Educational Textbooks: Digital Library Doubt Support System, ***3.Rahul Gupta***, ***8.Gaurav Dabas***, ***8.Himanshu Yadav*** and ***8.Nijdat Hasnain***, CSE, DTU

66. RANDOM FOREST CLASSIFIER BASED NETWORK INTRUSION DETECTION SYSTEM, ***8.Aadhar Dutta***, IT, DTU

67. Ranking the Factors Influencing e-Trading Usage in Agricultural Marketing, Sanjay Chaudhary and ***3.P. K. Suri***, DSM, DTU

68. Resolving Data Overload and Latency Issues in Multivariate Time Time-Series IoMT Data for Mental Health Monitoring, Divya Gupta, MPS Bhatia, ***3.Akshi Kumar***, CSE, DTU

69. Review on Spray-Assisted Solar Desalination: Concept, Performance and Modeling, ***3.Anil Kumar***, ***8.1.Ravi Kant*** and ***3.1.Samsher***, Mechanical and Environment, DTU

70. Rumour Stance Classification using A Hybrid of Capsule Network and Multi-Layer Perceptron, ***3.Akshi Kumar*** and ***7.Meghna Upadhyay***, CSE, DTU

71. SDE based Unified Scheme for Developing Entropy Prediction Models for OSS, Deepika, Ompal Singh, Adarsh Anand and ***3.Jagvinder Singh***, USME, DTU

72. Sentiment Analysis of Twitter Corpus related to COVID-19 induced Lockdown, ***Kritika Saini***, ***3.Dinesh Kumar Vishwakarma*** and ***6.Chhavi Dhiman***, Electronics & IT, DTU

73. Sentiment Analysis Using XLM-R Transformer and Zero-shot Transfer Learning on Resource-poor Indian Language, ***3.AKSHI KUMAR*** and VICTOR HUGO C. ALBUQUERQUE, CSE, DTU

74. Stability analysis of a logistic growth epidemic model with two explicit time-delays, the nonlinear incidence and treatment rates, ***6.Kanica Goel***, Abhishek Kumar and ***3.Nilam***, Mathematics, DTU

75. State of the art content based image retrieval techniques using deep learning: a survey, ***3.Rajiv Kapoor***, Deepak Sharma and Tarun Gulati, ECE, DTU

76. Survey and Comparison of Video Summarization Techniques, ***3.Anil Singh Parihar***, ***8.Ritvik Mittal***, ***8.Prashuk Jain*** and ***8.Himanshu***, CSE, DTU

77. SURVIVING INTERNATIONAL COMPETITION: ANALYSIS OF MEASURES EMPLOYED BY LOCAL COMPANIES, Vanshika Notani and ***6.Piyush Jain***, Production and Industrial, DTU

78. Sustainable human resource planning for hospitals in tier 2/3 cities: evidence from India, ***3.1.Sanjay Bhattacharya*** and Chandrani Bhattacharya, DSM, DTU

79. TCAD Analysis and Simulation of Double Metal Negative Capacitance FET (DM NCFET), ***6.Yash Pathak***, ***3.Bansi Dhar Malhotra*** and ***3.Rishu Chaujar***, Physics & Biotechnology, DTU

80. The Impact of COVID-19 on the Global Renewable Energy Industry, ***3.Sanjay Yadav***, ***8.Md. Tanweer Alam*** and ***8.1.Mohammed Sayeed Khan***, Mechanical, DTU

81. Thermodynamic and comparative analysis of ERC and ARC integrated wetethanol fueled HCCI engine for cogeneration of power and cooling, ***6.Mohd Asjad Siddiqui***, Abdul Khaliq and ***3.Rajesh Kumar***, Mechanical, DTU

82. Thidiazuron outpaces 6-benzylamino purine and kinetin in delaying flower senescence in Gladiolus grandiflora by alleviating physiological and biochemical responses, Madhulika Singh and ***8.Neha Tiwari***, Biotechnology, DTU

83. UndarkGAN: Low-light Image Enhancement with Cycle-consistent Adversarial Networks, ***3.Anil Singh Parihar***, ***8.Pratik Anand***, ***8.Aditya Sharma*** and ***8.Amol Grover***, CSE

84. Wideband Slotted Planar Inverted-F Antenna using Eccosorb MCS Absorber for Millimeter-Wave Applications, ***6.Akhilesh Verma*** and ***3.N S Raghava***, ECE, DTU

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

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

World Wide Web
https://doi.org/10.1007/s11280-021-00920-4

A Bi-GRU with attention and CapsNet hybrid model for cyberbullying detection on social media

Akshi Kumar[1] · **Nitin Sachdeva**[1]

Received: 31 March 2021 / Revised: 20 June 2021 / Accepted: 30 June 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

As a constructive mode of information sharing, collaboration and communication, social media platforms offer users with limitless opportunities. The same hypermedia can be transposed into a synthetic and toxic milieu that provides an anonymous, destructive pedestal for online bullying and harassment. Automatic cyberbullying detection on social media using synthetic or real-world datasets is one of a proverbial natural language processing problem. Analyzing a given text requires capturing the existent semantics, syntactic and spatial relationships. Learning representative features automatically using deep learning models efficiently captures the contextual semantics and word order arrangement to build robust and superlative predictive models. This work puts forward a hybrid model, Bi-GRU-Attention-CapsNet (Bi-GAC), that benefits by learning sequential semantic representations and spatial location information using a Bi-GRU with self-attention followed by CapsNet for cyberbullying detection in the textual content of social media. The proposed Bi-GAC model is evaluated for performance using F1-score and ROC-AUC curve as metrics. The results show a superior performance to the existing techniques on the benchmark Formspring.me and MySpace datasets. In comparison to the conventional models, an improvement of nearly 9% and 3% in F-score is observed for MySpace and Formspring.me dataset respectively.

Keywords CapsNet · Bi-GRU · Cyberbullying · Deep learning · Social media

This article belongs to the Topical Collection: *Special Issue on Synthetic Media on the Web*
Guest Editors: Huimin Lu, Xing Xu, Jože Guna, and Gautam Srivastava

✉ Akshi Kumar
akshikumar@dce.ac.in

Nitin Sachdeva
nits.usit@gmail.com

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

Published online: 22 July 2021

1 Introduction

Social media platforms like Facebook, Instagram, Twitter and more are giving people an opportunity to connect with each other across distances. Users have become individual information sources and clearly a part of a limitless sharing community that has broken the chains of the ancient information mass media order. The social media dynamics with pervasive internet access, no-restriction posting, hidden agendas and longitudinal-latitudinal reach has appended a polarized, radicalized, toxic and an abusive facet to it. Social media deception poses a serious threat to people and eventually society as disinformation, misinformation and malinformation contaminate the information streams [3, 34]. Misuse of social media to abuse and inflict harm taints this collaborative-communicative Web as an amplifying tool for negative narratives or events. Social media platforms now offer a completely new arena where bullies are assured anonymity and ready access to their targets incessantly with minimal retaliation. Cyberbullying [4] is a form of manipulation, belittlement, and targeted abuse using mean-spirited messages and negative electronic postings. It can be as straightforward as sending mean, hurtful, rude texts or instant messages to as devious as spreading secrets or rumours about people online. Though bullying in electronic form can have multiple-dimensions, such as exclusion, harassment, outing, trickery, cyberstalking, dissing, fraping, masquerading, trolling and flaming [16, 31], the obvious intention to hurt and harm is common. This inveterate nuisance creates mental, emotional and physical risks for the bullied. The targets (victims of cyberbullying) feel overwhelmed, powerless, vulnerable, unsafe, worthless, humiliated, isolated, depressed, embarrassed, vengeful and at times suicidal. An accurate detection can facilitate timely intervention by alarming the moderators to take countermeasures. But content moderation practices on these platforms by human moderators is often inconsistent and done in a non-transparent manner. It also suffers from biasing and may apprehend freedom of expression online. Moreover, spotting bullying instances explicitly as well as spotting victims is tricky too. Blocking and reporting might augment the reality if the bully is within the same professional or personal community, for example, a classmate. Sadly, the scale and impact of cyberbullying can be seen across social media platforms even though its awareness is at an all-time high. Figure 1 shows how social media platforms are hotbeds of cyberbullying activities for young people.

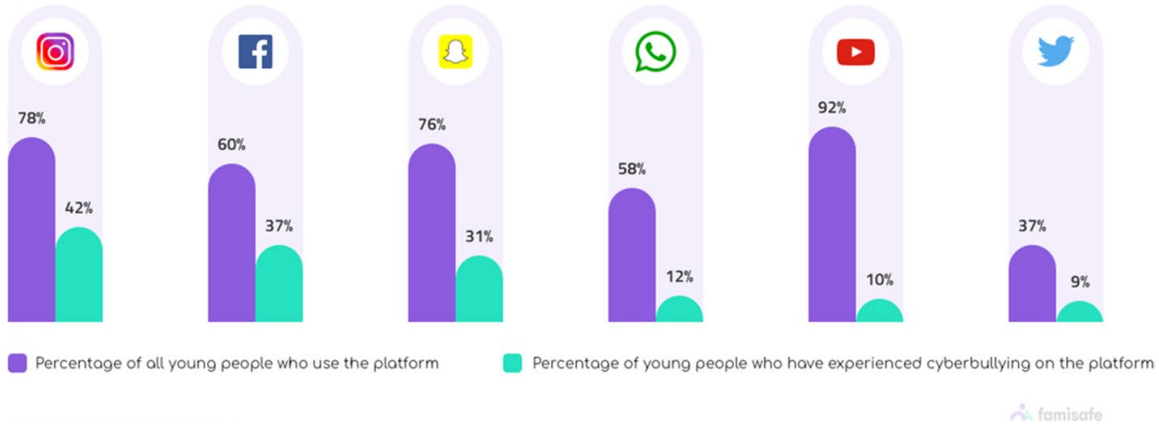


Figure 1 Statistics on share of social media platforms where cyberbullying occurs. https://www.ditchthelabel.org/

With the volume and variety of user-generated content on complex social media platforms the challenges to detect cyberbullying in real-time have amplified. The influx of content makes it challenging to timely regulate online expression. Moreover, the anonymity and context-independence of expressions in online posts can be ambiguous or misleading. The automated cyberbullying detection has attracted growing interest over the past decade as it facilitates combating toxic online behaviour. A lot of research has been done on detecting cyberbullying in textual data using a myriad of features [31]. Many datasets have been made open-source to facilitate research enthusiasts. Efforts from an interdisciplinary research space of computational social science and computational linguistics are required to deal with this antagonistic online delinquency. As a classical problem in natural language processing (NLP), cyberbullying detection in real-time user generated content needs high-level semantic analysis. Most of the earlier attempts on cyberbullying detection rely on manual feature extraction methods [36]. Such methods are not only time-consuming and cumbersome, but often fail to correctly capture the meaning of the sentence. Few lexicon-based methods by maintaining a list of offensive, abusive and hateful words have also been used, but are quite limited in scope [10]. Recent research focuses on the application of deep learning models for various NLP tasks and has reported state-of-the-art results [37]. Basically, deep architectures are neural networks with multiple processing layers of neurons with each layer having a specific task [2]. Utilizing deep learning models trivializes the need of explicit feature extraction techniques as these models are highly skillful and fast in retrieval of essential features and patterns by themselves. With minimal human intervention these models report superior results than the conventional machine learning models. Various deep learning architectures have contributed significantly in computational analytics of text [37]. Pre-trained word embeddings like Word2Vec, GloVe, ELMo, fastText that represent text in vector forms and deep neural networks such as CNN, RNN, GRU, LSTM, CapsNet & hierarchical networks that automate the task of feature extraction demonstrate best practices for solving text classification problems [17, 22].

In this paper, we propose a Bi-GRU-Attention-CapsNet (Bi-GAC) model to detect bullying in online textual content. The model benefits by learning both sequential semantic representations and spatial location information in textual content and acquires better generalization and prediction capabilities for cyberbullying tasks. The improved text representation and feature learning offers a robust model which can avoid the vanishing gradient problem in comparison to baseline neural models. Capsule networks (CapsNet) [20, 40] have the competence to express the semantic meanings in a wider vector space using which it captures the input instantiation parameters, such as word order and their semantic representation. The feature encoding ability of capsules and dynamic routing allow aggregating the important information. Simultaneously, bi-directional gated recurrent unit (Bi-GRU) [8, 9] are known for their sequential modeling capabilities which learns relationships between words from both directions. GRUs are easy and fast to train on smaller sequence data like social media posts, making them apposite for various text classification tasks on social media. Further, self-attention mechanism [8] helps to model dependencies between different parts of the sequence and captures important information using the mutual input interaction. Self-attention based neural networks basically extract neural features for the input sentence [32]. Thus apprehending the dexterity of Bi-GRU, self-attention and CapsNet, we put forward a hybrid model, Bi-GAC to classify online posts into bullying categories. The pre-trained ELMo word embedding [28] is used to create the input embedding matrix. Bi-GAC uses the fully connected output layer with sigmoid activation to finally classify the positive as positive (+1) of bullying or negative (-1) of bullying.

The model is validated on two benchmark datasets, Formspring.me [26] and MySpace [5] and compared with state-of-the-art model using the F1-score as performance measure. An ablation study is also done to ratify the results. Thus, the primary work undertaken in this research includes:

- Building a Bi-GRU-Attention-CapsNet (Bi-GAC) model for bully content classification
- Validating improved classifier performance in small sequences like social media posts by capturing semantic information, context and dependencies
- Evaluation of Bi-GAC on two benchmark datasets

The paper is structured as follows: Sect. 2 briefs about the related studies from pertinent literature followed by the details of the proposed Bi-GAC model in Sect. 3. Section 4 discusses the results whereas the final Sect. 5 presents the conclusion and future work.

2 Related studies

Social misbehavior on online portals has been studied across various dimensions such as hate-speech, aggression detection, comment toxicity and detecting cyberbullying [31]. Earlier works on cyberbullying are primarily available to understand the socio-cognitive aspects and behavioural intentions of bullies and victim experiences with the help of questionnaires and online surveys [14]. Studies are available on cause and effects of cyberbullying on teenagers, adolescents, school, college, ethnic and minority groups [13, 35]. Two surveys in 2019 were reported on automatic cyberbullying detection. Rosa et al. [29] conducted an in-depth analysis of 22 studies on automatic cyberbullying. Kumar and Sachdeva [16] did a meta-analysis on application of soft computing techniques for cyberbullying detection on social multimedia. Studies reveal most of the work on cyberbullying detection has been done using hand-crafted features with machine learning. Recent works using deep learning models have reported superlative results [7]. Agrawal et al. [1] implemented CNN, LSTM and its variants using attention using Twitter, Formspring and Wikipedia datasets. Meng et al. [25] applied a two-branch parallel neural network with multi-head self-attention mechanism (MHSA), capsule network (CapsNet) and independent recurrent neural network (IndRNN). Paul et al. [27] did the comparative analysis with the slot-gated or attention-based DL models using BERT for CB identification. Liu et al. [23] proposed a model for multi-label text classification using ELMo and attention with GRU on Kaggle's toxic comment classification data. Krešnáková et al. [24] carried out experimentation on the same Kaggle's dataset using different text pre-processing technique with DL models such as CNN, GRU, Bi-LSTM + CNN, Bi-GRU + CNN. Özel et al. [26] implemented various ML models such as SVM, kNN etc. on Formspring.me dataset using varied feature selection methods. Zhao et al. [39] performed automatic detection of cyberbullying using bag-of-words and latent semantic features on Twitter dataset.

3 The Bi-GAC model for cyberbullying detection

The Bi-GAC model which gets its nomenclature from its constituent core components, realizes the complexities of textual data in user-generated content where data representation is learned as real-valued vectors. The model has two core components, namely representation

learning and classification. A Bi-GRU encoder is trained using ELMo embedding to generate a sequence context feature vector. This feature vector is flawed due to the existence of redundant and irrelevant features. Consequently, a self-attention mechanism is added to capture significant information. Next the CapsNet generates semantic representation using a dynamic routing algorithm which is finally used for classification of the posts. Figure 2 depicts the architecture of Bi-GAC model.

3.1 Embedding layer

Word embeddings are based on the idea of distributional meaning: the fact that semantically (or morphologically) related words tend to appear in similar contexts. These represent each word using a continuously valued, lower-dimensional vector so that it reserves semantic information of the word. Once word embeddings have been trained, we can use them to derive similarities between words, as well as other relations. The pre-trained ELMo (Embeddings from Language Models) embedding [28] are used in this work. ELMo has an advantage above other conventional embeddings such as GloVe and word2vec as it encapsulates context in the word feature representations. These are high-dimensional representations of words, based on the contexts that different words appear in. ELMo uses a 2-layer bidirectional LSTM for learning words and their context. This design allows ELMo to learn more context-dependent aspects of word meanings in the higher layers along with syntax aspects in lower layers. Figure 3 shows an example of how an ELMo specific representation is generated by combining the bidirectional hidden representations.

3.2 Encoding layer

Recurrent neural network (RNN) models [6] have been popularly used in the field of natural language processing. But, seq2seq encoder/decoder architecture suffers from the vanishing gradients problem, that is the inability of the RNN unit to reminisce values

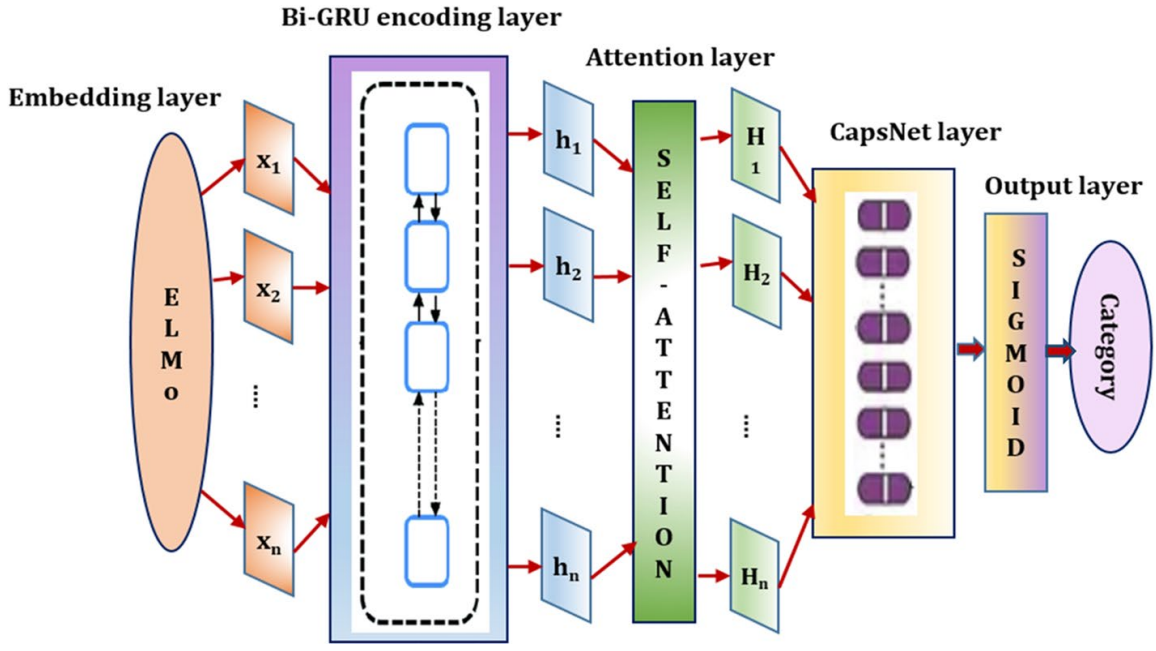


Figure 2 System architecture of proposed model

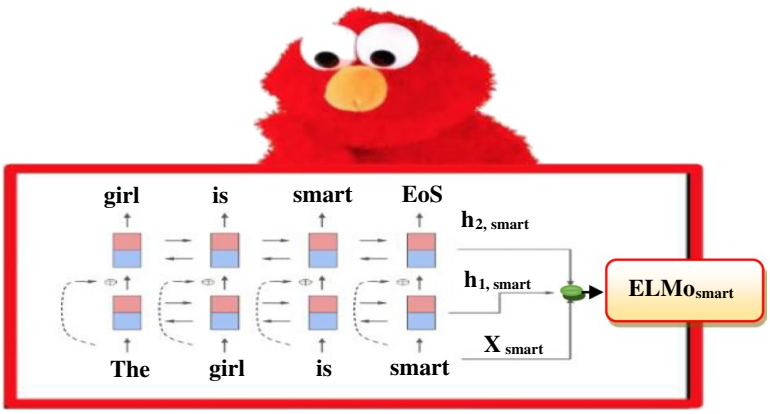


Fig. 3 ELMo-specific representation for "smart"

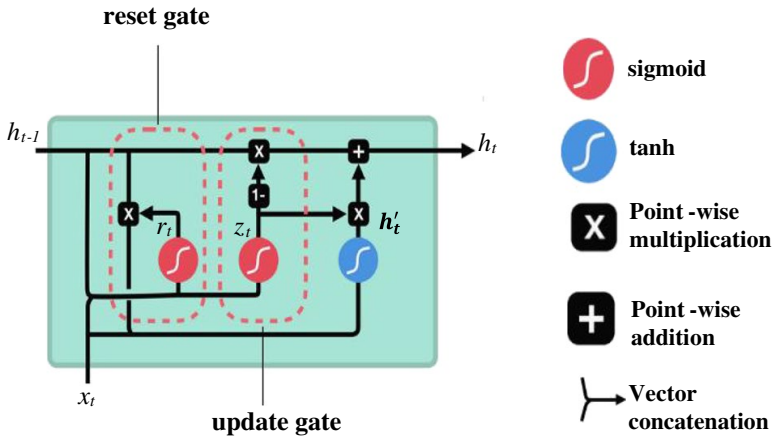
that showed initially in the sequence [11]. But this is vital within a typical NLP task like text classification, as few words rely on words that appear very early in the sentence. As a solution, both gated recurrent unit (GRU) & long-short-term memory (LSTM) solve the problem of vanishing gradients in RNN, by replacing the neurons of the hidden layer with the memory-units to store early sequence data [33]. GRU uses few training parameters, less memory, needs fewer data to generalize, execute faster and train faster than LSTM whereas LSTM is more accurate on dataset using longer sequence. As social media posts are shorter sequence, GRU is an apt choice for encoding. A GRU is a gating mechanism in RNN similar to a LSTM unit but without an output gate. The two gates of a GRU are as follows:-

- Update Gate: It determines how much of the past knowledge needs to be passed along into the future.
- Reset Gate: It determines how much of the past knowledge to forget.

The reset gate sits between the previous activation and the next candidate activation to forget the previous state, and the update gate decides how much of the candidate activation to use in updating the cell state. These gates aid in dealing with the long-term-dependencies. They forward and backward pass the information from the previous state to the next state. The update gate keeps a track of retaining all the important features and also aids in solving the long term temporal dependency issues. A vector having values from 0 to 1 is received via update gate with point-wise multiplication-operation (PMO). It uses sigmoid activation function for squashing values b/w 0 and 1. It basically helps in updating or forgetting (or disappearing) the data as the PMO would result in 0 for any multiplication with a vector of zeros. In such a scenario, the resulting values would be considered to have 'disappeared or be forgotten'. Contrastingly, for any multiplication with a vector of ones would result in the 'same value or be kept'. This mechanism would eventually help in retaining the relevant data and forgetting the non-relevant ones. The other gate which is the reset gate is normally used to ensure the amount of past information that can be forgotten. Figure 4 depicts the architecture of a basic GRU cell.

In this research, we use a bidirectional GRU's (Bi-GRU) encoder which takes the input sequence and encapsulates the information as the internal state vectors. A Bi-GRU allows capturing information from both previous time steps and later time steps to make predictions about the current state. Bi-GRU enables apprehending meaning and context for the sentences than a simple GRU. The forward GRU $\vec{f}$ reads the sentence s_i from w_{i1} to w_{iT} as given in (1):

Figure 4 GRU Cell architecture



$$\vec{h}_{\mathrm{it}} = \overrightarrow{GRU}\left(X_{it}\right),\ \mathrm{t} \in [1, \mathrm{T}] \tag{1}$$

The backward GRU $f^{\leftarrow}$ reads sentences from w_{iT} to W_{i1}, as given in (2):

$$\overleftarrow{h}_{it} = \overleftarrow{GRU}\left(X_{it}\right),\ t \in [T, 1] \tag{2}$$

The annotation of the word w_{it} is calculated by combining the forward and backward hidden states i.e., $h_{\mathrm{it}} = [\vec{h}_{\mathrm{it}}, \overleftarrow{h}_{\mathrm{it}}]$.

3.3 Attention layer

Attention mechanism allows output to focus attention on input while producing output [12, 21] whereas a self-attention model allows inputs to interact with each other, that is, calculate attention of all other inputs with respect to one input. Self-attention is good at modeling dependencies between different parts of the sequence. Self- attention includes both location and observation value information and replaces conditioning on the entire sequence with pairwise comparisons (the importance of one word and its position conditional on some other word and its location) given as vector representations of both. Consider the sentence "*The dog didn't eat the food because it was too full*". The word *"it"* refers to the dog. If we replace *"full"* with *"much"*, the word *"it"* now refers to the food. Attention mechanism helps to understand this, that is, in the former case there's high attention linking *"it"* and *"dog"* but in the latter case high attention shifts to *"food"*. In this research, a self-attention mechanism [32] as shown in Figure 5 is applied on the outputs of the GRU layer.

3.4 CapsNet layer

Conventionally, artificial neurons output a scalar, real-valued activation that loosely represents the probability of an observation. CapsNets [15, 40] replace the scalar output feature detectors with vector output capsules and max-pooling with routing-by-agreement. A capsule is a collection of neurons which represent distinctive properties of the same entity as outputs. Capsule's activity vector represents the instantiation parameters of a specific type of entity which adequately captures the important and relevant information (features). The CapsNet is divided into two layers: primary capsule layer and digit capsule layer. As the

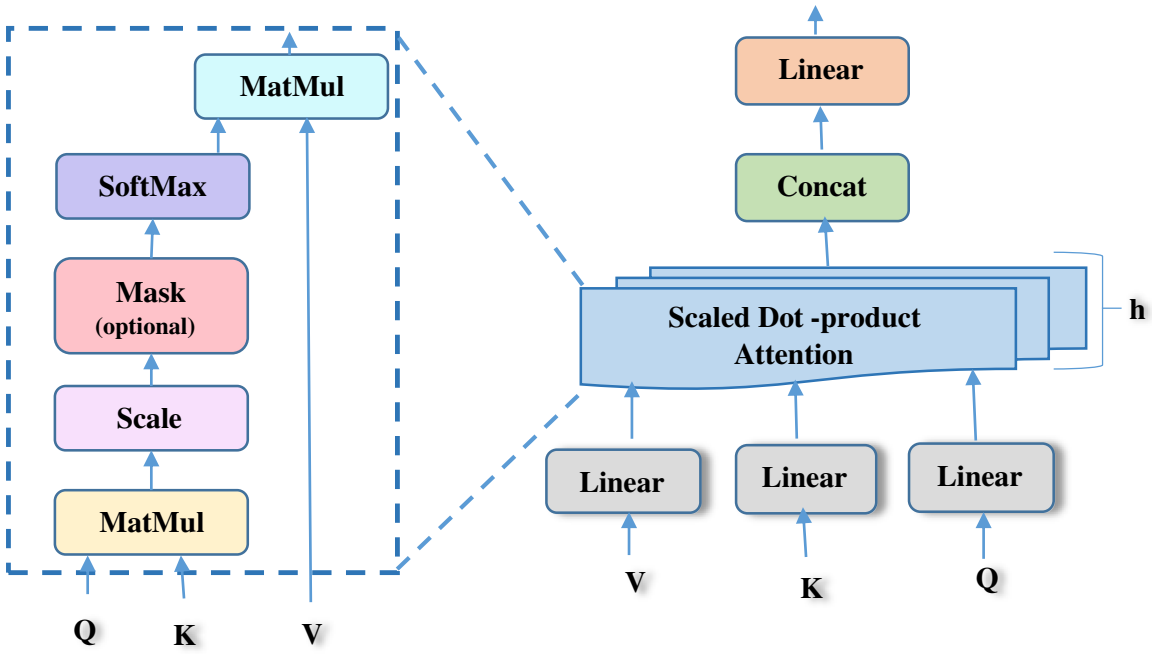


Figure 5 Self-Attention [32]

preceding Bi-GRU attention layer outputs high-level feature representation, these feature maps are input to the primary caps and subsequent digit caps layer as shown in Figure 6. Capsules with vector outputs are generated by the primary caps layers as given (3).

$$u_{j|i} = W_{ij} u_i \tag{3}$$

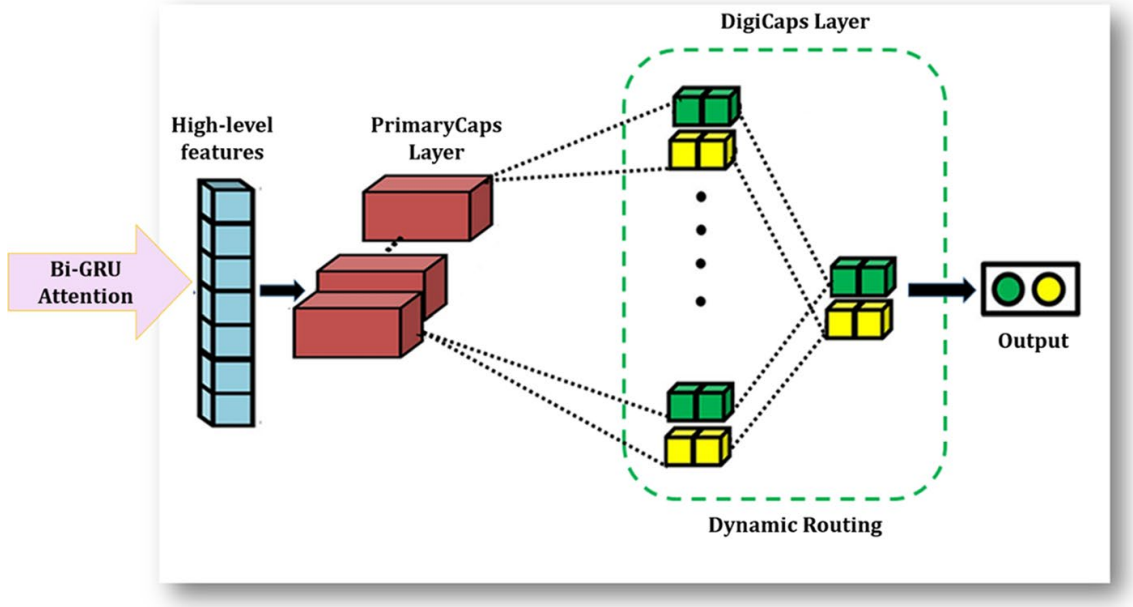


Figure 6 CapsNet Architecture

where, 'i' is in the current lower-level primary-caps and 'j' is in the next-level layer. Connection-weight updating (as shown in Eq. 4) will occur during network training using dynamic routing algorithm.

$$s_j = \sum_i c_{ij} u_{j|i} \tag{4}$$

The key operations within the capsule are as follows [19]:

- Multiplying input vector matrix and weights to encode significant features and their relationships within the text.
- Dynamic routing for sending output from one low-level capsule to a higher level capsule.
- Weighted input vectors summation.
- "Squash" function to add non-linearity using. This function takes a vector and "squashes" it to have a maximum length of 1, and a minimum length of 0 while retaining its direction.

Figure 7 summarizes the operations within a capsule.
The dynamic routing algorithm [30] is given below (algorithm 1).

Algorithm 1. Dynamic routing algorithm

procedure ROUTING($\hat{u}_{j|i}$, r, l)
  for all capsule i in layer l and capsule j in layer (l + 1): $b_{ij} \leftarrow 0$.
  for r iterations **do**
    for all capsule i in layer l: $c_i \leftarrow \text{softmax}(b_i)$
    for all capsule j in layer (l + 1): $s_j \leftarrow \sum_i c_{ij} \hat{u}_{j|i}$
    for all capsule j in layer (l + 1): $v_j \leftarrow \text{squash}(s_j)$
    for all capsule i in layer l and capsule j in layer (l + 1): $b_{ij} \leftarrow b_{ij} + \hat{u}_{j|i} . v_j$
return v_j

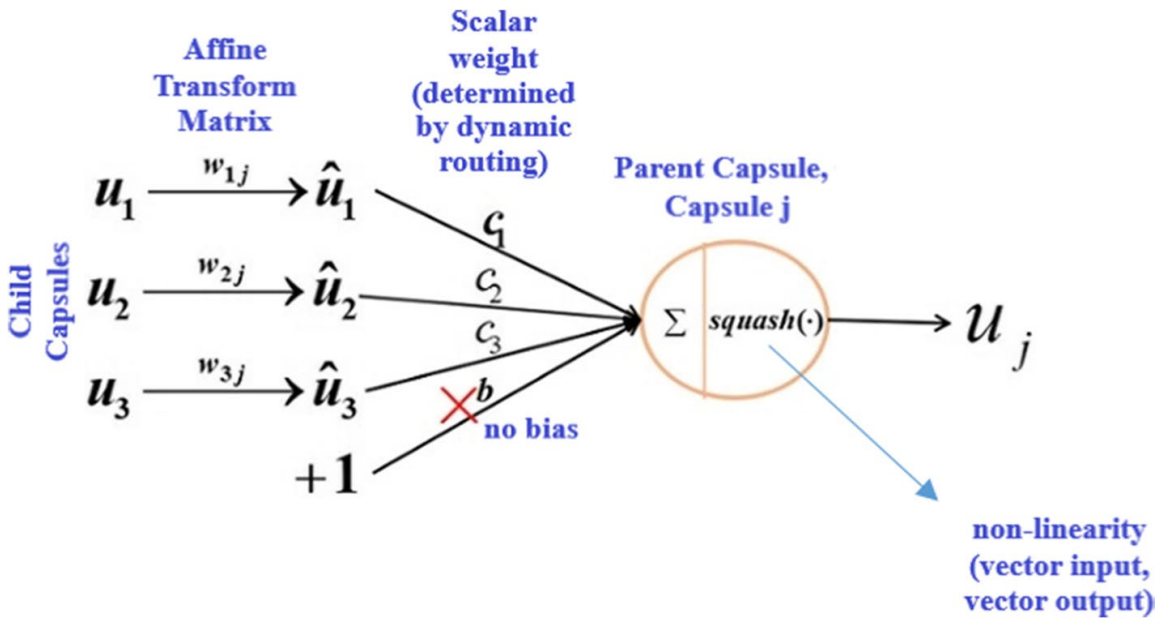


Figure 7 Operations within a Capsule [22]

Table 1 Sample distribution in datasets

Dataset	Bullying samples	Non-Bullying samples	Total samples
Formspring.me [26]	836	12,288	13,124
MySpace [5]	357	1396	1753

Table 2 F-1 Score for Bi-GAC

Dataset	F1-Score
Formspring.me	94.03
MySpace	93.89

3.5 Output layer

The final prediction layer of the proposed model is the fully-connected layer with sigmoid activation that eventually helps in obtaining the probabilities for the binary classification (target classes: bullying and non-bullying). It basically maps a real valued number using a threshold to a probability (i.e. to a no. b/w 0 to 1).

4 Computational results

Various social media datasets from Twitter, YouTube, MySpace, Kongregate, Formspring and Slashdot have been created for automatic cyberbullying detection. In this work, two benchmark datasets, namely, question-answering Formspring.me and thread-style MySpace social network cyberbullying classification datasets are used for experimentation. The datasets are labelled for cyberbullying and non-bullying type and the sample distribution is as given in Table 1.

Evidently, the datasets suffer from class imbalance and we over-sampled the data from the bullying class (sporadic class) thrice to deal with this skewness. This technique is usually used for balancing the corpus in similar studies [1]. The training: test ratio was taken as 70:30 and tenfold cross validation is used on each dataset. Using 10 folds means that in each iteration of cross-validation the validation-set would be approximately 10% of the size of the total dataset. The results are then averaged across the folds, using suitable performance measures. We used the Scikit-learn library and Keras deep learning library with Theano backend. The hyperparameters in experiments were set as follows: The Bi-GRU layer had 50 units and a dropout value of 0.2. The CapsNet layer also had a drop out of 0.2 and had 3 iterations for dynamic routing. Adam optimizer was used with the learning rate of 0.0001. The Bi-GAC model was evaluated using the F-1 score. The results for both datasets are shown in Table 2.

The results are also compared with the existing techniques used on both datasets. The performance comparison of the proposed Bi-GAC model with the existing techniques is shown in Tables 3 and 4 for MySpace and Formspring.me datasets respectively. A superior performance is observed for the proposed Bi-GAC model.

Table 3 Comparison of Bi-GAC with existing works on MySpace dataset

Reference Study	Techniques	F1-score
Kumar & Sachdeva [18]	k-nearest neighbour	67.7
Kumar & Sachdeva [18]	J48	69.6
Zhang et al. [38]	Logistic Regression	78
Zhang et al. [38]	Support Vector Machine	79
Zhang et al. [38]	CNN	85
Proposed Model	Bi-GAC	93.89

Table 4 Comparison of Bi-GAC with existing works on formspring.me dataset

Reference Study	Techniques	F1-score
Agrawal & Awekar [1]	Random Forest	29.8
Agrawal & Awekar [1]	Naive Bayesian	35.9
Agrawal & Awekar [1]	Support Vector Machine	42.2
Agrawal & Awekar [1]	Logistic Regression	44.8
Agrawal & Awekar [1]	Bi-LSTM	86
Paul & Saha [27]	RNN + LSTM	88
Paul & Saha [27]	Bi-LSTM with Attention	91
Paul & Saha [27], Agrawal & Awekar [1]	CNN	91
Proposed Model	Bi-GAC	94.03

Table 5 Ablation architectures

Dataset	Bi-LSTM + Attention + CapsNet	Bi-GRU + Attention + CNN	Bi-GRU + Attention + CapsNet
Formspring.me	92.67	91.83	94.03
MySpace	93.10	92.35	93.89

We also perform an ablation study to learn about the network by removing and/or replacing parts of the complex neural network architecture and study the model performance. The two variations studied are: using Bi-LSTM instead of Bi-GRU (Bi-LSTM + Attention + CapsNet) and using CNN instead of CapsNet (Bi-GRU + Attention + CNN). Table 5 presents the F1-score of these variations where the proposed Bi-GAC (Bi-GRU + attention + CapsNet) outperforms the other two.

Figure 8 shows the ROC (Receiver Operating Characteristic)—AUC (Area under curve) curve. ROC curve tells us about how good the model can distinguish between two classes. The AUC score gives a good idea of how well the model performs.

5 Conclusion

Social media serves as a virtual playground for bullying and allows the offenders to be mysterious, tough to trace and shielded from confrontation making the detection of bullying instances a challenging task. It can become a source of misery for victims and

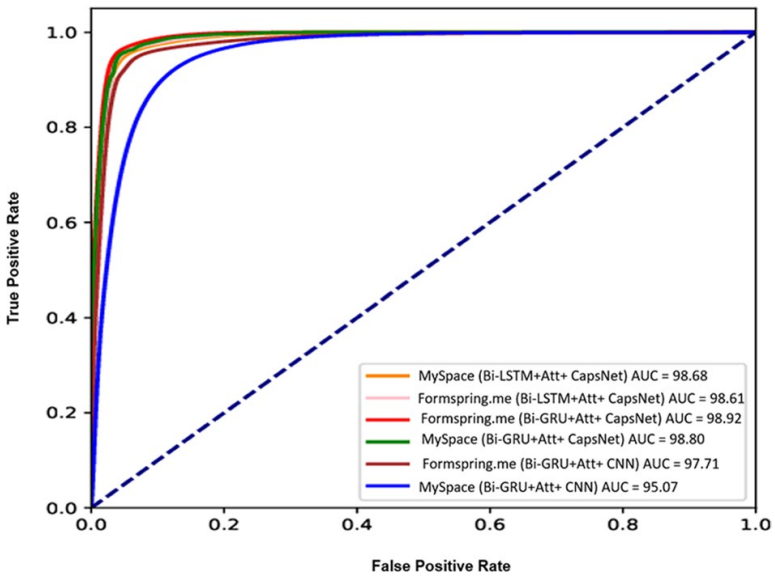


Figure 8 ROC-AUC for MySpace and Formspring.me datasets

often lead to mental distress and depression, and in extreme situations, even suicide. It is necessary to be both responsive and proactive to such a toxic environment on social media platforms. This research proposed a hybrid deep learning model for cyberbullying classification task which combines the advantages of self-attention based Bi-GRU encoder and capsule network. ELMo contextual embeddings are used as input. The model achieves a superior F1-score of 94.03 and 93.89 for Formspring.me and MySpace benchmark cyberbullying datasets respectively.

As a future direction, we would like to use transformer-based models like BERT, especially on multimodal and multilingual cyberbullying datasets. Fine-grain classification into cyberbullying categories has not been explored as yet and we would like to extend the model to multi-class cyberbullying categories. Cyberbullying divides the stakeholders into categories of offender, victim and by-stander (assistants, reinforcers, defenders and outsiders) and therefore combining user-based features with text-based features to train models for real-time cyberbullying needs investigation.

Author's contributions All the authors have contributed equally in the research and manuscript preparation.

Data availability Benchmark publicly available datasets have been used.

Code availability Can be made available on request.

Declarations

Ethics approval The work conducted is not plagiarized. No one has been harmed in this work.

Consent to participate All the authors have given consent to submit the manuscript.

Consent for publication Authors provide their consent for the publication.

Conflict of interest The authors certify that there is no conflict of interest in the subject matter discussed in the manuscript.

References

1. Agrawal, S., Awekar, A.: Deep learning for detecting cyberbullying across multiple social media platforms. In: European Conference on Information Retrieval, pp. 141–153. Springer, Cham (2018)
2. Ballal, N, Saritha, SK: A study of deep learning in text analytics. In: Shukla, R, Agrawal, J, Sharma, S, Chaudhari, N, Shukla, K (eds.) Social networking and computational intelligence. Lecture notes in networks and systems, vol. 100. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-2071-6_16
3. Bounegru, L, Gray, J, Venturini, T, Mauri, M: A field guide to 'fake news' and other information disorders. A field guide to" fake news" and other information disorders: a collection of recipes for those who love to cook with digital methods. Public Data Lab, Amsterdam (2018)
4. Campbell, MA: Cyber bullying: an old problem in a new guise? Aust. J. Guid. Couns. **15**(1), 68–76 (2005)
5. Çiğdem, A.C.I., Çürük, E., Eşsiz, E.S.: Automatic detection of cyberbullying in FORMSPRING. Me, Myspace and Youtube Social Networks. Turk J Eng **3**(4), 168–178 (2019)
6. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)
7. Dadvar, M., Eckert, K.: Cyberbullying detection in social networks using deep learning based models; a reproducibility study. arXiv preprint arXiv:1812.08046 (2018)
8. Deng, J., Cheng, L., Wang, Z.: Self-attention-based BiGRU and capsule network for named entity recognition.arXiv preprint arXiv:2002.00735 (2020)
9. Gangwar, A. K., Ravi, V.: A novel BGCapsule network for text classification. arXiv preprint arXiv:2007.04302 (2020)
10. Hang, O.C., Dahlan, H.M.: Cyberbullying lexicon for social media. In: 2019 6th International Conference on Research and Innovation in Information Systems (ICRIIS), pp. 1–6. IEEE (2019)
11. Hochreiter, S: The vanishing gradient problem during learning recurrent neural nets and problem solutions. Internat. J. Uncertain. Fuzziness Knowledge-Based Systems **6**(02), 107–116 (1998)
12. Jain, D, Kumar, A, Garg, G: Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature-rich CNN. Appl. Soft Comput. **91**, 106198 (2020)
13. John, A., Glendenning, A. C., Marchant, A., Montgomery, P., Stewart, A., Wood, S., ... Hawton, K.: Self-harm, suicidal behaviours, and cyberbullying in children and young people: systematic review. J. Med. Internet Res. **20**(4), e129 (2018)
14. Kowalski, RM, Limber, SP: Psychological, physical, and academic correlates of cyberbullying and traditional bullying. J. Adolesc. Heal. **53**(1), S13–S20 (2013)
15. Kim, J, Jang, S, Park, E, Choi, S: Text classification using capsules. Neurocomputing **376**, 214–221 (2019)
16. Kumar, A., Sachdeva, N.: Cyberbullying detection on social multimedia using soft computing techniques: a meta-analysis. Multimed Tools Appl **78**(17), 23973–24010 (2019)
17. Kumar, A, Jaiswal, A: A deep swarm-optimized model for leveraging industrial data analytics in cognitive manufacturing. IEEE Trans. Industr. Inf. **17**(4), 2938–2946 (2020)
18. Kumar, A., Sachdeva, N.: Cyberbullying checker: online bully content detection using Hybrid Supervised Learning. In: International Conference on Intelligent Computing and Smart Communication 2019, pp. 371–382. Springer, Singapore (2020)
19. Kumar, A., Sachdeva, N.: Multi-input integrative learning using deep neural networks and transfer learning for cyberbullying detection in real-time code-mix data. Multimedia Systems, 1–15 (2020)
20. Kumar, A., Sachdeva, N.: Multimodal cyberbullying detection using capsule network with dynamic routing and deep convolutional neural network. Multimedia Systems, 1–10 (2021)
21. Kumar, A., Sangwan, S.R., Arora, A., Nayyar, A., Abdel-Basset, M.: Sarcasm detection using soft attention-based bidirectional long short-term memory model with convolution network. IEEE Access **7**, 23319–23328 (2019)
22. Kumar, A., Srinivasan, K., Cheng, W.H., Zomaya, A.Y.: Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data. Inf Process Manag **57**(1), 102141 (2020)
23. Liu, W., Wen, B., Gao, S., Zheng, J., Zheng, Y.: A multi-label text classification model based on ELMo and attention. In: MATEC Web of Conferences, Vol. 309, pp. 03015. EDP Sciences (2020)
24. Maslej-Krešňáková, V, Sarnovský, M, Butka, P, Machová, K: Comparison of deep learning models and various text pre-processing techniques for the toxic comments classification. Appl. Sci. **10**(23), 8631 (2020)

25. Meng, Z, Tian, S, Yu, L: Regional bullying text recognition based on two-branch parallel neural networks. Autom. Control. Comput. Sci. **54**(4), 323–334 (2020)
26. Özel, SA, Sarac, E: Effects of feature extraction and classification methods on cyberbully detection. Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi **21**(1), 190–200 (2016)
27. Paul, S., & Saha, S. (2020). CyberBERT: BERT for cyberbullying identification.*Multimedia Systems*, 1–8.
28. Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations.arXiv preprint arXiv:1802.05365 (2018)
29. Rosa, H, Pereira, N, Ribeiro, R, Ferreira, PC, Carvalho, JP, Oliveira, S, Coheur, L, Paulino, P, Simão, AV, Trancoso, I: Automatic cyberbullying detection: a systematic review. Comput. Hum. Behav. **93**, 333–345 (2019)
30. Sabour, S., Frosst, N., Hinton, G. E.: Dynamic routing between capsules. arXiv preprint arXiv:1710.09829 (2017)
31. Sangwan, S.R., Bhatia, M.P.S.: D-BullyRumbler: a safety rumble strip to resolve online denigration bullying using a hybrid filter-wrapper approach. Multimedia Systems, 1–17 (2020)
32. Shao, Y, Lin, JCW, Srivastava, G, Jolfaei, A, Guo, D, Hu, Y: Self-attention-based conditional random fields latent variables model for sequence labeling. Pattern Recogn. Lett. **145**, 157–164 (2021)
33. Sherstinsky, A: Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D **404**, 132306 (2020)
34. Shrivastava, G., Kumar, P., Ojha, R.P., Srivastava, P.K., Mohan, S., Srivastava, G.: Defensive modeling of fake news through online social networks. IEEE Trans Comput Soc Syst **7**(5), 1159–1167 (2020)
35. Smith, PK, Mahdavi, J, Carvalho, M, Fisher, S, Russell, S, Tippett, N: Cyberbullying: Its nature and impact in secondary school pupils. J. Child Psychol. Psychiatry **49**(4), 376–385 (2008)
36. Van Hee, C, Jacobs, G, Emmery, C, Desmet, B, Lefever, E, Verhoeven, B, Hoste, V: Automatic detection of cyberbullying in social media text. PLoS ONE **13**(10), e0203794 (2018)
37. Young, T, Hazarika, D, Poria, S, Cambria, E: Recent trends in deep learning based natural language processing. IEEE Comput. Intell. Mag. **13**(3), 55–75 (2018)
38. Zhang, A., Li, B., Wan, S., Wang, K.: Cyberbullying detection with birnn and attention mechanism. In: International Conference on Machine Learning and Intelligent Communications, pp. 623–635. Springer, Cham (2019)
39. Zhao, R., Zhou, A., Mao, K.: Automatic detection of cyberbullying on social networks based on bullying features. In: Proceedings of the 17th international conference on distributed computing and networking, pp. 1–6 (2016)
40. Zhao, W., Ye, J., Yang, M., Lei, Z., Zhang, S., Zhao, Z.: Investigating capsule networks with dynamic routing for text classification. arXiv preprint arXiv:1804.00538 (2018)

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

Proceedings of the ASME 2021
International Mechanical Engineering Congress and Exposition
IMECE 2021
November 1-5, 2021, Virtual, Online

IMECE2021-69923

arXiv:2107.03361v1 [physics.flu-dyn] 7 Jul 2021

A COMPARATIVE STUDY OF VARIOUS DEEP LEARNING TECHNIQUES FOR SPATIO-TEMPORAL SUPER-RESOLUTION RECONSTRUCTION OF FORCED ISOTROPIC TURBULENT FLOWS

T.S.Sachin Venkatesh[1], **Rajat Srivastava**[2], **Pratyush Bhatt**[2], **Prince Tyagi**[2], **Dr. Raj Kumar Singh**[2]

[1]Department of Applied Physics, Delhi Technological University, New Delhi, India
[2]Department of Mechanical Engineering, Delhi Technological University, New Delhi, India

ABSTRACT

Super-resolution is an innovative technique that upscales the resolution of an image or a video and thus enables us to reconstruct high-fidelity images from low-resolution data. This study performs super-resolution analysis on turbulent flow fields spatially and temporally using various state-of-the-art machine learning techniques like ESPCN, ESRGAN and TecoGAN to reconstruct high-resolution flow fields from low-resolution flow field data, especially keeping in mind the need for low resource consumption and rapid results production/verification. The dataset used for this study is extracted from the 'isotropic 1024 coarse' dataset which is a part of Johns Hopkins Turbulence Databases (JHTDB). We have utilized pre-trained models and fine tuned them to our needs, so as to minimize the computational resources and the time required for the implementation of the super-resolution models. The advantages presented by this method far exceed the expectations and the outcomes of regular single structure models. The results obtained through these models are then compared using MSE, PSNR, SAM, VIF and SCC metrics in order to evaluate the upscaled results, find the balance between computational power and output quality, and then identify the most accurate and efficient model for spatial and temporal super-resolution of turbulent flow fields.

Keywords: super-resolution, transfer learning, isotropic turbulent flow fields

1 INTRODUCTION

In the recent years that have gone by, deep learning models have been intensively used to tackle numerous problems in fluid mechanics. There is an exponential increment in the amount of data available through experiments and simulations due to the rapid improvement in the performance of computational architectures and advances in experimental measurement capabilities. Machine learning provides a wealth of techniques that can learn from this data to solve the prominent challenges and problems in fluid dynamics such as experimental data processing, reduced-order modeling, optimization and turbulence closure modeling. Convolutional Neural Networks serve as a powerful tool for the extraction of fluid dynamic features and thus aid in the prediction of flow fields. Thus, various models have been developed based on CNN to refine the results obtained through CFD analysis quickly with remarkable accuracy.

Super-resolution is an important set of image processing techniques that enhances or upscales the resolution of an image or video. These methods thus enable us to reconstruct high-fidelity images from low-resolution ones. In the past few years, with the accelerated growth of techniques of deep learning, various deep learning-based super-resolution models ranging from the classic Convolutional Neural Networks (CNN) to more recent and promising Generative Adversarial Networks (GAN) have been explored and developed which often achieve state-of-the-art performance. It is also to be noted that we undertook this endeavor to not actually solve flow fields or predict them, there are already many out there which are state of the art, but instead

Copyright © 2021 by ASME

we focused on the visualization of the flow field itself which is very often the problem associated with low resource simulations. These techniques when employed on flow fields enabled us in reconstruction of the high-resolution flow (HR) fields from the low- resolution (LR) fluid flow data, thus saving an enormous amount of time and computational power to perform the high-resolution flow field simulations.

Transfer learning is a technique that utilizes pre-trained models and allows us to mix them too, to minimize the computational resources and the time required for the implementation of the super-resolution models. Thus, transfer learning allows models that have been conditioned for one task to be repurposed for another particular task. The Develop model approach and the Pre-trained methodology are the two most widely used methods in transfer learning. In the Develop model approach, a similar predictive modeling problem with plenty of data is solved and any relationship between input and output data is selected, and then a skillful model is developed, trained, and tested for the first problem. This model is then fit on the source task is utilized as the starting point for the model which we need to develop to derive a solution for the problem of interest. This involves either using the entire model or parts of the model, depending on the problem and the modeling technique used. On the basis of the input-output pair data required for the task of interest, the model may also be customised or optimised. In a pre-trained model approach, a pre-trained source model is chosen from available models already trained on large and challenging datasets released by various research institutions. Then, this model acts as an initial point for development of the model specific to the problem statement. Either the entire model or parts of it are utilized and tuned to achieve the desired results.

2 LITERATURE REVIEW

In 2019, a data-driven approach was presented by Sekar, Qinghua, Chang, and Khoo for the prediction of incompressible laminar steady flow fields over airfoils developed by combining CNN and deep Multilayer Perceptron (MLP) [1]. Subgrid modeling assisted by machine learning for large eddy simulation (LES) was proposed by Maulik in 2019 [2]. Combined CNN and Long short-term memory (LSTM) were used by Hasegawa for developing a reduced-order model (ROM) for a two-dimensional unsteady wake behind various bluff bodies in 2020 [3]. In 2019, Fukami, Fukagata, and Taira [4] performed super-resolution to reconstruct the high-resolution flow field of homogeneous decaying isotropic turbulence on low resolution turbulent flow field data using CNN and hybrid down sampling skip multiscale (DSC/MS) models. The multiscale model showed remarkable accuracy in the generation of HR flow fields from LR snapshots. Then in 2020, they performed super-resolution analysis based on DSC/MS machine learning model using spatial super-resolution and temporal inbetweening techniques which was able to recover high-resolution turbulent flow fields from low resolution or coarse flow data in space and time [5]. In 2019, Deng, He, Liu, Kim [6] utilized advanced neural networks like super-resolution generative adversarial networks (SRGAN) and enhanced-SRGAN (ESRGAN) to enhance the spatial resolution of turbulent flow behind two side by side cylinders. Various other deep learning models such as static convolutional neural network (SCNN) and Multiple temporal path convolutional network (MTPC) have also been utilized to fully capture features in different time ranges by Liu, Tang, Huanga and Lu [7]. Techniques for super-resolution and denoising of fluid flow using a physics informed CNN without using any HR labels for training have also been demonstrated by Gao, Sun, Wang [8]. Several flow SR problems relevant to cardiovascular applications were proposed in the study to evaluate the method's effectiveness and merit. In 2020, Stengel, Glaws, Hetinger, and King developed an adversarial deep learning model to super resolve the atmospheric turbulent winds and solar outputs from global climate models up to fifty times [9]. Xu, Luo, Wang and You [**?**] used 3D super-resolution generative adversarial network (3D-SR-GAN) so as to develop a data-driven 3D super-resolution approach which can enhance the spatial resolution of a three dimensional turbulent jet flame structure by two times in all the spatial direction [10]. Meta-Transfer Learning for Zero-Shot Super-Resolution (MZSR) [11], is a technique that finds a generic initial parameter which is suitable for internal learning has been proposed by Sho in 2020. This method thus allows the network to quickly adapt to a given image condition. In 2019, Zhang, Wang, Lin and Qi designed an end-to-end deep learning model [12] which is able to enrich the high-resolution details of the image by adaptive transfer of the the texture from the reference images according to their textural similarity by conducting a multilevel matching in the neural space.

3 MATERIALS AND METHODS

3.1 Super resolution

Super-resolution is an approach to improve the resolution/quality of both spatial and temporal data. An image or video might be in Low-Resolution form either due to downsampling or degradation. The mathematical function $LR = (Degrade)HR$ states the relation between the two aspects of the data. We here observe that the Low Resolution data can be obtained by downsampling the High Resolution data. However here the objective is on the contrary. We have to obtain the HR from the LR and that too in the absence of the downsampling/degrading function. In recent years deep learning methodologies have been proven effective in obtaining the HR i.e., High-Resolution image from the LR i.e., Low Resolution or degraded image. Deep learning models and techniques for super-resolution utilized in our study are explained in the subsequent sections.

Copyright © 2021 by ASME

3.1.1 ESPCN

ESPCN has an Efficient Sub Pixel convolutional layer added to the CNN which enhances the resolution towards the end of the network. The network of ESPCN can be represented as an efficient sub-pixel convolutional layer [13] that includes a convolution having a fractional stride of $1/r$ in LR space with W_s as a filter having size K_s. Weights that are not in alignment with the pixels are to be omitted for calculations. This results in the number of activation patterns to be r^2. These activated patterns have the greatest number of activated weights in accordance with the location.

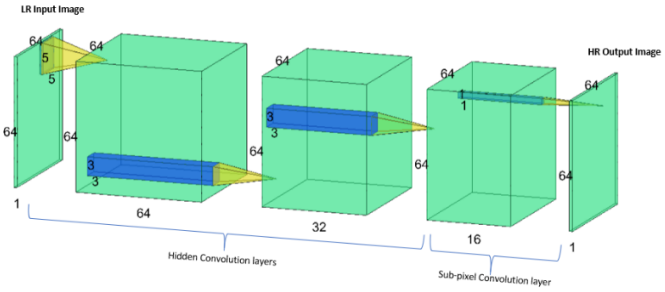


FIGURE 1: ESPCN network architecture

$$I^{SR} = f^{L}\left(I^{LR}\right) = PS\left(\mathrm{W}_{L} * f^{L-1}\left(I^{LR}\right)\right) + \mathrm{b}_{L} \tag{1}$$

where PS is the periodic shuffling operator and W_L, b_L are the weights and biases respectively. The PS or periodic shuffling operator shuffles the elements of the tensor. In certain deconvolution procedures, we pad the images with zeros until convolution, which may have a negative impact on the outcome. While performing pixel shuffle at the network's final layer to recover the LR image, no padding operation is needed. The above equation can be mathematically represented as:

$$PS(T)_{x,y,c} = T_{\left[\frac{x}{r}\right],\left[\frac{y}{r}\right],c.r.mod(y,r)+c.mod(x,r)+c} \tag{2}$$

Thus, W_L having a shape of $n_{L-1} * r^2 * C * k_L * k_L$, this new layer together with CNN layers is termed as Efficient Sub Pixel Convolutional Network or ESPCN. In this paper, the Low-Resolution videos are generated by downsampling the High-Resolution Data so the actual High-Resolution data will be useful in generating the MSE or Mean Squared Error. This MSE is used to calculate the difference between the Super-Resolution Videos generated through the network and the actual High-Resolution Videos. The loss function in ESPCN is given by

$$1(\mathrm{W}_{1:L}, b_{1:L}) = \frac{1}{r^2 HW} \sum_{x=0}^{rH} \sum_{y=0}^{rH} \left(I^{HR}_{x,y} - f^{L}_{x,y}\left(I^{LR}\right)\right)^2 \tag{3}$$

3.1.2 ESRGAN

Enhanced Super-Resolution Generative Adversarial Networks better known as ESRGAN, as the name implies is an upgraded SRGAN [14] which is a super-resolution technique for single images based on generative adversarial network . It employs a perceptual loss mechanism that includes an adversarial loss and a material loss. In ESRGAN the network's overall high-level architecture nature is preserved, but a few new ideas are introduced and updated, resulting in a more effective network. The Residual-in-Residual Dense Block (RRDB) is introduced into the Generator network configuration, which increases the network's power and makes training simpler.

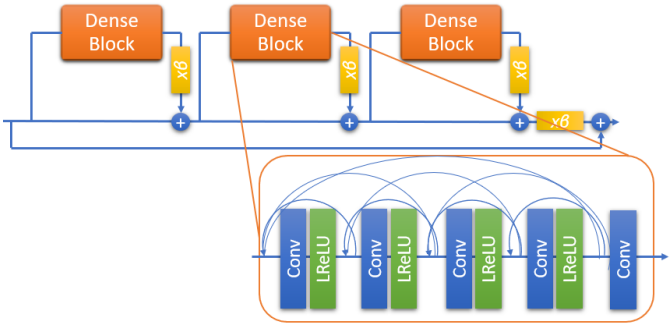


FIGURE 2: Residual in Residual Block (RRDB) of ESRGAN

The ESRGAN model retains the SRGAN high-level architecture framework and uses the new basic component called RRDB [15]. The proposed RRDB structure is more complex and diverse as compared to the initial residual structure in SRGAN. In addition, to the improved design, it employs a number of techniques to facilitate the preparation of a very profound network: 1) Residual scaling, which is the method of decreasing residuals by proliferating them by a constant ranging from 0 to 1 prior addition to the main path to avert the erratic behavior; 2) Decreased initialization, since it was found factually that remnant structure is simpler to train as the initial parameter variance decreases.

In addition to the traditional discriminator, ESRGAN employs the relativistic GAN, which attempts to estimate the likelihood that the actual picture is more realistic than a false image.

$$\begin{aligned} D_{Ra}\left(x_r, x_f\right) &= \sigma\left(C\left(I_r\right) - E\left[C\left(I_f\right)\right] \rightarrow 1\right. \\ D_{Ra}\left(x_f, x_r\right) &= \sigma\left(C\left(I_f\right) - E\left[C\left(I_r\right)\right] \rightarrow 0\right. \end{aligned} \tag{4}$$

Copyright © 2021 by ASME

where D_{Ra} is Relativistic Average Discriminator, I_r and I_f denote Real image and Fake image respectively and E denotes the operation of taking an average of all data in mini-batch. We build a more effective perceptual loss L_{percep} by controlling features before activation rather than after, as SRGAN does. Using features after they have been enabled causes differences in recreated brightness as compared to the ground-truth picture.

$$L_G == L_{percep} + \lambda L_G^{Ra} + \eta L_1 \tag{5}$$

Where L_1 represents the loss that computes the 1-norm distance from reconstructed image to ground truth, λ, η are the loss term balancing coefficients.

3.1.3 TecoGAN

GANs (generative adversarial networks) have shown to be remarkably efficient at solving complicated distributions. Explicitly applying GANs to sequence generation without carefully designed restrictions, on the other hand, usually results in strong artifacts over time. Conditional video generation functions, in particular, have extreme learning difficulties. It presents a recurrent adversarial generative model for VSR that consists of three structures: a recurrent generator, a flow estimation network, and a spatial-temporal discriminator. The generator G, is used to generate significant video frames from low-resolution inputs frequently [16]. The flow estimation network F, learns flow adjustment between frames to support both the generator and the spatial-temporal model.

$$\mathrm{v}_t = \mathrm{F}\left(\mathrm{i}_{t-1}, \mathrm{i}_t\right), \mathrm{g}_t = G\left(\mathrm{a}_t, W\left(\mathrm{g}_{t-1}, \mathrm{v}_t\right)\right) \tag{6}$$

The g_T, high-resolution Output, is generated by the generator from the LR data input x_T and then sequentially uses the previously produced HR output g_{T-1}. The architecture of the generator can be represented as:

In this case, the network F has been trained to evaluate the motion v_t from frame i_{t-1} to i_t and W stands for warping.

The architecture of the discriminator can be seen in the Fig 4 [17]. This sequence of the model is passed with two input sets: i) the actual structure, ii) the generated structure. The shape of both sets is the same, having equal numbers of High-resolution data and corresponding Low-resolution frames. The actual structure has the high resolution frames with natural data and on the other hand, the generated Structure is made up of the data generated by the G. The objective of having two input sets is that the model can be penalized by the Discriminator for the structures generated having less spatial-temporal details or vice versa.

3.2 Transfer Learning

One of the major challenges faced while training a neural network is to figure out the exact amount of the dataset and also

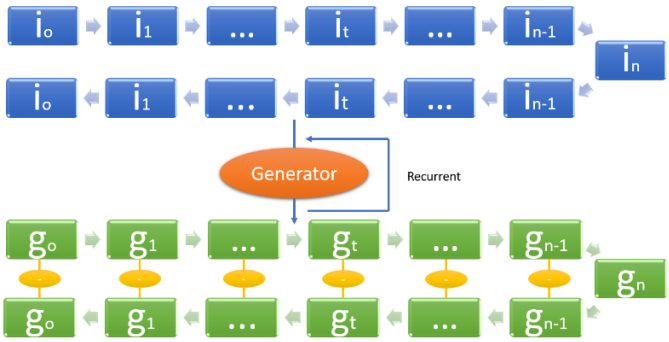


FIGURE 3: Frame Recurrent Generator for VSR

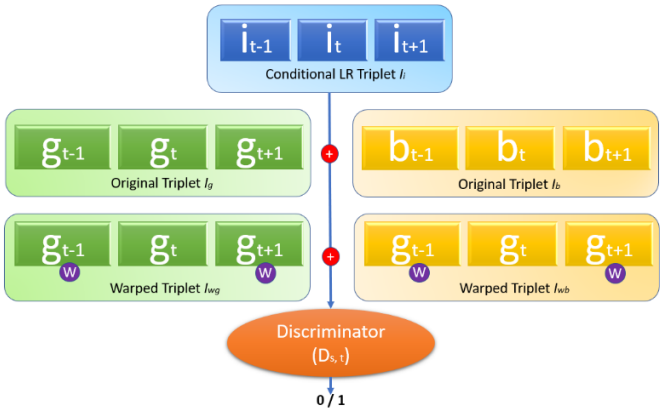


FIGURE 4: Conditional Discriminator for VSR

the quality of the dataset. Training large models can be very difficult as it not only requires a vast amount of data set but also enormous computation power. Transfer learning provides a good advantage in overcoming these challenges. It is the repeated usage of the pre-trained model for a new problem statement similar to the previous one [18]. The intention in transfer learning is to save the information that a model has acquired while solving a problem with enormous amounts of labeled data and use it to resolve another similar problem with very little data available to train. Rather than starting from scratch, it starts from a pattern having similarities with the sample solved beforehand. In this, we prevent the first few layers from training. The first few layers are common in general. The number of layers to be frozen depends upon the similarity of the problem. We only need to put in new layers for the unrecognized features. In case of large inaccuracy, the layers can be retrained from start. The application of transfer learning can be easily seen in computer vision, NLP as they require huge amounts of data to train on. Transfer Learning not only saves time as well as saves us from the complicated computations which require a heavy usage setup.

Copyright © 2021 by ASME

3.3 Metrics

3.3.1 Mean Squared Error (MSE)

Mean Squared Error commonly called MSE is one of the widely used loss functions used for calculating the mean squared error of each pixel between two comparative images. The MSE [19] between two images I (x, y) and I' (x, y) can be depicted as:

$$MSE = \frac{1}{MN} + \sum_{x=1}^{M-1} \sum_{y=1}^{N-1} \left[I'(x,y) - I(x,y) \right]^2 \tag{7}$$

A lower value of MSE means both images are quite identical and vice versa. For example, an MSE of 0 represents both images as identical. The only concern with MSE is that it is extremely dependent on image intensity scaling.

3.3.2 Peak Signal to Noise Ratio (PSNR)

Peak Signal-to-Noise Ratio is a comparison and benchmarking metric, calculated as the ratio of the highest potential power of a signal to the power of unwanted noise that reduces the accuracy of its representation. It is most widely used to assess the accuracy of lossy compression codec restoration. PSNR is measured in logarithmic values with decibel(dB) as a reference scale [20] and mostly uses MSE for its computation:

$$PSNR = 10\log_{10} \frac{\mathrm{MAX}_I^2}{MSE} = 20\log 10\,(\mathrm{MAX}_I) - 10\log 10(\mathrm{MSE}) \tag{8}$$

Here, the maximum possible pixel value of the image is given by MAX_I. Conventional PSNR values in highly compressed images and video compressions are from 30 and go upto 50 dB, with pixel value of 8 bits, with higher being stronger. Thus, a higher value signifies that the images are more detailed.

3.3.3 Spectral Angle Mapper (SAM)

A pixel vector in N-dimensional space of spectra has both magnitude and an angle which is measured from the respected coordinate axis. Spectral Angle Mapper (SAM) is a spectral classification that matches pixels to reference spectra using only the n-D angular component of the pixel vector so when comparing the two spectra the multidimensional vector is defined for both and the angle formed between them is calculated. Smaller angles indicate further similarity to the reference continuum. Both spectra are treated as matched if the angle is below a threshold value and unclassified if the angle is further away from maximum threshold.

3.3.4 Visual Information Fidelity (VIF)

The Visual Information Fidelity (VIF) index [21] is a comprehensive image caliber evaluation index that has grounds on natural scene statistics and the definition of image data derived from the human visual system. To measure the data exchanged among the trial and reference images, the VIF indicator engages natural scene statistical (NSS) models in combination with a distortion (channel) model. Furthermore, the VIF index is founded on the conjecture that this mutual knowledge is an allegiant feature that is closely in relation with the visual consistency. A higher VIF index is obtained when the images are almost similar. In comparison to antecedent approaches focusing on human visual system (HVS) responsiveness to error and composition analysis, this computational perspective employed in an informative theory context produces a full reference (FR) quality evaluation (QA) technique that is independent of any HVS or observing geometrical variable, nor any constants that need to be optimised, thus remaining competitive with cutting-edge QA methods.

3.3.5 Spatial Correlation Coefficient (SCC)

Spatial correlation is a statistical dependency of spatial direction between two signals. It means there is a relationship between the average signal gain obtained and the angle of arrival of a signal. Natural images have a high degree of spatial similarity but real-world situations, the correlation will be lower, but high odds will also be exceeded. This correlation is directly dependent on the eigenvalues of matrices which are in correlation. Higher the correlation coefficient higher the similarity in images. The eigenvectors demonstrate the spatial direction of the matrix. High spatial correlation represents that the eigenvalue spread is sizeable and vice versa.

4 DATASET

The dataset contains 100 high resolution (HR) videos and corresponding 100 4x degraded low resolution (LR) videos of forced isotropic turbulent flow fields. Each video has a duration of 8s and dimensions 64 x 64 pixels. Each high-resolution video was obtained by stacking 200 timestamps of forced isotropic flow fields obtained by plotting pressure for 64x64 points. Consecutive timestamps are separated by a time step of 0.02s. The HR videos are converted into LR by first downscaling and then upscaling the HR snapshots by 4 times using bicubic interpolation and then stacking these 200 snapshots to obtain the LR video. The HR 64x64 snapshots or timestamps are obtained from the 'isotropic 1024 coarse' dataset which is a part of Johns Hopkins Turbulence Databases (JHTDB). As stated in the description of the database, the data was obtained by direct numerical simulation (DNS) of forced isotropic turbulence on a 1024^3 periodic grid, with viscosity(ν) equal to 0.000185 and Taylor Reynolds number (R_λ) fluctuating around 433, by utilizing a pseudo-spectral parallel code.

Analyzing the time convergence of the viscous term was accomplished using an integrating factor. The nonlinear term is written in vorticity form, and the remaining terms are combined

Copyright © 2021 by ASME

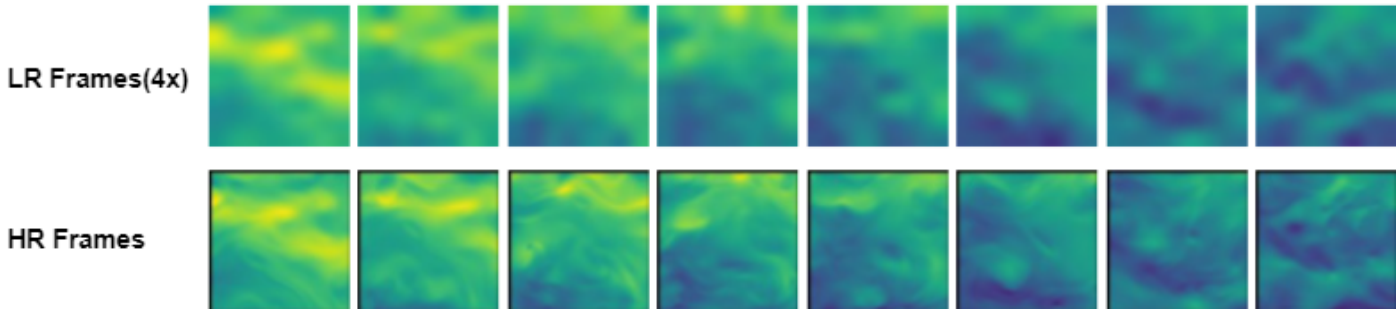


FIGURE 5: 4x degraded LR video frames (top) of forced isotropic turbulence and their corresponding HR video frames (bottom)

using a second-order Adams-Bashforth scheme. After the simulation reaches a statistically stationary state, 5028 frames of data, which consists of pressure and the 3 components of the velocity vector, are generated and inserted inside the database. The entire duration of the stored data is about five large-eddy turnover times. The isotropic coarse dataset thus possesses 5028 timesteps, for time t between 0 and 10.056 s. The dataset of 100 HR and 100 LR videos is segmented in the ratio of 80:10:10 for training, validation, and testing respectively.

5 RESULTS AND DISCUSSION

We performed exhaustive evaluations of different models on our dataset with different techniques of architecture handling and model training. As a preliminary test of the model's capabilities, we use the pre-trained weights to obtain results and use them as a benchmark. We then use transfer learning on these models to partially train and tweak them. Results obtained through the use of both these methods are discussed for the various models used in our study in the following section. To ensure uniformity in results and performance, we employ an adaptive data rate with an initial value of 0.01 which scales by 10%. We also implement the Adam optimizer in all models as it synergizes well with the adaptive learning rate and also works best in low memory/computing power scenarios.

5.1 ESPCN

We chose this model because of its ease of use and a comparatively simple architecture. Since it is based on Convolutional Neural Networks, albeit an advanced one, its architecture is straightforward and relies mostly on its Sub Pixel Convolution layer to do most of the work. Therefore the pre trained model which was trained on the CDVL database consisting of videos set in everyday scenarios performs a poor job at upscaling our turbulence images. The results generated although bring back the pressure differences, aren't very clear in defining their boundaries and the depth.

We then froze the first two layers of the model and trained the final Sub Pixel layer on our dataset for 10 epochs, each with a batch size of 80. This rapid training produced highly detailed images as compared to the ones produced by the pre trained model. After some fine tuning, the produced images could bring back a some lost details like the enhanced colour variance, but it still couldn't produce a concrete result as the output images lacked detail in areas with rapid change in pressure. We also encountered random artifacts as the model hit overfitting in less than 15 epochs which inversely affected the image quality and the model statistics.

5.2 ESRGAN

ESRGAN produces commendable results particularly with respect to our dataset. It utilizes its Generative Adversarial Net-

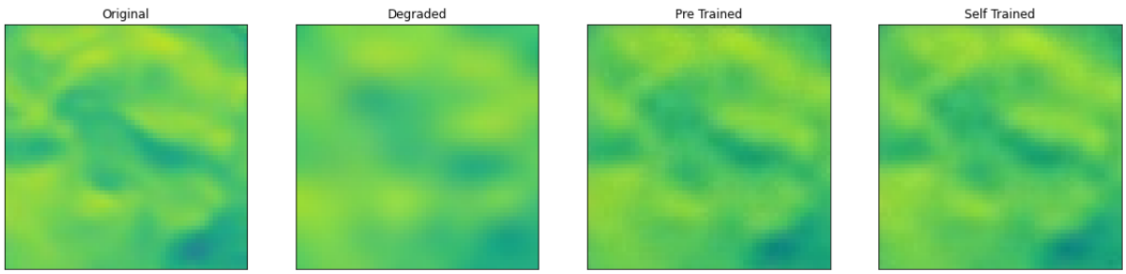


FIGURE 6: ESPCN results

Copyright © 2021 by ASME

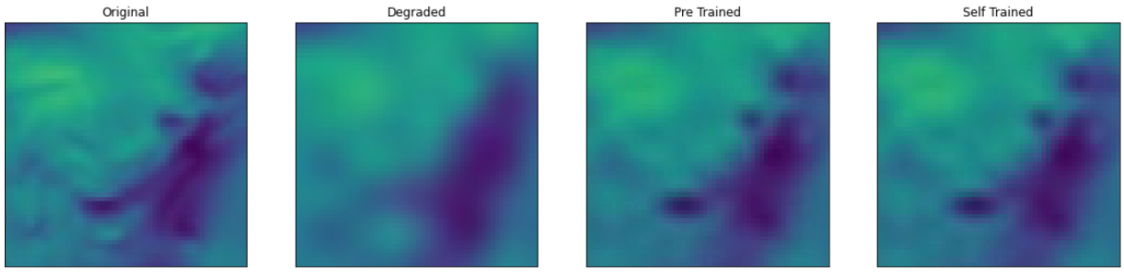


FIGURE 7: ESRGAN results

work to upscale images to satisfaction, but shows minimal improvement when it is partially trained. It is able to differentiate between high and low pressure regions and upscales them as whole blocks but the model compensates for it by averaging the neighboring pixels in areas of minimal change defeats the purpose of the model.

To fully harness the capability of this model, we trained different layers partially and found that freezing all the layers except for the last layer of the generator and the whole discriminator produces the best results in an efficient manner. Since the generator is constantly being updated with counter results from the discriminator, we train the discriminator to classify our dataset. But this resulted in minimal to average improvement depending on different images.

This is visible in the results presented in Fig 7, where both the pre trained model and the self trained model are able to identify areas of differing pressure, highlight it and upscale it but the pre trained model leaves out the minute intricacies such as the waves and curves. In comparison, the self trained model does all that the pre trained model does but also account for the small scale changes like the curves and waves that the pre trained model left out.

The other caveat of this model is that it is trained to super resolve images and not videos. Therefore, it does not account for temporal coherence and instead trades it for enhancing the quality of a single image, one at a time. Thus, even though ESRGAN scores well in the single image criteria, it loses out in preserving the flow.

5.3 TecoGAN

Of all the different models we examined and shortlisted, TecoGan performs extraordinarily well even with limited amount of resources and computational power. Since a heavy emphasis is laid on temporal coherency, TecoGAN learns from the accompanying frames of the dataset to learn the changes in pressure with time, therefore it not only enhances the low resolution data but also preserves the 'flow' of the video which other models don't.

It compensates for the data loss in low pressure regions or the 'dips' by examining the near future frames to produce those details while upscaling LR images. The pre trained model of TecoGAN was really well trained as it could produce above average results out of the box as compared to the results from other pre trained models. Since the pre trained model was already capable enough, we only had to fine tune the generator, thus we focused on training the discriminator similar to ESRGAN.

The self trained model produced high resolution images which were almost as good as the ground truth and were very similar to their corresponding original images. This is evident by the value of MSE for TecoGAN's self trained version. The pre trained model was able to reconstruct the features but they were still not properly defined but the self trained model was able to exactly replicate the features right to the minute changes while preserving the boundary. The boundary problem is of prime importance as it may highly distort/skew the pressure field. This added advantage of being able to exactly replicate the image and its features complements TecoGAN's flow preservation very well and produces a highly synergic model.

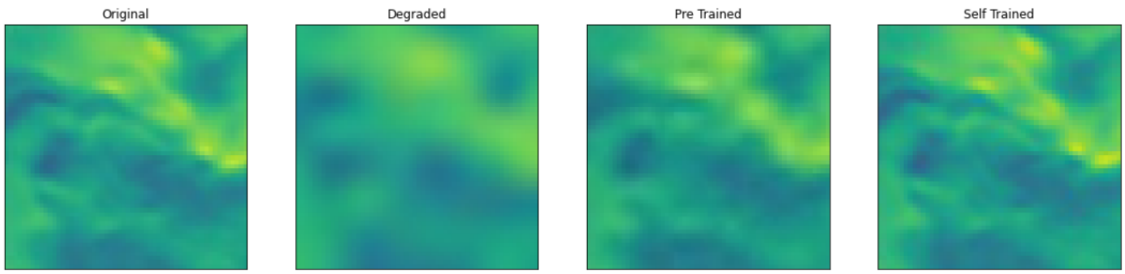


FIGURE 8: TecoGAN results

Copyright © 2021 by ASME

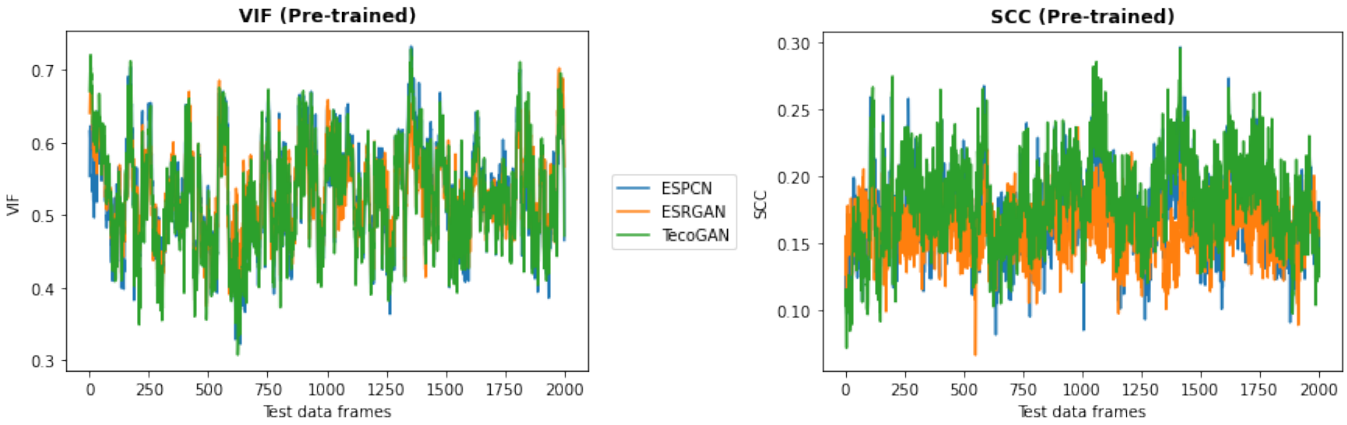


FIGURE 9: Frame to frame statistics of VIF and SCC metrics of results generated by pre trained models

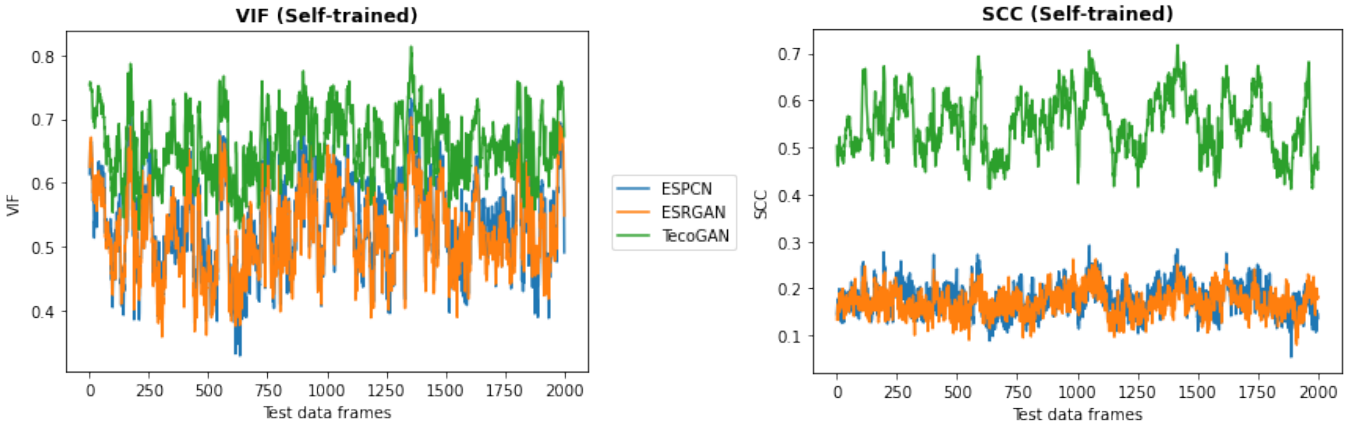


FIGURE 10: Frame to frame statistics of VIF and SCC metrics of results generated by self trained models

Metrics	Degraded	ESPCN		ESRGAN		TecoGAN	
		Pre-Trained	Self-Trained	Pre-Trained	Self-Trained	Pre-Trained	Self-Trained
MSE	30.9911	26.8544	21.3932	16.1058	14.0273	21.8246	5.5169
PSNR	34.3949	34.5287	34.8281	36.4730	37.0000	35.0206	41.2876
SAM	0.0598	0.0443	0.0424	0.0380	0.0365	0.0387	0.0277
VIF	0.3593	0.5272	0.5321	0.5275	0.5283	0.5289	0.6610
SCC	0.0761	0.1745	0.1790	0.1619	0.1714	0.1833	0.5506

TABLE 1: Median metrics of 10 videos (test set) of the dataset obtained from the pre and self trained models of ESPCN, ESRGAN and TecoGAN

Copyright © 2021 by ASME

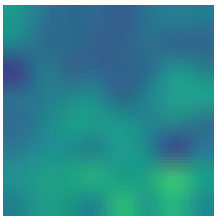
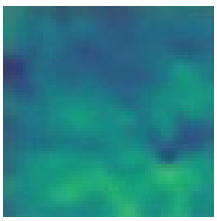
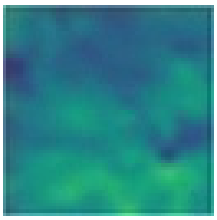
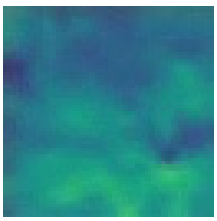
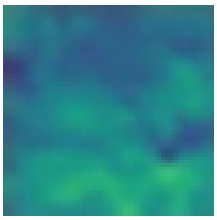
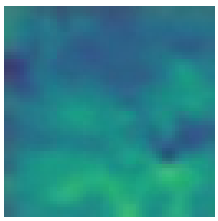
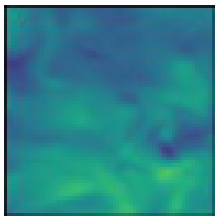
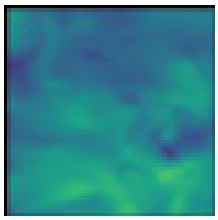
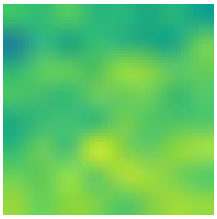
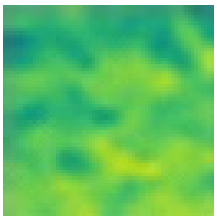
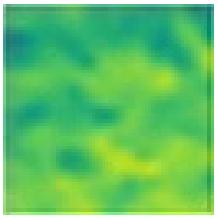
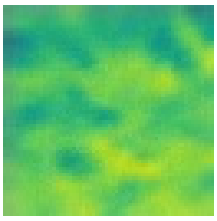
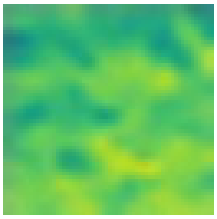
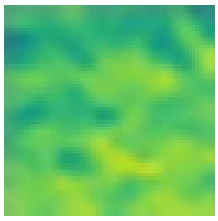
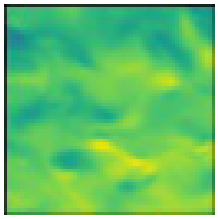
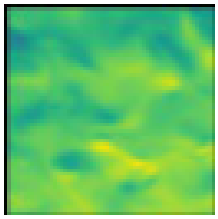
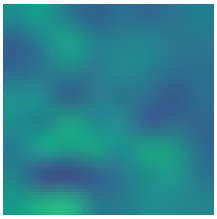
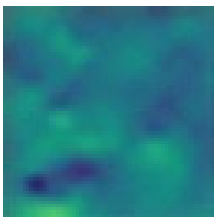
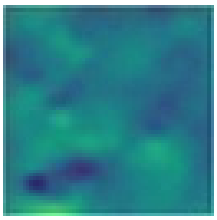
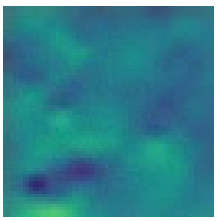
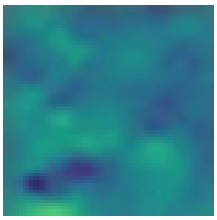
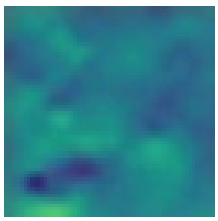
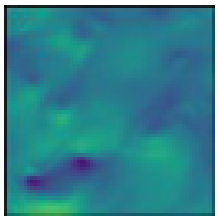
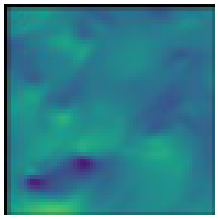
Bicubic 4x	ESPCN	ESRGAN	TecoGAN	Ground Truth
	 	 	 	
	 	 	 	
	 	 	 	

TABLE 2: Visual comparison of some Super-Resolution images generated from different models. The upper sub-row of images are generated from pre trained models while the lower sub-row of images are generated from self trained models

Copyright © 2021 by ASME

6 CONCLUSION

Our study was able to conclude that in the low resource settings, transfer learning can be used for super-resolution which is a computationally expensive task and it can be used to recover high-res turbulent flow fields from low-res data, by training only a select number of layers and fine tuning models according to our datatset. TecoGAN produced the best results in terms of VIF for pretrained weights as well as when the model was trained specifically for forced isotropic turbulent flow fields. Thus transfer learning incorporated super resolution can enhance the quality of low-res simulations while making a slight trade-off in the simulation quality for a great reduction in the required computational time and resources for processing. All layers of the model can be further trained to achieve better results in high resource settings.

ACKNOWLEDGMENT

The authors would like to thank the Fluid Systems Laboratory of Delhi Technological University for providing computing facilities. We are thankful to our professor Dr. Raj Kumar Singh who provided expertise which greatly assisted us in the research and improved the paper significantly. We would also like to thank other members of the Fluid Mechanics Group for directly or indirectly helping during the research work.

REFERENCES

[1] Sekar, V., Jiang, Q., Shu, C., and Khoo, B. C., 2019. "Fast flow field prediction over airfoils using deep learning approach". *Physics of Fluids,* ***31***(5), May, p. 057103.

[2] San, O., and Maulik, R., 2018. "Extreme learning machine for reduced order modeling of turbulent geophysical flows". *Phys. Rev. E,* ***97***, Apr, p. 042322.

[3] Hasegawa, K., Fukami, K., Murata, T., and Fukagata, K., 2020. "Machine-learning-based reduced-order modeling for unsteady flows around bluff bodies of various shapes". *Theoretical and Computational Fluid Dynamics*, May.

[4] Fukami, K., Fukagata, K., and Taira, K., 2019. "Super-resolution reconstruction of turbulent flows with machine learning". *Journal of Fluid Mechanics,* ***870***, July.

[5] Fukami, K., Fukagata, K., and Taira, K., 2021. "Machine-learning-based spatio-temporal super resolution reconstruction of turbulent flows". *Journal of Fluid Mechanics,* ***909***, Feb., p. A9.

[6] Deng, Z., He, C., Liu, Y., and Kim, K. C., 2019. "Super-resolution reconstruction of turbulent velocity fields using a generative adversarial network-based artificial intelligence framework". *Physics of Fluids,* ***31***(12), Dec., p. 125111.

[7] Liu, B., Tang, J., Huang, H., and Lu, X.-Y., 2020. "Deep learning methods for super-resolution reconstruction of turbulent flows". *Physics of Fluids,* ***32***(2), Feb., p. 025105.

[8] Gao, H., Sun, L., and Wang, J.-X., 2020. "Super-resolution and denoising of fluid flow using physics-informed convolutional neural networks without high-resolution labels". *arXiv e-prints*, Nov., p. arXiv:2011.02364.

[9] Stengel, K., Glaws, A., Hettinger, D., and King, R. N., 2020. "Adversarial super-resolution of climatological wind and solar data". *Proceedings of the National Academy of Science,* ***117***(29), July, pp. 16805–16815.

[10] Gupta, P., Tyagi, P., and Singh, R. K., EasyChair, 2021. Analysis of generative adversarial networks for data driven inverse airfoil design. EasyChair Preprint no. 5975.

[11] Soh, J. W., Cho, S., and Cho, N. I., 2020. "Meta-Transfer Learning for Zero-Shot Super-Resolution". *arXiv e-prints*, Feb., p. arXiv:2002.12213.

[12] Zhang, Z., Wang, Z., Lin, Z., and Qi, H., 2019. "Image Super-Resolution by Neural Texture Transfer". *arXiv e-prints*, Mar., p. arXiv:1903.00834.

[13] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., and Wang, Z., 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network.

[14] Ledig, C., Theis, L., Huszar, F., Caballero, J., Aitken, A. P., Tejani, A., Totz, J., Wang, Z., and Shi, W., 2016. "Photo-realistic single image super-resolution using a generative adversarial network". *CoRR,* ***abs/1609.04802***.

[15] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Loy, C. C., Qiao, Y., and Tang, X., 2018. "ESRGAN: enhanced super-resolution generative adversarial networks". *CoRR,* ***abs/1809.00219***.

[16] Chu, M., Xie, Y., Mayer, J., Leal-Taixé, L., and Thuerey, N., 2020. "Learning temporal coherence via self-supervision for gan-based video generation". *ACM Transactions on Graphics,* ***39***(4), Jul.

[17] Chu, M., Xie, Y., Leal-Taixé, L., and Thuerey, N., 2018. "Temporally coherent gans for video super-resolution (tecogan)". *CoRR,* ***abs/1811.09393***.

[18] Olivas, E. S., Guerrero, J. D. M., Sober, M. M., Benedito, J. R. M., and Lopez, A. J. S., 2009. *Handbook Of Research On Machine Learning Applications and Trends: Algorithms, Methods and Techniques - 2 Volumes*. Information Science Reference - Imprint of: IGI Publishing.

[19] Wang, Z., and Bovik, A., 2002. "A universal image quality index". *IEEE Signal Processing Letters,* ***9***(3), pp. 81–84.

[20] Huynh-Thu, Q., and Ghanbari, M., 2008. "Scope of validity of psnr in image/video quality assessment". *Electronics Letters,* ***44***, pp. 800–801.

[21] Sheikh, H., Bovik, A., and de Veciana, G., 2005. "An information fidelity criterion for image quality assessment using natural scene statistics". *IEEE Transactions on Image Processing,* ***14***(12), pp. 2117–2128.

Copyright © 2021 by ASME

2021 IEEE 12th Energy Conversion Congress & Exposition - Asia (ECCE-Asia) | 978-1-7281-6344-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/ECCE-Asia49820.2021.9479170

A High Gain, High-Efficiency Dual-Band Cascode Class-F PA for Ka-Band Application

Priya Puja
Department of electronics and communications
Delhi Technological University
Delhi, India
pujapriya1802@gmail.com

Kaur Gurjit
Department of electronics and communications
Delhi Technological University
Delhi, India
gurjitkaur@du.ac.in

Kumar Rajesh
Department of electronics and communications
Indian Institute of Technology
Dhanbad, India
rajeshict42@gmail.com

***Abstract*— this paper presents a high gain and high-efficiency dual band Cascode class-F power amplifier (PA) for Ka- band applications using 0.25µm GaN technology. Cascode Class-F PA is combined with a non-uniform step impedance transmission line and magic tee (MTEE) to illustrate a second harmonic balance matching technique. A second harmonic balance matching technique utilized to enhance the gain up to 30dB at the saturation power region. This technique also enhances the resonant circuits band properties. Magic tee helps to enhance an internal matching design. For maximum output power, drain efficiency, and PAE the fundamental and second harmonic impedances are to be accurately improved for the particular operating frequency. Measured results represent maximum output power of 45.1dBm, drain efficiency of 87% and power added efficiency (PAE) of 73.4%. The fundamental and second harmonic balance levels are adjusted at -21dBc to -16dBc. The simulated and the measured results of the proposed power amplifier show the right constancy.**

Keywords—Class-F, Harmonic Balance, Power added efficiency (PAE), Magic Tee (MTEE), Cascode.

I. INTRODUCTION

Recent trends shows that now various communication service providers are looking at Ka-band communication infrastructures for further technology enhancement in the present base station networks infrastructure. The Ka-band is becoming the most promising band for more data rate than any other lower frequencies band. Hence, Ka-band can play a very crucial role in enlarging future telecommunication technologies, like satellite communication or IP traffic. Fig. 1 represents various Ka-band applications, where cellular back haul has emerging as the major issue in coming 5G technologies base stations. The power amplifier (PA) is an essential block of transceiver system, which is used to enhance the strength of electrical signal in a wireless transmitter system. The GaN technology comes with many advantages such as higher battery life and low power consumption at a low cost for portable devices.

Gallium nitride transistor (GAaN HEMT) provides high electron mobility and operates at microwave frequencies. A GaN PA with PAE of 82% along with a harmonic tuned network (HTN), have the ability to manage the first four harmonics [1][2]. The main aim of the research is to improve efficiency and linearity while increasing the bandwidth [3][4][5]. At dual band mode, class F PA have reached to efficiency of 58% and output power of 51%[6][7]. Transformer based power combiner method is used to increase the output power of 24.8dBm and stacked topology were used to enhance efficiency of 24.3%[8]. Several design techniques for ka-band wide bandwidth and low chip area is discussed for high power amplifiers[9]. As the demand for class –F PA increase, various methods are shown using an infinite number of odd or even harmonics [10]. A frequency tunable load with a two stage common-emitter (CE) power amplifier has accomplished PAE to 55.9% [11]. However, at 21.5GHz frequency, 3-stage high power amplifiers (HPA) have achieved output power of 39.5dBm and PAE of 35% [12].

This paper proposes a fully integrated dual-band Class-F PA for Ka-band applications. A proposed second harmonic balance technique will provide high gain and also improving the band properties of the PA circuit. This paper is arranged as follows: Section II includes design and consideration of PA while Section III would be to focus on the result and discussion, and finally, the conclusion is followed in Section IV.

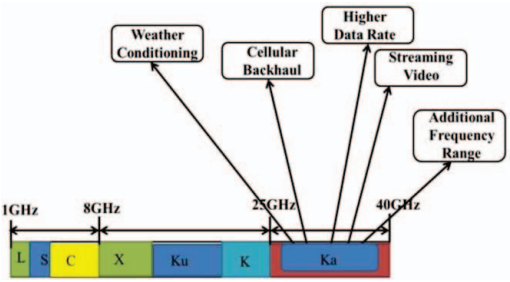


Fig. 1: Various Band configurations with Ka-band applications

II. DESIGN AND CONSIDERATION

In this segment, the design and consideration of the proposed Cascode Class-F PA will be discussed. The design issue of Class-F PA is considered in subsection *A* while designing a dual-band Class-F PA in sub-section *B*.

A. Design Issues of the Class-F PA

The main components of every PA circuit design are input and output matching networks. These networks are designed using inductors and capacitors. Efficiency of power amplifier design is affected by the capacitive and inductive losses of capacitors and inductors. Impedance adjustment is one of the major challenges for PA design at higher frequencies for every harmonic matching network. The network should be terminated as per the order of harmonics. The maximum number of harmonics that could be optimized in a circuit is till 4th order. The maximum number of harmonics presents in a class- F circuit design comes up with a trade-off between linearity and efficiency, which reduce the power utilization. Even so, the maximum number of harmonics in class-F design is limited to three as higher

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:22:18 UTC from IEEE Xplore. Restrictions apply.

harmonics shows lower effect on waveform shaping. It was observed that the respective values of absolute efficiency of the class-F PA is reduced 90.7%. Lumped elements which are used to design input and output matching network also induce some limitations at high frequency design of PA. To meet the respective value of impedance at 2nd harmonic are analyzed using transmission line and magic tee in input-output networks. For PA design magic tee is coming as a new research area and it is enhancing the interest for more improvement in efficiency and output power in transmitting and receiving of signals without any limitations. The length of transmission line is analyzed for every harmonics in matching networks so that it creates no issues on-chip

B. Class-F PA with harmonic balance network

An ideal class-F PA used a high-quality (Q) harmonic control network, which can only match harmonics till third order. This ideal harmonic control network has limited bandwidth and lower capability to handle high order harmonics. So, harmonic control network have high influence on efficiency and output power of Class- F PAs. An enhanced circuit of harmonic balance PA design is shown in Fig.2, where idea of non-uniform transmission line and magic tee is shown. Also, detailed input and output harmonic balance matching network is presented. The non-uniform step impedance transmission line MSL1 to MSL4 are connected with a Magic TEE1 in series configuration. An input matching network is used to short the second harmonic impedance for even harmonics while an open circuit for odd harmonics. Likewise, the non-uniform step impedance transmission lines from MSL6 to MSL10 are connected to Magic TEE2 with an output matching network. An output matching network is used to short the second harmonic impedance for even harmonic while open circuit for odd harmonics. A non-uniform step impedance short circuited transmission line helps in enhancing resonant circuits for band effects. The Magic TEE helps in enhancing the internal matching network. Designing and implementation of non-uniform step transmission line and magic Tee (MTEE) is proving very beneficial for a high efficiency and high output power for a PA circuit design. For Cascode class-F PA topology the load impedance are arrange at fundamental and 2nd harmonic frequency band i.e. f_o and $2f_o$. High efficiency of the PA demands upon the 2nd harmonic matching network. The harmonic balance network can control till 2nd order input impedance at 40GHz. The input impedance Zin_2 can be expressed as:

$$Z_{in2} = jZ_8 \frac{Z_2 \tan(6n/n_1) + Z_4 \tan(6n/n_2)}{Z_4 - Z_2 \tan(6n/n_1)\tan(6n/n_1)} \quad (1)$$

Where, $Z_2$and Z_4 is the characteristic impedance of MSL2 and MSL4 respectively. MSL1 and MSL4 are the stepped impedance transmission line, which helps in reducing reflection created by impedance mismatch. The impedance value of the 2nd harmonic network is arranged correctly at a high impedance region over a whole frequency range.

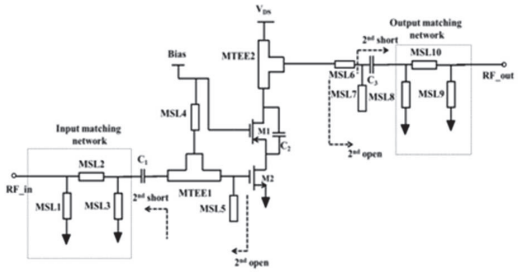


Fig. 2: Proposed dual-band Class-F PA with harmonic balance network

III. RESULTS AND DISCUSSION

For the proposed Cascode Class-F PA the gate voltage and drain bias voltage are place at -3.1V and 28V respectively. Simulation of class-F PA is done in the Advanced Design System (ADS) using S-parameters and harmonic balanced simulator. Fig. 3 depicts the S-parameters of Class-F reverse isolation (S11) and forward gain (S21) whose values are -35.4dB and 29.8dB respectively. The dual band respond is depicts at 27.8GHz and 36GHz at resonant frequency. Fig. 4 represents the input impedance at the fundamental frequency and 2nd harmonic frequency. Fig. 5 represent the continuous wave (CW) simulation results i.e. PAE, gain, and drain efficiency versus output power of the proposed Cascode PA.

The output results shows that at the peak power of 45.1 dBm output power the drain efficiency and PAE is 89.4% and 73.4% respectively. Input impedances are matched to get high gain. The output at fundamental and 2nd harmonic impedances is matched to get required impedances for large saturated drain efficiency.

Fig. 6 represents the adjacent channel leakage ratio (ACLR) versus input power of the proposed Cascode-F PA design, where results show ACLR of 16dBc and 21dBc at 25GHz and 40GHz. It has realized -40dB performance, which are summarized and differentiated to previously reported work with frequency bands and output power levels state-of-the-arts as shown in Table. 1.

TABLE1 PERFORMANCE COMPARISON OF PA WITH OTHER REPORTED ONES

References	*f* (GHz)	P_{out} (dBm)	DE (%)	PAE (%)	Gain (dB)	ACLR (dBc)
[9]	24-35	11.4	-	55.6	16	-
[10]	24-35	39.5 dB	-	35	24.3	-25
This work	25-40	45	89.4	73.4	29.4	16-21

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:22:18 UTC from IEEE Xplore. Restrictions apply.

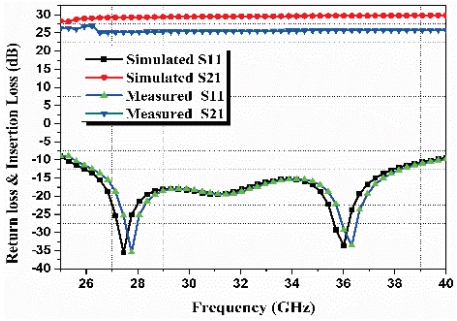


Fig. 3. Measured and simulated S-parameters

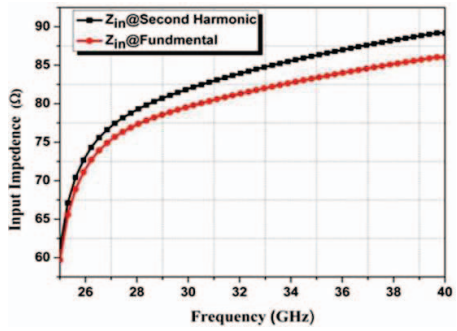


Fig. 4. Input impedance versus frequency

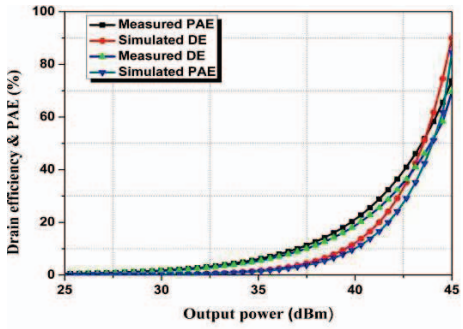


Fig. 5. Simulated and measured characteristics drain efficiency, PAE, and gain

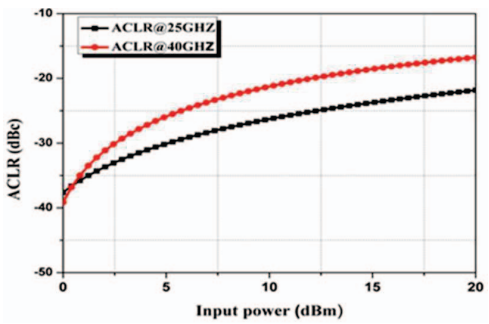


Fig. 6. ACLR variation versus input power

IV. CONCLUSION

This work concluded that a Cascode dual band class-F PA with harmonic balance network for Ka band applications. Fundamental and the 2nd harmonic frequency are analyzed with simulated and measured results for a Cascode class-F PA design. A Cascode topology helped in improving gain stability and enhances efficiency. The simulated and measure efficiency and output power of Cascode class-F PA is considered to be constant about 87% efficiency at output power 45.1dBm.

REFERENCES

[1] V. Carrubba *et al.*, "On the extension of the continuous class-F mode power amplifier," *IEEE Trans. Microw. Theory Tech.*, vol. 59, no. 5, pp. 1294–1303, 2011, doi: 10.1109/TMTT.2011.2117435.

[2] F. H. Raab, "Class-F Power Amplifiers with Maximally Flat Waveforms," *IEEE Trans. Microw. Theory Tech.*, vol. 45, no. 11, pp. 2007–2012, 1997, doi: 10.1109/22.644215.

[3] J. H. Kim, S. J. Lee, B. H. Park, S. H. Jang, J. H. Jung, and C. S. Park, "Analysis of high-efficiency power amplifier using second harmonic manipulation: Inverse class-F/J amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 59, no. 8, pp. 2024–2036, 2011, doi: 10.1109/TMTT.2011.2157354.

[4] G. Nikandish, E. Babakrpur, and A. Medi, "A harmonic termination technique for single-and multi-band high-efficiency class-F MMIC power amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 62, no. 5, pp. 1212–1220, 2014, doi: 10.1109/TMTT.2014.2315591.

[5] M. Thian, A. Barakat, and V. Fusco, "High-Efficiency Harmonic-Peaking Class-EF Maximum Operating Frequency," *IEEE Trans. Microw. Theory Tech.*, vol. 63, no. February, pp. 659–671, 2015.

[6] S. N. Ali, P. Agarwal, S. Gopal, S. Mirabbasi, and D. Heo, "A 25-35 GHz neutralized continuous class-F CMOS power amplifier for 5g mobile communications achieving 26% modulation PAE at 1.5 Gb/s and 46.4% peak PAE," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 66, no. 2, pp. 834–847, 2019, doi: 10.1109/TCSI.2018.2860019.

[7] S. N. Ali, P. Agarwal, J. Baylon, and D. Heo, "Reconfigurable high efficiency power amplifier with tunable coupling coefficient based transformer for 5G applications," *IEEE MTT-S Int. Microw. Symp. Dig.*, no. 1, pp. 1177–1180, 2017, doi: 10.1109/MWSYM.2017.8058811.

[8] Y. Chang, B. Z. Lu, Y. Wang, and H. Wang, "A Ka-Band Stacked Power Amplifier with 24.8-dBm Output Power and 24.3% PAE in 65-nm CMOS Technology," *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 2019-June, pp. 316–319, 2019, doi: 10.1109/mwsym.2019.8700937.

[9] K. Q. H. Amplifiers, "On Design of Wideband Compact-Size," pp. 1–12, 2016.

[10] S. Shakib, H. C. Park, J. Dunworth, V. Aparin, and K. Entesari, "A Highly Efficient and Linear Power Amplifier for 28-GHz 5G Phased Array Radios in 28-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 51, no. 12, pp. 3020–3036, 2016, doi: 10.1109/JSSC.2016.2606584.

[11] K. Ma, T. B. Kumar, and K. S. Yeo, "A Reconfigurable K-/Ka-Band Power Amplifier With High PAE in 0.18-<formula formulatype="inline"><tex Notation="TeX">μ</tex></formula>m SiGe BiCMOS for Multi-Band Applications," *IEEE Trans. Microw. Theory Tech.*, vol. 63, no. 12, pp. 4395–4405, 2015, doi: 10.1109/tmtt.2015.2495129.

[12] Y. S. Noh and I. B. Yom, "A Linear GaN High Power Amplifier MMIC for Ka-Band Satellite Communications," *IEEE Microw. Wirel. Components Lett.*, vol. 26, no. 8, pp. 619–621, 2016, doi: 10.1109/LMWC.2016.2585553.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:22:18 UTC from IEEE Xplore. Restrictions apply.

2021 2nd International Conference for Emerging Technology (INCET)
Belgaum, India. May 21-23, 2021

A Novel 8.5 GHz Parallel Coupled, Half Wavelength Bandpass Filter with Dumbbell-Shaped Defected Ground Structure

Aditya Agrawal*
Department of Electronics and Communications
Delhi Technological University
New Delhi, India
aditya.aggie1999@gmail.com

Adya Agarwal*
Department of Electronics and Communications
Delhi Technological University
New Delhi, India
adyaagarwal4@gmail.com

Ojas Aryam*
Department of Electronics and Communications
Delhi Technological University
New Delhi, India
ojasaryam@gmail.com

Asok De*
Department of Electronics and Communications
Delhi Technological University
New Delhi, India
asok.de@dtu.ac.in
**All authors have contributed equally*

2021 2nd International Conference for Emerging Technology (INCET) | 978-1-7281-7029-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/INCET51464.2021.9456441

***Abstract*—A novel 8.5 GHz parallel coupled bandpass filter is presented with dumbbell shaped defected ground structure which is placed in the ground plane. The circuit structure is simulated using ADS. The microstrip filter structure is analyzed by gradient variation of the line that passes through the midpoints of the conducting strips, which is varied linearly with different values of slope. The effects on the bandwidth and cut-off frequency are observed to understand the desired gradient for different uses of the linear bandpass filter. The measured and analyzed results are in good agreement with the theoretical results. The proposed structure has compact size, good bandwidth, and high cut off frequency.**

Keywords— Bandpass Filter, Parallel-Coupled, Microstrip, Defected Ground Structure (DGS)

I. INTRODUCTION

With the shift to 5G networks, it has made high frequency filters a necessity. Recently the analysis of microstrips filters with defected ground structure is popular because of its efficient design and compact size. DGS in some of the frequencies can provide a rejection band because there is an increase in the inductance for the microstrip line. In this paper a bandpass microstrip filter is proposed with a fractional bandwidth of 0.15 and cut off frequency of 8.5 GHz. In this paper various configurations are applied for the design by changing the gradient of the slope that cuts through centre points of the parallel coupled filters. This helps in obtaining appropriate gradient for the filter design according to different usability with minimal changes to the filter design.

A. Microstrip Filter

It is a conducting strip with width and length which is placed on top of a substrate with some dielectric constant ε_r, width w and thickness t. In microstrip lines, the transverse components are mostly considered because the longitudinal components are small and therefore, negligible. Microstrips lines behave similar to the transverse transmission lines and therefore, quasi-TEM approximations can be applied to the microstrip lines.

The dielectric substrate which is homogeneous replaces the air which has inhomogeneous dielectric. There are different kinds of losses in microstrips such as radiation losses, conductor losses and dielectric losses. However, the magnetic losses play a role for substrates which are magnetic such as ferrites.

The metallic enclosures are required for most microstrip circuits. The conducting side and top walls are present which affects the effective dielectric constant and the characteristic impedances. These are some of the few effects that must be considered while designing microstrip filters.

B. Microstrip Discontinuities

There are many discontinuities that are encountered in microstrips such as bends, steps, gaps, open-ends and junctions. The effect of these discontinuities should be taken into consideration and modelled accurately. Microstrips with right-angle bends can be made using T-network which is equivalent to capacitance and inductance. Gaps in microstrips are shown using equivalent series (C_g) and shunt capacitance (C_p). For open ends, the field does not abruptly stop but is slightly extended further because of the effect of the fringing field. This is modelled with a shunt capacitance (C_p) or can be modelled with a transmission line of equivalent length l.

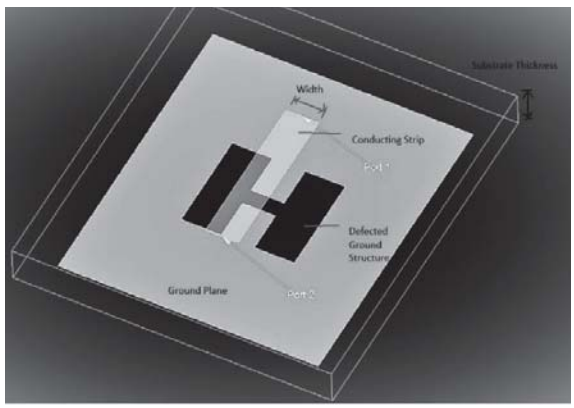


Fig. 1. General Microstrip Structure with Defected Ground Structure

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:27:37 UTC from IEEE Xplore. Restrictions apply.

C. *Parallel-Coupled, Half Wavelength Resonator Filters*

Microstrip bandpass filter with parallel coupled conducting strips has been presented in this paper. Adjacent resonators are placed such that they are along half of the length and parallel to each other. This kind of filter is convenient for building filters with a wide bandwidth as they give large coupling for a space between the resonators [1]. In [2], the suppression of 1st spurious harmonic in parallel-coupled bandpass filter with DGS has been studied.

II. Equivalent Circuit and Layout of Dumbbell DGS

Various close-packed and high-efficiency components were described [3] using structures called the defected ground structures where a defect is engraved into the ground plane of the microstrip. There are many applications of DGS in microstrip filters with various configurations [4-11]. Dumb-bell DGS is formed with 2 rectangle-shaped defects that are positioned within the ground plane and connected through a perpendicular channel that is carved in the ground. The DGS was placed according to the position and structure of our parallel coupled bandpass filter so that it doesn't overlap with the adjacent resonator which can cause a failure in the operation of the filter.

The layout of the Defected Ground Structure used is shown below in Fig. 2. There is a microstrip line above and a dumbbell shaped pattern that is carved into the ground plane.

In Fig. 2, the width of the microstrip line is 0.57 mm (w) and the characteristic impedance is 50 Ω (Z_0). The substrate used for the simulation is Rogers RT/Duroid 6010. It has substrate height h=0.635 mm, a dielectric loss tangent of 0.0023, and dielectric constant ε_r=10.2. The microstrip line is made of copper.

The dimensions of the dumbbell DGS that is used for reduction of size of the microstrip bandpass filter are a = 1.69 mm, b = 0.69 mm, and g = 0.285 mm. Simulation is done using Keysight ADS software and the simulation results are illustrated in Fig. 3.

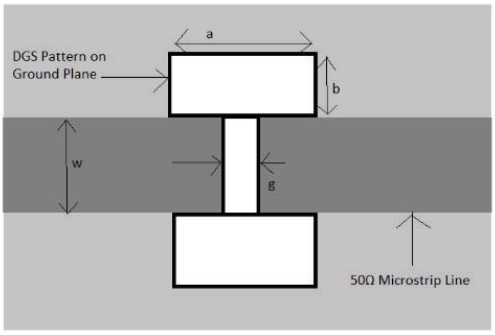


Fig. 2. Configuration of the dumbbell-shaped DGS unit.

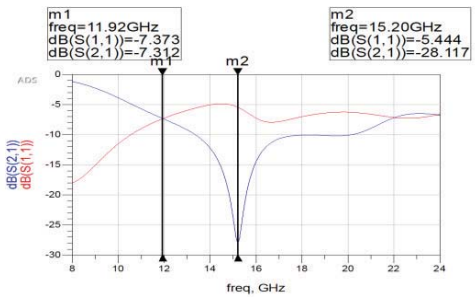


Fig. 3. Simulation result for dumbbell-shaped DGS with Microstrip Line.

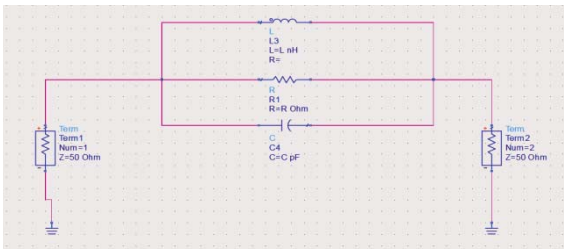


Fig. 4. Equivalent Circuit of dumbbell shaped DGS

A single parallel RLC resonator circuit (shown in Fig. 4) can be used to model the frequency characteristics of the proposed dumbbell DGS structure, where L signifies the effect of rectangle shape in the DGS unit and C signifies the effect of the channel between two rectangular shapes [8]. Changes in equivalent-circuit parameters due to variation of physical dimensions of the DGS are elaborated further in [12,13].

By using the modelling techniques in [6], the parameters of the equivalent circuit are found using (1), (2) & (3).

$$C = \frac{\omega_c}{2Z_0(\omega_0^2 - \omega_c^2)} \tag{1}$$

$$L = \frac{1}{4\pi^2 f_0^2 C} \tag{2}$$

$$R = \frac{2Z_0}{\sqrt{\frac{1}{|S_{11}(\omega_0)|^2} - \left(2Z_0\left(\omega_0 C - \frac{1}{\omega_0 L}\right)\right)^2} - 1} \tag{3}$$

Here Z_0 is the characteristic impedance of the microstrip line, f_c is the 3dB cut-off frequency, and f_0 is the resonant frequency. Values of cut-off frequency and resonant frequencies extracted from Fig. 3 are shown in Table I and the values of the parameters of the equivalent circuit, calculated using (1), (2) & (3) are also shown in Table I.

The simulation result of the equivalent circuit of dumbbell DGS (Fig. 4) is presented in Fig. 5. This result is similar to the simulation result obtained with DGS on microstrip line (Fig. 3).

TABLE I. Extracted Equivalent Circuit Parameters for Dumbbell DGS Unit

Cut-off Frequency, ω_c (GHz)	11.92
Resonant Frequency, ω_0 (GHz)	15.20
$\lvert S_{11}(\omega_0)\rvert$	0.534
L(nH)	0.514
C(pF)	0.213
R(Ω)	360

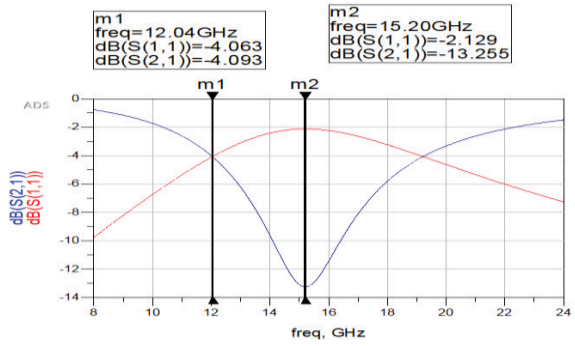


Fig. 5. Simulation Result of the equivalent RLC circuit of dumbbell shaped DGS

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:27:37 UTC from IEEE Xplore. Restrictions apply.

III. Conventional Parallel-Coupled Line Bandpass Microstrip Filter

This section is devoted to explaining the designing procedure of the parallel-coupled microstrip bandpass filter. The filter is designed with the centre frequency (f_0) chosen as 8.5 GHz and the fractional bandwidth (FBW) around f_0 as 0.15, which gives cut-off frequencies f_1=7.88 GHz and f_2=9.16 GHz for the bandpass. FBW is defined using (4).

$$FBW = \frac{\omega_2 - \omega_1}{\omega_0} \tag{4}$$

Filters can be designed using a variety of materials and by various methods [14]. The filter is chosen to be a five-pole Chebyshev Response bandpass filter with a 0.1-dB ripple in the passband. The characteristic impedance, Z_0=50 Ω. The substrate on which the filter is implemented is Rogers RT/Duroid 6010. It has substrate height h=0.635 mm, a dielectric loss tangent of 0.0023, and dielectric constant ε_r=10.2.

Element values for the low-pass prototype filter (Fig. 6) of a 5-pole Chebyshev filter with 0.1-dB ripple are g_0=g_6=1.0, g_1=g_5=1.1468, g_2=g_4=1.3712, g_3=1.975 [15].

Using these g-values, the low-pass prototype filter is converted to a band-pass filter. The 'g' elements transform to a series/parallel LC resonant circuit for given low-pass prototype circuit. For 'g' representing inductance, the elements for series LC resonator in the bandpass filter are found using (5).

$$L_s = \left(\frac{\Omega_c}{FBW\omega_0}\right)\gamma_0 g \tag{5a}$$

$$C_s = \frac{1}{(\omega_0^2 L_s)} \tag{5b}$$

For 'g' representing the capacitance, the elements for parallel LC resonator in the bandpass filter are found using (6).

$$C_p = \left(\frac{\Omega_c}{FBW\omega_0}\right)\frac{g}{\gamma_0} \tag{6a}$$

$$L_p = \frac{1}{(\omega_0^2 C_p)} \tag{6b}$$

The cut-off frequency for low-pass prototype Ω_c is chosen as 1 rad/s and impedance scaling factor γ_0=Z_0/g_0.

The RLC Bandpass Chebyshev Filter derived is shown in Fig. 7 and the values of the parameters are given in Table II.

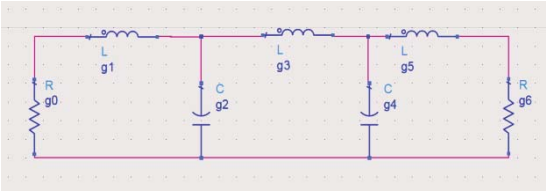


Fig. 6. Five-pole Low-pass prototype

TABLE II. Parameters of Chebyshev Bandpass Filter

i	L_i (nH)	C_i (pF)
1 and 5	7.1576	0.0489
2 and 4	0.1024	3.4233
3	12.3267	0.02844

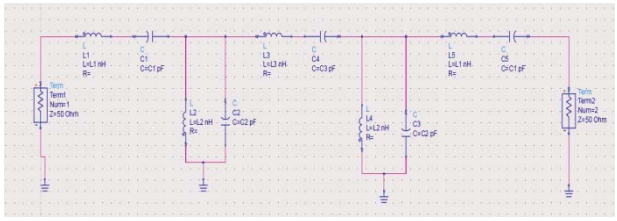


Fig. 7. Chebyshev Bandpass Filter

To design the parallel-coupled microstrip bandpass filters with half-wavelength line resonators, first the characteristic admittances of the J-inverters ($J_{j,j+1}$) are calculated using the following equations [16]:

$$\frac{J_{01}}{Y_0} = \sqrt{\frac{\pi\, FBW}{2\, g_0 g_1}} \tag{7}$$

$$\frac{J_{j,j+1}}{Y_0} = \frac{\pi FBW}{2}\frac{1}{\sqrt{g_j g_{j+1}}} \qquad j=1 \text{ to } n-1 \tag{8}$$

$$\frac{J_{n,n+1}}{Y_0} = \sqrt{\frac{\pi}{2}\frac{FBW}{g_n g_{n+1}}} \tag{9}$$

Where, Y_0= (1/50) ℧ is the characteristic admittance of the terminating line, g_i are the values of the elements of the low-pass prototype filter, and FBW is fractional bandwidth. Next, the even- and odd-mode characteristic impedances of the coupled microstrip line resonators are calculated using the following equations [16]:

$$(Z_{0e})_{j,j+1} = \frac{1}{Y_0}\left[1 + \frac{J_{j,j+1}}{Y_0} + \left(\frac{J_{j,j+1}}{Y_0}\right)^2\right] \quad j = 0 \text{ to } n \tag{10}$$

$$(Z_{0o})_{j,j+1} = \frac{1}{Y_0}\left[1 - \frac{J_{j,j+1}}{Y_0} + \left(\frac{J_{j,j+1}}{Y_0}\right)^2\right] \quad j = 0 \text{ to } n \tag{11}$$

Where Z_{0e} signifies the even mode impedance and Z_{0o} signifies the odd mode impedance. The calculated values of the impedances are shown in Table III. Using these values, the physical dimensions of the Microstrip Coupled Lines are calculated using the inbuilt calculator in ADS, LineCalc. The coupled line 1, 2 and 3 are symmetric to coupled lines 6, 5 and 4, respectively. The calculated physical dimensions for each coupled line are given in Table IV. The dimensions of microstrip line (calculated using LineCalc) are w_0=0.599322 mm & l_0=3.30971 mm.

TABLE III. Circuit Design Parameters of Parallel-Coupled Bandpass Filter

j	$J_{j,j+1}/Y_0$	$(Z_{0e})_{j,j+1}$	$(Z_{0e})_{i,i+1}$
0	0.4533	82.93666	37.60916
1	0.1879	61.16003	42.37045
2	0.1432	58.18392	43.86609

TABLE IV. Microstrip Design Dimensions of Parallel-Coupled Bandpass Filter

Coupled Line (j)	w_i (mm)	s_i (mm)	l_i (mm)
1 and 6	0.363721	0.151062	3.44776
2 and 5	0.543624	0.501387	3.36031
3 and 4	0.564447	0.666008	3.34475

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:27:37 UTC from IEEE Xplore. Restrictions apply.

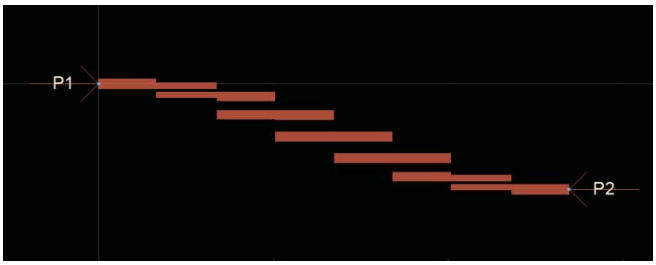


Fig. 8. Top view of the Configuration of the Parallel-coupled Microstrip Bandpass Filter

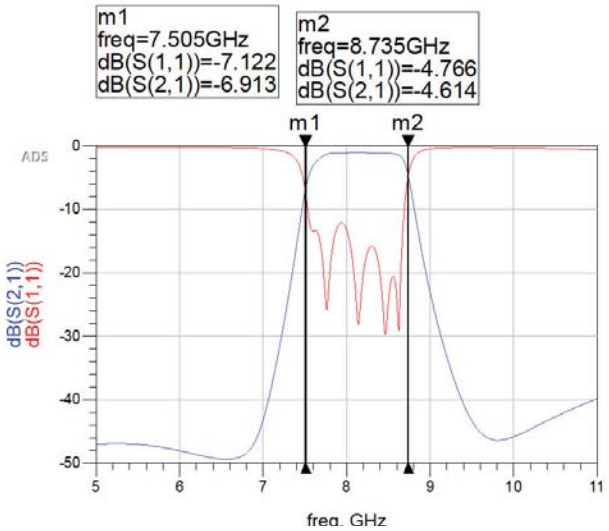


Fig. 9. Simulation of the Parallel-coupled Microstrip Bandpass Filter.

The parallel-coupled microstrip bandpass filter is designed and simulated using Keysight ADS software. The layout is given in Fig. 8 and the simulation result is given in Fig. 9. The simulation result in Fig. 9 has a centre frequency of 8.097 GHz (4.74 % error) and an FBW of 0.1519 (1.28 % error) around the centre frequency.

IV. Proposed Parallel Coupled Line Microstrip Bandpass Filter with Dumbbell DGS

For the proposed Microstrip Bandpass Filter with Defected Ground Structures (DGS), a DGS with similar specifications to the one described in Section II, is etched in the ground plane under each conducting strip at their centre points as shown in Fig. 10. All DGS sections have the same dimensions.

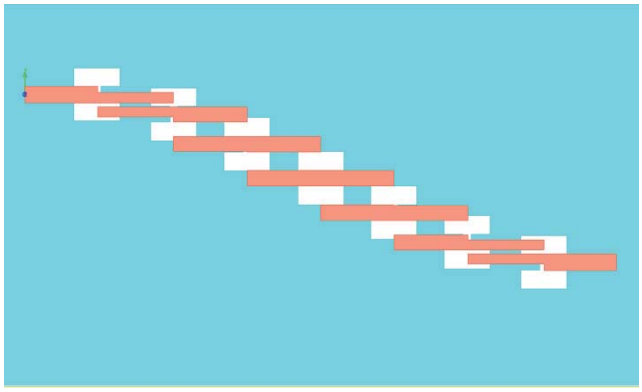


Fig. 10. Layout of the Proposed Parallel-coupled Microstrip Bandpass Filter with Dumbbell DGS.

When the Microstrip Bandpass Filter with f_0=8.5GHz has been simulated with the proposed DGS Filter Configuration (Fig. 10), the centre frequency of the simulation decreases significantly to 7.368 GHz, but the FBW remains the same around the centre frequency with a very insignificant difference.

A different parallel-coupled microstrip bandpass filter is designed & simulated using the same specifications and method as Section III but with a different centre frequency of f_0=10.5 GHz. The calculated physical dimensions for each coupled line for this new bandpass filter is given in Table V. The total length of the microstrip filter has decreased by 5.25 mm.

The microstrip parallel-coupled bandpass filter with f_0=10.5 GHz is simulated both without DGS and with DGS according to proposed design in Fig. 10. The simulation results of both are illustrated in Fig. 11 & Fig. 12.

The simulation in Fig. 11 has a centre frequency of 9.889 GHz and an FBW of 0.1522. The simulation in Fig. 12 has a centre frequency of 8.732 GHz (2.73 % error w.r.t 8.5 GHz) and an FBW of 0.1603 (6.89 % error w.r.t 0.15) around the centre frequency.

Therefore, the inclusion of DGS section in ground plane of the filter has decreased the centre frequency of the bandpass filter. The new design with DGS can simulate a bandpass filter of 8.5 GHz with smaller dimensions for the microstrip lines with the same fractional bandwidth.

TABLE V. Microstrip Dimensions of Parallel-Coupled Bandpass Filter with Centre Frequency = 10.5 GHz

Coupled Line (j)	w_j (mm)	s_j (mm)	l_j (mm)
1 and 6	0.368268	0.154844	2.77839
2 and 5	0.549329	0.517338	2.70358
3 and 4	0.569806	0.687885	2.69033

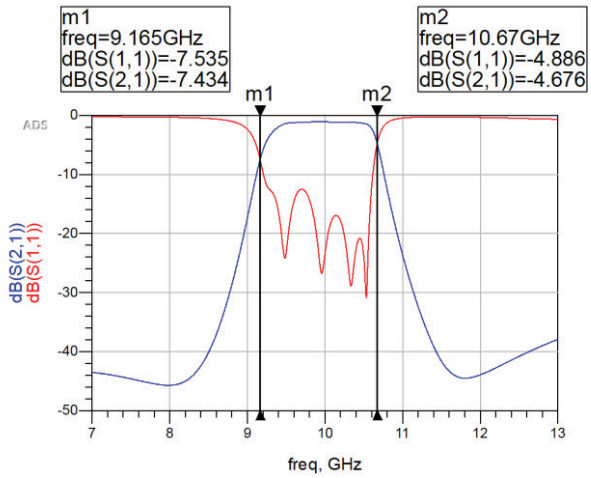


Fig. 11. Simulation of Parallel-coupled Microstrip Bandpass Filter (f_0=10.5 GHz) without Dumbbell DGS.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:27:37 UTC from IEEE Xplore. Restrictions apply.

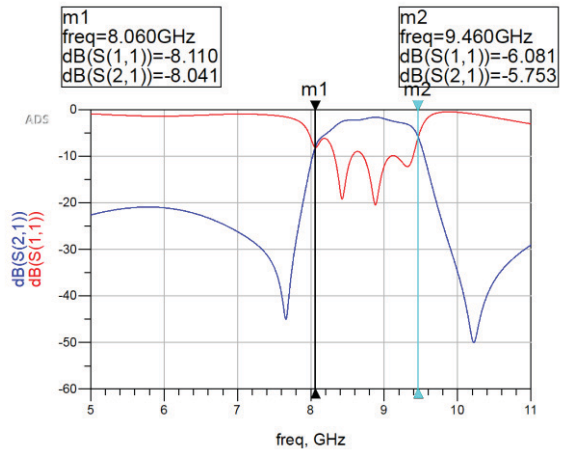


Fig. 12. Simulation of the new Microstrip Bandpass Filter with Dumbbell Defected Ground Structure.

TABLE VI. COMPARING THE RESULTS BEFORE AND AFTER INCLUDING DGS IN GROUND PLANE

Filter	Results		
	Centre Frequency (f_0) (GHz)	*FBW around f_0*	*Total Length (mm)*
f_0=8.5 GHz without DGS	8.097	0.1519	26.925 mm
f_0=10.5 GHz with DGS	8.732	0.1602	21.666 mm

V. EFFECT OF VARYING GRADIENT OF LINE THROUGH CENTRE POINTS OF MICROSTRIP LINES

In this section, the positions of the microstrip lines in the filter developed in section IV is varied along the x-axis, keeping the positions fixed w.r.t the y-axis. A straight line is drawn using the centre points of the first and the last microstrip sections, as shown in Fig. 13. The slope of this line is found to be -0.375 (absolute values are considered). All the other microstrips sections are moved so that their centres lie on this straight line. This filter is then simulated. The simulation has a centre frequency of 8.563 GHz and an FBW of 0.0794 around the centre frequency.

The slope of this straight line is varied and the positions of the microstrip sections are changed to fit the different lines, keeping the position of the top left microstrip section fixed. Two examples with slopes -0.6 & -0.25 are shown in Fig. 14 & Fig. 15. The centre frequencies and bandwidths for different slopes are shown in Table VII.

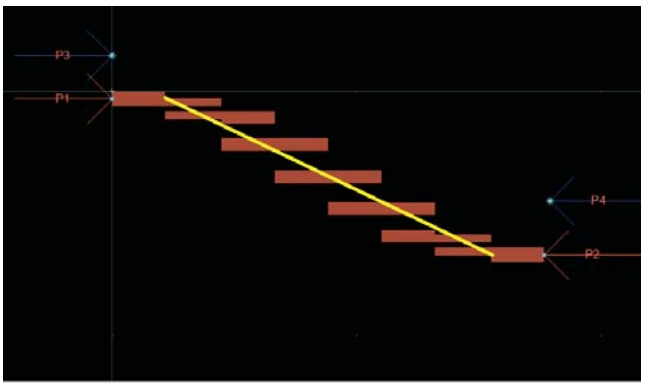


Fig. 13. Straight line drawn using the centre points of the first and the last microstrip sections.

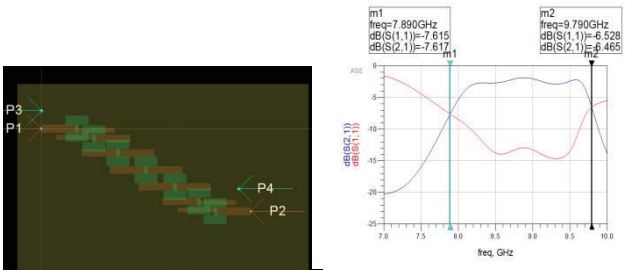


Fig. 14. Simulation Results for slope, m=-0.6

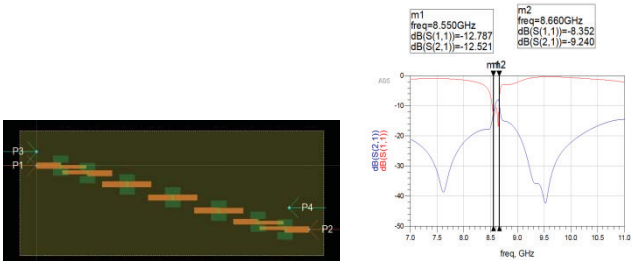


Fig. 15. Simulation Results for slope, m=-0.25

TABLE VII. RESULTS FOR DIFFERENT SLOPES OF STRAIGHT LINE

Slope (m)	Results	
	Centre Frequency (f_0) (GHz)	*FBW around f_0*
-0.6	8.7888	0.2162
-0.5	8.6718	0.1614
-0.375	8.5633	0.0794
-0.25	8.6048	0.0128

It is observed from Table VII that the bandwidth of the filter can be manipulated though variation the slope of the line that passes through the centre points of the microstrip sections. As the slope is decreased from -0.6 to -0.25, the bandwidth decreases. A very narrow bandwidth bandpass filter is obtained at slope -0.25.

VI. CONCLUSION

The study was aimed to design a filter for a centre frequency of 8.5 GHz and obtaining a minimized area for fabrication with the help of dumbbell shaped defected ground structure and after analysing data from different variations, we were able to obtain the desired result. By analysing the output waveforms for different microstrip filter design for different centre frequency, it is noticed that the increase in the centre frequency causes decrease in the area required for microstrip. To make the microstrip compact we decided to design a microstrip filter with centre frequency of 8.5 GHz. The second part of the analysis includes the variation in the gradient values for the slope that passes through the centre points of the conducting strips by keeping the upper corner constant. The bandwidth can be manipulated according to our desired output while keeping the centre frequency constant. The increase in the value of gradient causes an increase in the bandwidth but with minimal changes to the microstrip structure. Similarly, a narrow bandwidth filter can be obtained by decreasing the gradient which would result in lower bandwidth. This filter design can be further improved in the future by analysing it

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:27:37 UTC from IEEE Xplore. Restrictions apply.

with different functions along which the conducting strips are positioned.

REFERENCES

[1] S. B. Cohn. "Parallel-Coupled Transmission-Line-Resonator Filters." IRE Transactions on Microwave Theory and Techniques, vol. 6, pp. 223-231, Apr. 1958.

[2] D. Ahn, J. Park, C. Kim, J. Kim, Y. Qian, and T. Itoh, "A design of the low-pass filter using the novel microstrip defected ground structure," IEEE Trans. Microwave Theory Tech., vol. 49, pp. 86–93, 2001.

[3] L. H. Weng, Y. C. Guo, X. W. Shi, and X. Q. Chen, "An Overview on Defected Ground Structure." Progress In Electromagnetics Research, vol. 7, pp. 173-189, 2008

[4] V. Radisic, Y. Qian, R. Coccioli, and T. Itoh, "Novel 2-D photonic bandgap strucutre for microstrip lines," IEEE Microwave Guided Wave Lett., vol. 8, pp. 69-71, Feb. 1998.

[5] C. S. Kim, J. S. Park, D. Ahn, and J. B. Lim, "A novel one-dimensional periodic defected ground structure for planar circuits," IEEE Microwave Guided Wave Lett., vol. 10, pp. 131-133, Apr. 2000.

[6] D. Ahn, J. -. Park, C. -. Kim, J. Kim, Y. Qian and T. Itoh, "A design of the low-pass filter using the novel microstrip defected ground structure," in IEEE Transactions on Microwave Theory and Techniques, vol. 49, no. 1, pp. 86-93, Jan. 2001, doi: 10.1109/22.899965.

[7] R. Qiang, Wang and D. Chen, "Anovel Microstrip Bandpass Filter with Two Cascaded PBG Structures," IEEE AP-S Digest, pp. 510-513, 2001.

[8] J. P. Kim and W. S. Park, "Microstrip lowpass Filter with Multislots on Ground Plane," Electronic Letters, vol. 37, No. 25, 6 Dec. 2001.

[9] M. Kahrizi, T. Sarkar and Z. Maricevic, "Dynamic Analysis of a Microstrip line Over a Perforated Ground Plane," IEEE Trans. Microwave Theory Tech., vol. 42, No. 5, pp. 820-825 , May 1994,

[10] X. Yi, P. Yeh and J. Hong, "Photonic Bandgap Structure in Thin Film Waveguides," SBMO/IEEE MTT-S IMOC 99 Proceedings, pp. 451-453. 1999.

[11] S. Jeong, "Amplifier Design Using /4 High Impedance Bias Line with Defected Ground Structure (DGS)," IEEE MIT-S Digest, pp. 1161-1163. 2002.

[12] J. I. Park, C. S. Kim, J. S. Park, Y. Qian, D. Ahn, and T. Itoh, "Modeling of photonic bandgap and its application for the low-pass filter design," in Asia–Pacific Microwave Conf. Dig., Dec. 1999, pp. 331–334.

[13] D. Ahn, J. S. Park, C. S. Kim, Y. Qian, and T. Itoh, "A design of the low-pass filter using the novel microstrip defected ground structure," IEEE Trans. Microwave Theory Tech., vol. 49, pp. 86–93, Jan. 2001

[14] D. K. Mahor, "Roll-off Improvement of Open Circuit Stub Band-Stop Filter." International Journal of Research Antennas and Microwave Engineering, Vol. 3, No. 3, pp 112-118, July-Dec 2012

[15] D.M. Pozar, Microwave engineering, 2nd ed., Wiley, New York, 1998

[16] J. I. Park, C. S. Kim, J. Kim, Y. Qian, D. Ahn, and T. Itoh, "Modeling of Photonic Bandgap and Its Application for the Low-pass Filter Design." in Asia-Pasific Microwave Conference Proceedings, pp. 331-334. 1999,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:27:37 UTC from IEEE Xplore. Restrictions apply.

Ain Shams Engineering Journal xxx (xxxx) xxx

Contents lists available at ScienceDirect

Ain Shams Engineering Journal

journal homepage: www.sciencedirect.com

Civil Engineering

Aerodynamic mitigation by corner modification on square model under wind loads employing CFD and wind tunnel

Nikhil Gaur [a], Ritu Raj [b,*]

[a] Delhi Technological University, New Delhi, India
[b] Delhi Technological University, New Delhi, India

ARTICLE INFO

Article history:
Received 13 September 2020
Revised 31 May 2021
Accepted 9 June 2021
Available online xxxx

Keywords:
CFD
Aerodynamic mitigation
Corner modifications
k-ε turbulence model
Power spectral density

ABSTRACT

This paper explores a recent application of the computational fluid dynamics technique "CFD" for wind analysis and its comparisons with the conventional wind tunnel experimentations. This study is centered on the wind response of square and corner cut-shaped building models and its optimization caused by the variation of the wind incidence angle. Extensive rigid model experiments of two building models of length scale 1:100 have been performed in the boundary layer wind tunnel. The numerical analysis has been carried out with the standard k-ε turbulence model to evaluate the force coefficients, base moments, power spectra, external surface pressure coefficients, and flow field characteristics of the models with variable wind angles of incidence. The comparisons between experimental results and CFD analysis suggest the computational approach's viability in wind analysis of tall structures efficiently and accurately. A case study of aerodynamic mitigation by corner cut suggests minor modification techniques performance, efficiency, and limitations.

© 2021 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Engineering, Ain Shams University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The revolutionary scientific, commercial development and advancement of construction techniques, high strength materials, and innovation in architectural approach have led to the construction of super-tall structures. Contemporary tall buildings are remarkably much slenderer, flexible, and lighter than their former precedents. We have moved on from conventional and symmetrical cross-sections to more complex and non-uniform shapes of tall buildings. The shape of the structure plays an important role in the structure's resistance against the wind-induced load and response. The wind load for the tall buildings cannot be generalized due to the wide variability in shapes and surroundings for a building. These advancements in heights and shapes are accompanied by a lack of sufficient damping and low natural frequency of the structures. Such undesired conditions cause excessive vibration, bad serviceability, and occupant discomfort under the action of wind.

This increment of wind load with heights imparts problems related to vortex shedding excitations and wind-induced dynamic response in the range of wind gust. This oscillatory motion generated by dynamic motion is perceived by the occupants and can produce discomfort or a nuisance.

Therefore, the wind-induced building motions are also within the performance design scope, such as comfort level for building occupancy. It must satisfy the serviceability requirements as an occupant's discomfort feeling arises from the building's lateral motion.

The sharp corners of buildings can produce wind flow separation resulting in strong wind-structure interaction induced loads. The bluff structures are more prone to excessive wind loads. Due to flow separation and reattachment around bluff bodies, external shape affects wind interaction with tall buildings. To ensure the functional performance of tall flexible structures and control the wind-induced motion, various types of aerodynamic modifications to the shape and geometry of the buildings are possible. Some aerodynamic modifications in architectural design are effective in reducing the effects of the lateral wind forces significantly. These

* Corresponding author at: Dept. of Civil Engineering, Delhi Technological University, New Delhi 110042, India.
E-mail address: rituraj@dtu.ac.in (R. Raj).

Peer review under responsibility of Ain Shams University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.asej.2021.06.007
2090-4479/© 2021 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Engineering, Ain Shams University.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article as: N. Gaur and R. Raj, Aerodynamic mitigation by corner modification on square model under wind loads employing CFD and wind tunnel, Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2021.06.007

Nomenclature

C_{fy}	Lift force coefficient
Cp	Pressure coefficient
μ	Viscosity coefficient
μ_{eff}	Effective viscosity
μ_t	Turbulence viscosity
P	Air density
Ω	Specific (per unit k) turbulence dissipation rate
P	Buoyant production term
FC	Correction function of eddy viscosity
F	Re-distributed function in the turbulence model
K	Energy of turbulent fluctuations
P_k	Turbulence due to mean velocity
P_{kb}	Turbulence due to buoyancy
Re	Reynolds number
P'	Modified pressure
$C_{1\epsilon}$, $C_{2\epsilon}$,Cμ	Turbulence model constants
σε	Turbulent Schmidt number
SM	Sum of body forces
Tu	Turbulence intensity
U, V	Undisturbed flow velocity
K^2	Fluctuation energy of velocity components normal to the wall
PSD	Power spectral density
S(n)	PSD for lift coefficient
N	Frequency
σ_{Cfy}	Standard deviation of lift coefficient variation at time t
KE	Kinetic energy

aerodynamic modifications in building shape change the flow pattern around the building, thus moderating the wind responses.

Corner modification of building planes such as chamfer, recession, roundness, and slotting are common approaches in this category. In fact, horizontal corner modifications of building shapes have been proven to reduce wind effects on tall buildings effectively. These changes may improve the structural design, but structural measures or damping devices may still be required to serve the objective of wind restraint design in many cases. These improvements turn more beneficial as the corners are progressively rounded [1]. These methods alter the characteristics of the separated shear layers, thus reducing the drag and lift forces, and the wake size of the vortex [2].

A corner modification strategy can only be valid for a specific type and size of the plan [3] and wind direction [4]. Slotted and chamfered corners disturb the vortex shedding, whereas Recessed Corners make smaller separation zones and narrower wakes than a traditional square shape. Virote Boonyapinyo et al. [5] suggested chamfering and recession may produce reductions in the base moment up to the extent of 25%. Contrary to benefits, these modifications may have an adverse effect on structural stability too [6]. The models with corners recessed and chamfered effectively reduce wind loads than corner rounded models [12,13].

Many studies have been performed to investigate the effect of corner modifications and their impact on the aerodynamic forces including rounded corners (Carassale et al. [7]), chamfered corners (Gu and Quan [3]), recessed corners (Kawai [9], Tse et al. [4]), and finned corners (Kwok and Bailey [8], Kwok et al. [10]).

K.C.S. Kwok [10] measured the static and dynamic along-wind displacement responses and the standard deviation cross-wind displacement response at the top of the building model. In chamfered corners, the reductions were more than the slotted corners i.e., up to a 40% reduction in response than the plain building shape. These changes mean a suppressed vortex shedding process, hence a significant reduction in cross-wind response. In chamfered corners, the excitation processes were altered, both in frequency and in magnitude.

N.J. Jamieson et al. [11] conducted a study to investigate the effect of the different corner configurations on the magnitude and distribution of the peak pressure coefficient. Bevelling the corners weakened peak suctions that occur at the trailing edges of the side faces, and the peak suctions on the side faces move forward to the leading edges.

The aqua tower in Chicago, Taipei 101 skyscraper in Taiwan, and Yokohama Building in Japan are prime examples of tall buildings to adopt corner modifications as aerodynamic mitigation. These modifications have led to a 25% decrement in the base moment in Taipei 101 skyscraper. The aqua tower in Chicago has applied rounded corner modification to disrupt the wind flow around the tower. Ping, an International Finance Centre, is tapered first as building shrinks with the elevation. Then each corner of the building is beveled to create a rounded octagonal-shaped floor, effectively controlling the cross-wind effect. A reduction of 35% in wind loads and 32% in the overturning moment was observed [18].

It was evident from past research work that the conventional methods of wind analysis, both codal provisions as well as wind tunnel experiments have their shortcomings in terms of application and finances. Wind codes are restricted with height and conventional geometry provisions. The wind tunnel experiments are too expensive. Therefore, the sophisticated computational methods for numerical simulation have been introduced in wind analysis of buildings, but the reliability of such simulations is yet to be verified.

A series of works have been carried out in the area of wind engineering based on wind tunnel test and Computational Fluid Dynamics (CFD) technique earlier. Computational fluid dynamics analysis of flow characteristics past circular cylinder with confinement was performed by R Deepakkumar et. al. [26] to study flow separation, fluid mixing, and vortex identification.

Prasenjit dey et. al. [27] studied the effect of corner radius and Reynolds number on fluid flow over a square cylinder under steady laminar flow. The results suggested the variation in boundary layer thickness and the displacement thickness was proportional to the corner radius for a particular Reynolds number.

Ahmed Elshaer et al. [14] studied the application of Computational Fluid Dynamics "CFD" modeling for aerodynamic modification performance assessment. The comparison between the low dimensional CFD based methods and results of wind tunnel experiments were produced. The k-epsilon turbulence model was used to simulate turbulence. Higher values for the drag coefficient were produced as the bluffer body induced a larger wake size. A two-dimensional flow field was generated to analyze the models on "STAR CCM+" application using k-ϵ turbulence model. The numerical model was verified by calculating the drag force for the sharp edge square cylinder. The drag coefficients obtained from this study align well with the response of CFD models of Lee et al. [15] and Otsuki et al. [16].

The problem considered by Vikram C.K. et al. [17] was the flow past a single square cylinder, and two square cylinders of the pitch to perimeter ratios of 2, 4, and 6 with and with-out corner modifications carried out numerically by using a commercial CFD code

fluent. A square cylinder had more lift coefficient but lowers strouhal number when compared to chamfered and rounded corner. The lift and drag coefficient was lower for corner modified models because of the large tangential velocity of the square cylinder which made the width of the wake behind the corner modified model small.

This paper aims at utilizing computational fluid dynamics to study aerodynamic mitigation by corner modification, after verifying its reliability with respect to wind tunnel experiment data. Validation of the experimental and numerical data is necessary to justify the simulations for unconventional and irregular geometries. Comparison between force and moment coefficients of CFX simulation with wind tunnel results has been made for three corner modified sections in this study to validate the CFD simulation technique. Thereafter, the same CFX setup is used to study predetermined building corner modifications for a square section. Moreover, results obtained from this study can preliminarily offer an insight into other projects in case the geometrical properties of the models match since the tests provide some normalized parameters.

2. Wind tunnel experiment

The synchronous multi-pressure scanning system (SM-PSS) to measure pressure, high-frequency base balance (HFBB) test to measure base force components, and aeroelastic model which takes account of stiffness and damping properties of the structure are major three experiments performed to analyze wind structure interaction in a wind tunnel. In this study, Wind tunnel experimental data has been taken from the base force component studies of Dr. Ritu Raj et al. [23] work on wind loads on cross-shaped tall buildings.

2.1. Measurement technique

The base shear, base moments, and twisting moments acting on rigid models for several angles of attack of wind incidence varying from 0° to 180° at an interval of 15° are measured by a five component load cell. Average output has been recorded for 60-second observations at intervals of 1 s by the specialized data acquisition system.

2.2. Model description

Rigid models of plywood at a scale of 1:100 are used to model prototype buildings of cross-section 400 m^2 and height 60 m (Fig. 3). Three sections, one square (A) and two with corner cuts (B and C), have been modeled with the same height of 600 mm and a cross-sectional area of 40,000 mm^2 (Figs. 1 and 2). The models B and C have corner cuts of 50 mm and 100 mm respectively.

2.3. Flow characteristics

In this study, the experiments have been performed in an open circuit boundary layer wind tunnel at the Department of Civil Engineering, IIT Roorkee, India. The wind tunnel has a cross-section of 2 m (width) × 2 m (height) and a 15 m long test section (Fig. 5). 99 cm vortex generators, 16 cm barrier wall, and 70 mm, 50 mm, 38 mm cubical blocks (Fig. 4) were set up at upstream test length to produce terrain category-II as per IS 875 (Part 3)-1987 [24]. The models placed on load balance at 10.1 m distance from the upstream test section have been exposed to a wind velocity of 9.78 m/s. Atmospheric boundary layer with turbulence intensity 12% was formed in the wind tunnel.

3. Numerical study

Ansys 19.1 is used to visualize and create wind flow in the virtual wind tunnel tunnel. In the initial stages the structural model and the wind tunnel geometries are created with adequate domain and boundary conditions as per the recommendations by Franke et al. [19]. Stage two involves the meshing refinement as per grid independence study performed in Ansys 19.1 prepossessing meshing tool. The prepossessing is performed with fluent as a solver.

The mass and momentum equation needs to be solved in this simulation as the wind flow is considered to be in-compressible and turbulent, thus density and viscosity remain constant. A k-ε turbulence model is introduced with an appropriate pressure velocity coupling scheme to generate the this turbulence flow in CFX-Pre for a transient simulation. A CFX solver is used for post-processing. Furthermore, the simulation is fine turned with multiple modeling approaches by comparing the CFD analysis results wind tunnel experiment results.

The solver settings in CFX-Pre depend upon the governing equations of the turbulence model adopted. A convergence control of order 1e−6 is adopted for the simulation run. Multiple monitor points are incorporated on the building model to monitor the drag and lift pentameters for along and across wind responses. Ansys CFX-Post is introduced at the final stage of the simulation run to interpret the result and generate graphical solution post analysis. An elaborate explanation of the above process is detailed in the following sections.

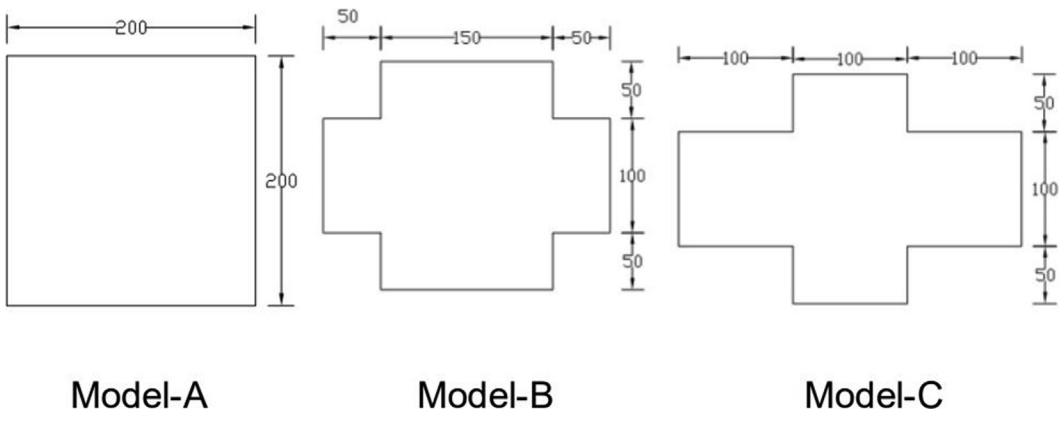


Fig. 1. Model cross-section.

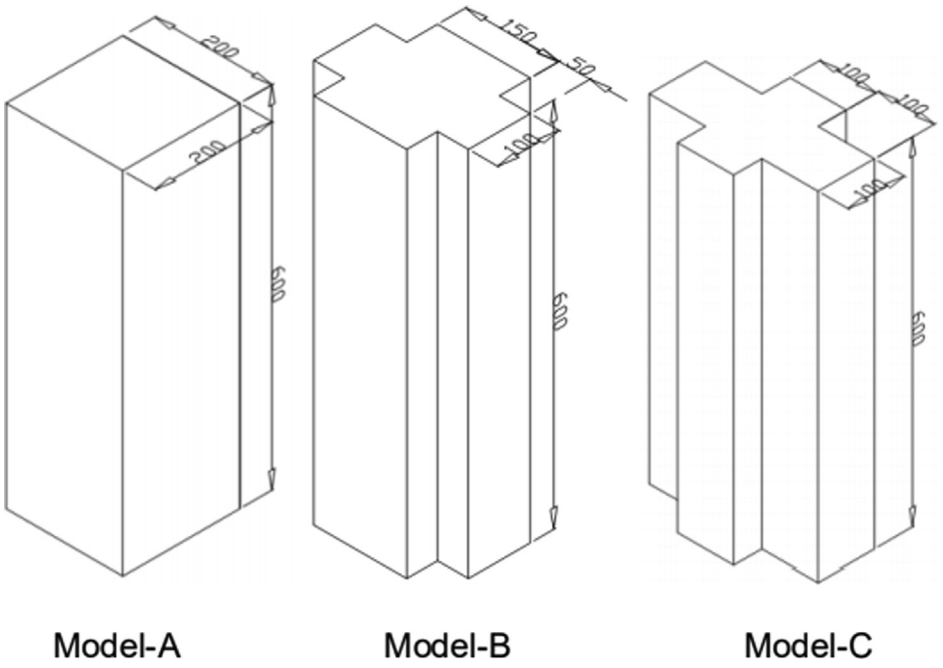


Fig. 2. Dimensions of rigid models.

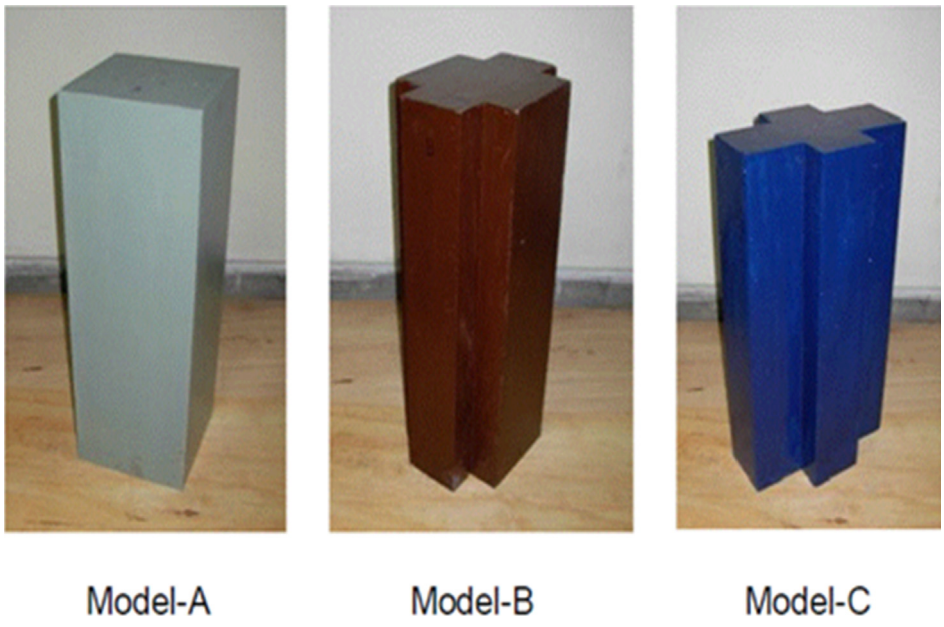


Fig. 3. Wooden models test sections.

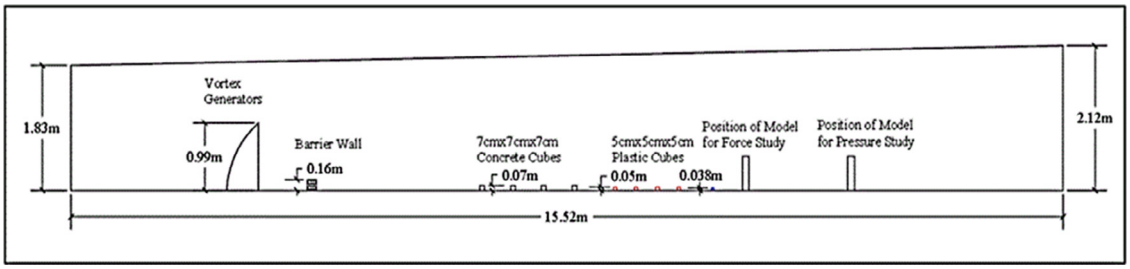


Fig. 4. Elevation view of tunnel section.

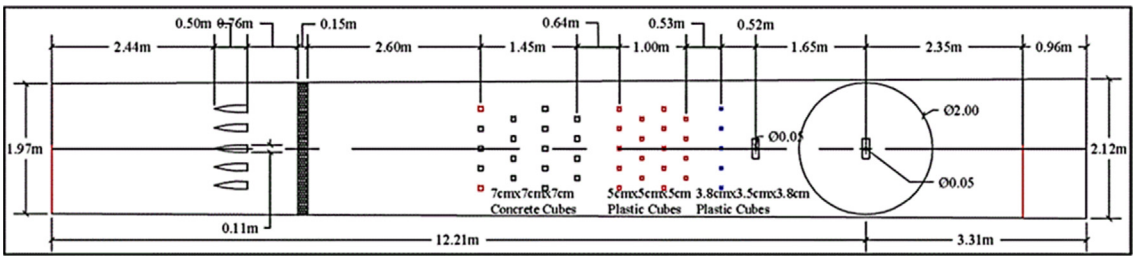


Fig. 5. Flow profile of wind tunnel.

Table 1
Turbulence model k-ε model constants.

	Cε1	Cε2	σε	Cμ
Standard Value	1.44	1.92	1.3	0.09
Modified Value [23]	1.489	2.801	0.373	0.146

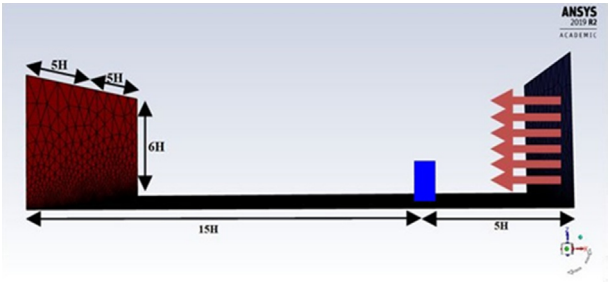


Fig. 6. Domain and model boundaries.

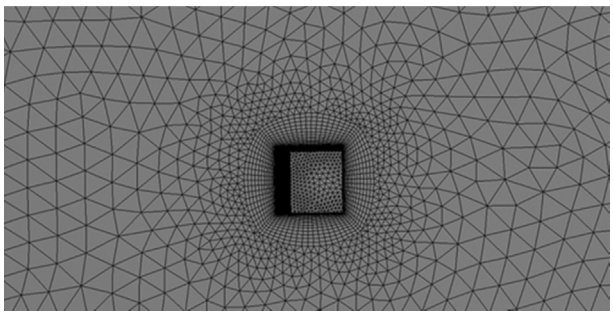


Fig. 7. Structured domain and meshing.

Table 2
Mesh characteristics.

Criteria	Value
Body sizing	0.03 m
Face sizing	0.12 m
Edge sizing	0.01 m
Inflation layer	20
Growth rate	1

3.1. Turbulence model

The CFD software package of Ansys provides multiple turbulence models. However, only a proper turbulence model flow profile generates accurate results as compared to wind tunnel experiments. The accuracy mainly depends upon the computational grids and discretization method incorporated. All the above factors are considered in this study. Numerical predictions are verified by their comparison to the results of a wind tunnel experiment. The standard k-ε turbulence model has been producing accurate results except for non-isotropic k-ε turbulence simulations [20].

Table 3
Boundary conditions and settings.

Boundary Condition	Setting	Distance
Inlet	Velocity = 9.78 m/s	3 m
Outlet	Static pressure = 0 Pa	9 m
Sidewalls	Free slip walls	3 m
Top wall	Free slip wall	3 m
Ground	No-slip wall	–

Carassale et al. [29] presented a case of the Messina Strait Bridge for the simulation of wind velocity fields over large domains by generating three components of turbulence coherent with the classical models of turbulence by the Monte Carlo technique.

Aboshosha et al. [30] proposed a turbulent inflow generator technique based on synthesizing random divergent-free turbulent velocities. This technique aligns both the turbulent spectra and coherence function with the target ABL flow. The base load spectra of LES arrangement aligned with the boundary layer wind tunnel response upto a great extent.

Smirnov et al. [31] proposed a random flow generation technique for inlet boundary generation in LES and particle tracking in LES/RANS turbulence systems by synthesizing divergence-free vector fields of fourier harmonics to represent turbulent velocity fluctuations.

In comparison with the LES, RNG k-ε model and SST model, it was evident that the standard k-ε turbulence model has produced more convergent results. The k-ε model relates the Reynolds stresses to the mean velocity gradients and turbulent viscosity, where 'k' is turbulence kinetic energy i.e., the variance of fluctuations in velocity and 'ε' is the turbulence eddy dissipation. The CFX solver solves the three components of velocity momentum equations, and turbulence kinetic energy and dissipation of turbulent kinetic energy equations. Bhattacharyya and Dalui [22] have also used two equations k-ε turbulence model for wind force and pressure coefficients study on the surfaces of an unsymmetrical E-shaped building using CFD.

The RNG k-ε produces adequate results too but fails in back wall result convergence. The k-ε turbulence is also limited to the fully developed turbulent and non-separated flows. To negate these limitations turbulence coefficients have been modified according to k-ε models developed by Mohammadreza Shirzadi [21]. This modified model produces a less viscous flow, hence more realistic flow features with complex geometry. Therefore, the two-equation k-ε turbulence model was introduced for better accuracy and less numerical effort. Generally, laminar and turbulent flows are represented by Navier-stokes equations for different turbulent length

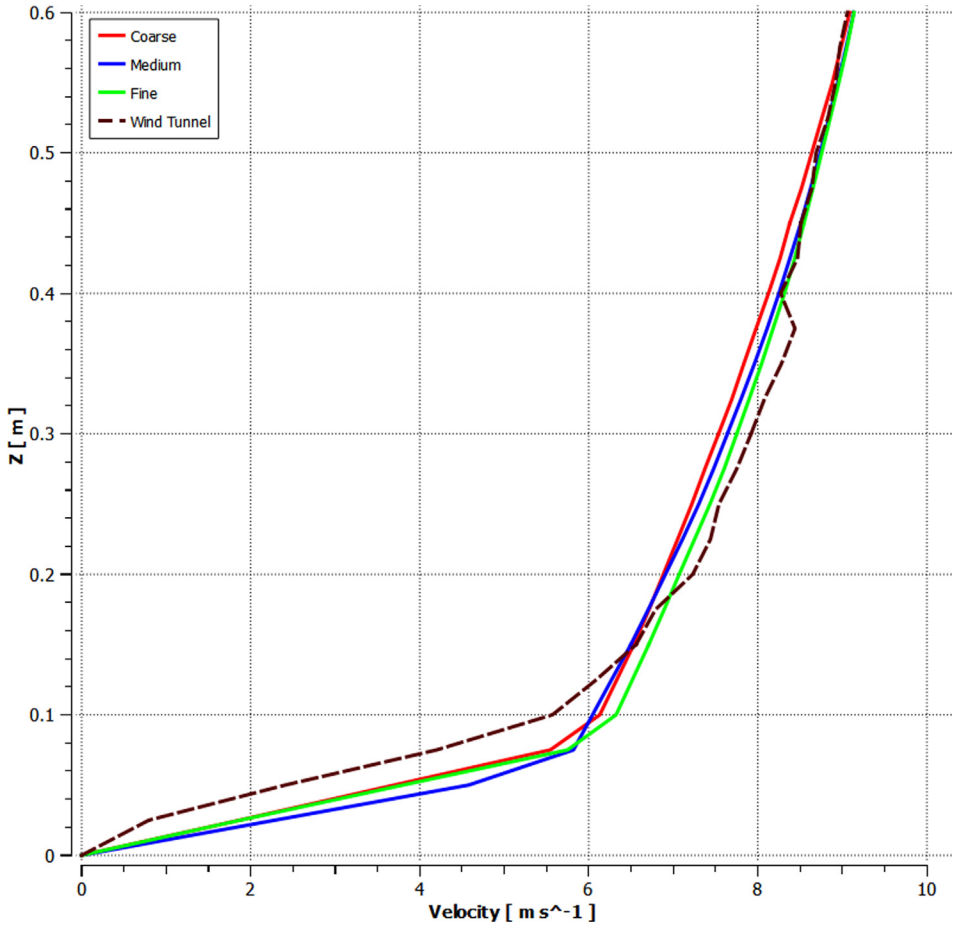


Fig. 8. Wind profile variation with meshing control.

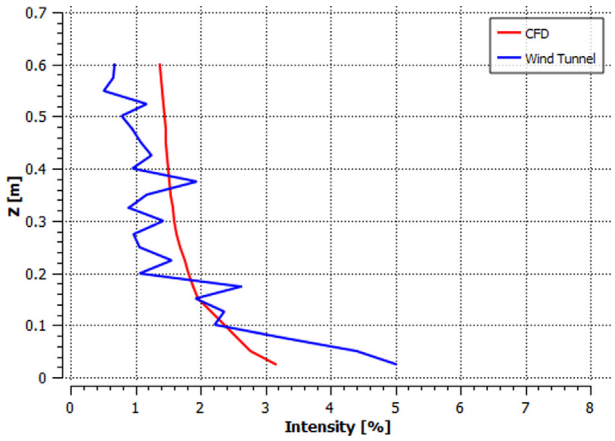


Fig. 9. Turbulence intensity for CFD and In-situ flow.

and time scales as mentioned below. In CFD, turbulence models alter the unsteady Navier-stokes equations by introducing averaged and fluctuating quantities to produce the Reynolds Averaged Navier-Stokes (RANS) equations, thus producing statistical turbulence models.

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial x_j}(\rho U_j) = 0 \tag{1}$$

$$\frac{\partial \rho U_j}{\partial t} + \frac{\partial (\rho U_j)}{\partial t} = -\frac{\partial P'}{\partial x_i} + \frac{\partial}{\partial x_j}[\mu_{eff}(\frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_j})] + S_M \tag{2}$$

The k-ε model is based on the eddy viscosity concept, so that

$$\mu_{eff} = \mu + \mu_t \tag{3}$$

The turbulence viscosity is related to turbulence kinetic energy and dissipation such that

$$\mu_t = C_\mu \rho \frac{k^2}{\varepsilon} \tag{4}$$

Therefore, differential transport equations for the turbulence kinetic energy and turbulence dissipation rate as per differential transport equations [25] are

$$\frac{\partial (\rho k u_j)}{\partial x_j} = \frac{\partial}{\partial x_j}[(\mu + \frac{\mu_t}{\sigma_k})\frac{\partial k}{\partial x_j}] + P_k - \rho\varepsilon + P_{kb} \tag{5}$$

$$\frac{\partial (\rho \varepsilon u_j)}{\partial x_j} = \frac{\partial}{\partial x_j}[(\mu + \frac{\mu_t}{\sigma_\varepsilon})\frac{\partial \varepsilon}{\partial x_j}] + \frac{\varepsilon}{k}(C_{\varepsilon 1}P_k - C_{\varepsilon 2}\rho\varepsilon + C_{\varepsilon 1}P_{\varepsilon b}) \tag{6}$$

The standard and modified k-ε model constants are given below in Table 1

3.2. Domain and meshing

The domain size has been introduced as per suggestions by Franke et al. [19]. The upstream, side, and top clearance from model faces of model height H have been taken as 5H. The downstream clearance has been provided with length 15H so that vortex and backflow on the leeward side can be curbed (Fig. 6). The meshing is produced with physics preference CFD and solver CFX. Tetrahedral meshing with inflation control has been formed to generate

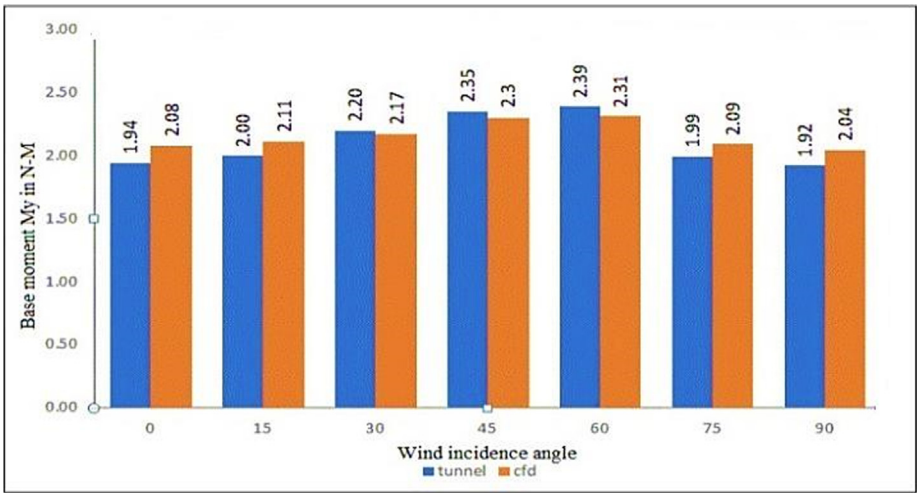


Fig. 11. Moment My values for model A.

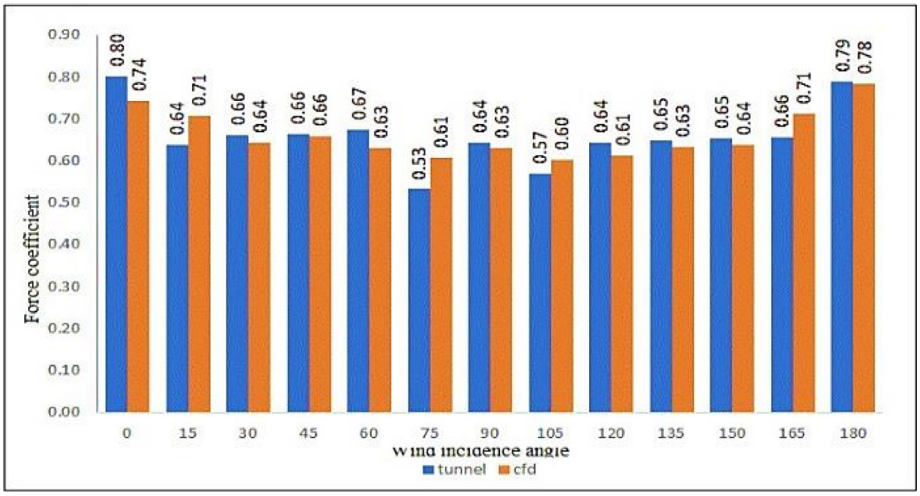


Fig. 12. Force coefficients C_d for model B.

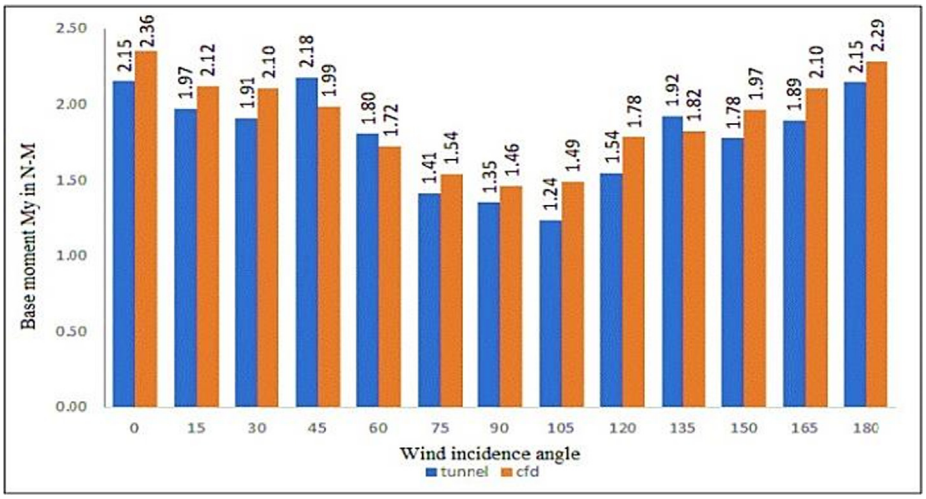


Fig. 13. Moment My values for model B.

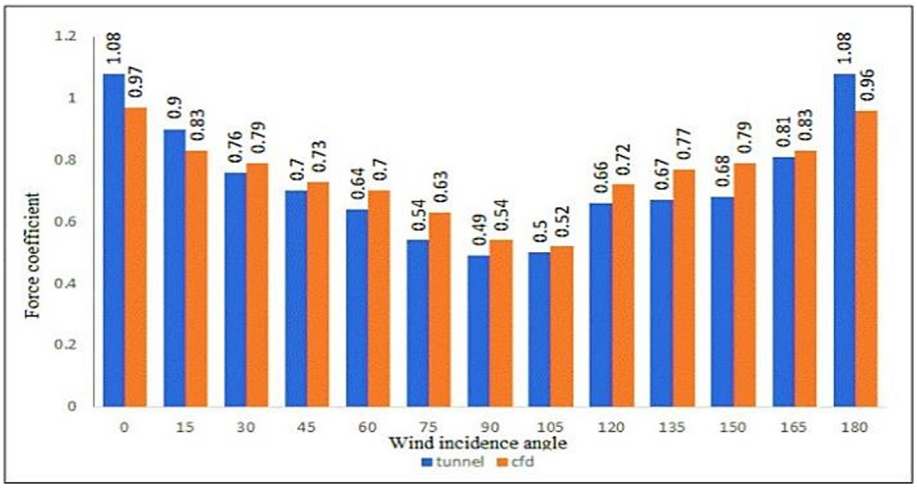


Fig. 14. Force coefficient C_d for model C.

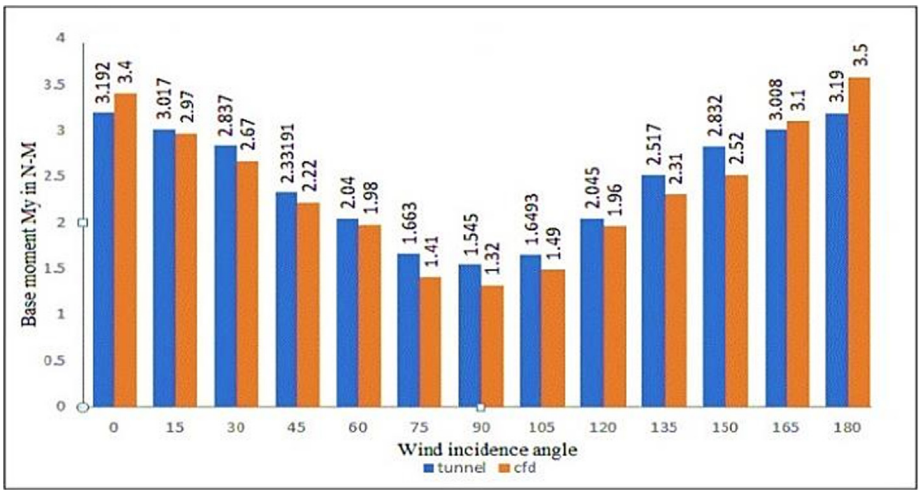


Fig. 15. Moment My values for model C.

Table 4
Error analysis of drag force response.

Model	R	MAE	MAPE	RMSD
A	0.493	0.063	8	0.075
B	0.809	0.033	5.2	0.041
C	0.953	0.067	9.2	0.075

Table 5
Error analysis of moment response.

Model	R	MAE	MAPE	RMSD
A	0.986	0.09	4.4	0.097
B	0.893	0.168	9.8	0.176
C	0.977	0.178	7.7	0.203

uniform flow around the model. The size function for the mesh is set as curvature to maintain the shape of the model. Skewness and orthogonality controls have been used to check the mesh quality.

Table 6
Force and moment observations of test models.

Model	Corner cut [mm]	Drag force [N]	Lift force [N]	Moment [N-M]
A	0	5.70	0.010	1.92
B	6.25	5.20	−0.058	1.72
C	12.5	4.22	0.006	1.46
D	18.75	4.44	−0.019	1.52
E	25	4.88	0.025	1.56
F	31.25	5.34	−0.011	1.80
G	37.5	5.50	−0.003	1.84
H	43.75	5.72	−0.023	1.90
I	50	5.84	−0.021	1.95

The ANSYS (CFX) uses the finite volume method to discretize the region of interest into the finite number of cells. The discretization square model is shown in Fig. 7.

Face sizing: An unstructured tetrahedral element size of 0.01 m has been provided to the model walls. A 20 layered inflation is also provided to form uniform flow. Further refinement is conducted on the bottom boundary near the model domain for the accuracy of the model results. Mesh characteristics are listed in Table 2.

3.3. Boundary conditions

The wind tunnel experiment boundary conditions have been introduced so that numerical and experimental results can converge. Boundary layer wind flow is generated by incorporating inlet velocity with power-law at the inlet boundary. The gauge pressure at the outlet has been considered as 0 Pa. The ground surface and model surface are provided with no-slip condition with ground roughness 0.0024 m. The sidewalls and top wall of the tunnel domain have free slip condition so that no force generates there. The boundary wall conditions are mentioned in Table 3.

3.4. Grid independence study

Unstructured grids are incapable of condensing mesh nodes, thus providing less accurate computational results. About 0.4 million grid cells are introduced in the domain, out of which 0.16 million are on the tunnel floor. The cell height of the grid is also allotted according to y+ study.

Grid independent studies are performed to generate grid-independent numerical solutions. The mesh must be divided into higher orders to reduce vortex sizes in wake regions and achieve accuracy in CFD simulation. It is observed that simulations with a higher number of nodes produce better accuracy. However, a higher number of nodes will increase simulation time proportionally. The results of convergence study for finer base sizing with constant face sizing and inflation suggest that mesh and results are better converged in case of finer sizing. The variation in the wind profile with the finer meshing are presented in Fig. 8. The velocity profile align with the wind tunnel profile the most with the medium to finer meshing control.

4. Results and discussion

4.1. Comparative study

A comparison with the wind field quality between wind tunnel test and the virtually simulated wind tunnel is provided to compare the accuracy of the CFD simulations. The turbulence intensity has been taken as a common parameter for numerical and experimental studies to compare the in-situ measured turbulence characteristics with the numerical characteristics. A study by Cao et al. [28] on wind characteristics of a strong typhoon Maemi suggested that the turbulence intensity decreases with wind speed and remain almost constant when the wind speed becomes high. The numerical and experimental turbulence intensity responses for this study align with the findings by Cao et al. [28] as evident from the Table.

A new approach for analyzing wind speed was proposed by Xu et al. [32] in which a wind sample is decomposed into the sum of a deterministic time-varying mean wind speed plus a stationary fluctuating wind component using the empirical mode decomposition. The spectral density functions of fluctuating wind components obtained by this approach has a lower amplitude in the low-frequency range than those gained by the traditional approach. However, the spectral density functions of fluctuating wind components obtained in the higher-frequency range are very similar in this approach. The traditional approach has been adopted in this study to evaluate the turbulence intensity as the wind response in high frequency region is prominent for flow characteristics. The intensity of CFD simulation and the in-situ condition is presented in the Fig. 9.

A comparative study between numerical simulations and the wind tunnel experiment results has been carried out. It was observed that the CFD simulations were at par with the tunnel results. Force coefficient (drag) and moment in Y direction have been compared for this analysis. The convergence appears to decrease with corner cut models because of the complexity of meshing. Orthogonality and skewness in corner cut models also affected the results. A finer face and mesh sizing in the corner cut models resulted in more converged results. Interpolation has been adopted for significant discrepancies in results. The comparative graphs for force coefficient and moments have been presented in Figs. 10–15.

4.1.1. Error analysis

A statistical error analysis has been performed to obtain the convergence accuracy in computational simulation and wind study. The Pearson's correlation coefficient (R), mean absolute error (MAE), mean absolute percentage error (MAPE), root mean standard deviation (RMSD) values have been calculated to compare the results. The force coefficient and moment coefficients have been shown in Table 4 and Table 5 respectively. The RMSD values suggest better convergence for the rectangular model than the corner cut model. The results indicate the accuracy of CFD simulation in more complex geometries may vary but are within acceptable limits. Therefore, this modeling can be trusted with future simulations.

4.2. Case study

In present work, implications of corner configuration and performance of computational fluid dynamics in the study of wind interactions have been performed. A case study has been conducted based on the results of the numerical simulations to study aerodynamic mitigation by minor modification in terms of corner-cutting. The same flow, domain, and boundary conditions have been applied to new models with improved face and edge meshing to compare corner modification effects on structures. A square section of dimension 200 mm has been taken for this study. The height of the section has been taken as 600 mm. A corner cut at the 6.25 mm rate on all sides has been applied in the square model, and force, moment and pressure variations in the CFX solver have been observed. The results listed in Table 6 suggest significant force and moment deductions for 12.5 mm, 18.75 mm, and 25 mm corner cut cases. These results are further explored in subsections.

4.2.1. Power spectral density

This section explores the power spectral density function for a more nuanced study of the across wind response. The power spectra are used for a better understanding of across wind responses of test models of smaller magnitudes. Determination of the power spectral density (PSD) of wind load is important in the accurate analysis of structures in time domain. Tao et al. [33] presented a study on the analysis and modelling of the longitudinal evolutionary power spectral density (EPSD) of typhoon winds by examining high-frequency sub-ranges with the help of spectral analysis. An analytical approach was developed to determine the EPSD model with the high-frequency sub-range.

Tao et al. [34] in their next study analysed typhoon-induced non-stationary responses of the stonecutters cable-stayed bridge in Hong Kong with time-invariant wind coherence. The comparison with wind experiment results indicate the enhanced framework for buffeting analysis with time-varying wind spectra and coherence is very accurate and should be adopted in the non-stationary buffeting.

Tao et al. [35] in their study of spectral modelling of typhoon winds used spectral modification to the isotropic turbulence to produce a nexus between longitudinal and lateral spectra of typhoon winds. A unified semi-stationary wind velocity model

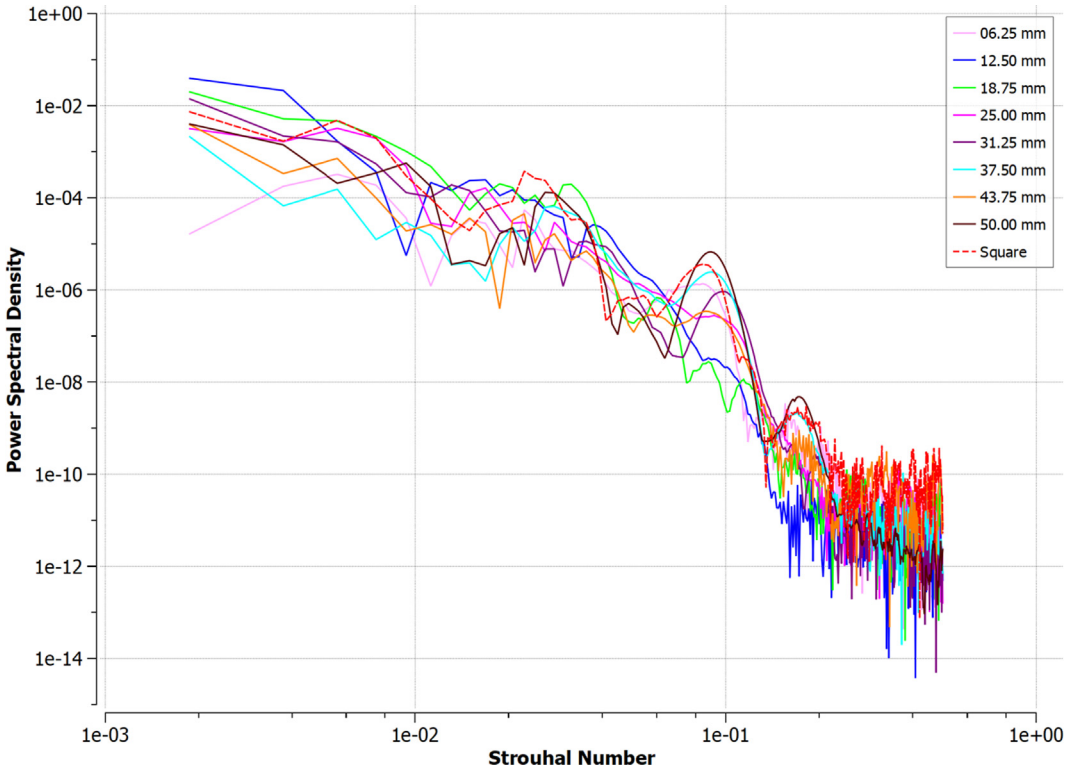


Fig. 16. PSD for across wind response.

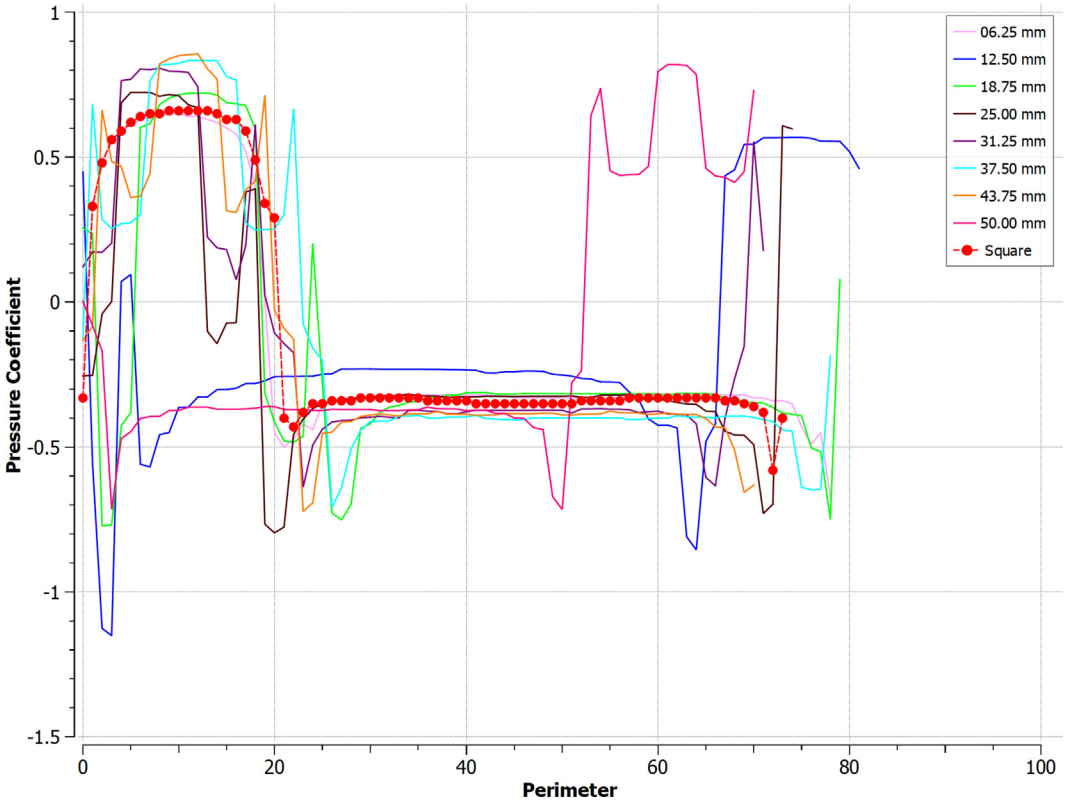


Fig. 17. Pressure coefficient curve for polyline at 540 mm.

was generated by treating turbulence in a semi-stationary manner to overcome the limitations due to deviations from the measured spectra . This model was further evolved into a time–frequency form to fit the measured EPSD. The comparison between measured and fitted EPSDs suggest this approach is very effective without additional subjective operations on the EPSD model.

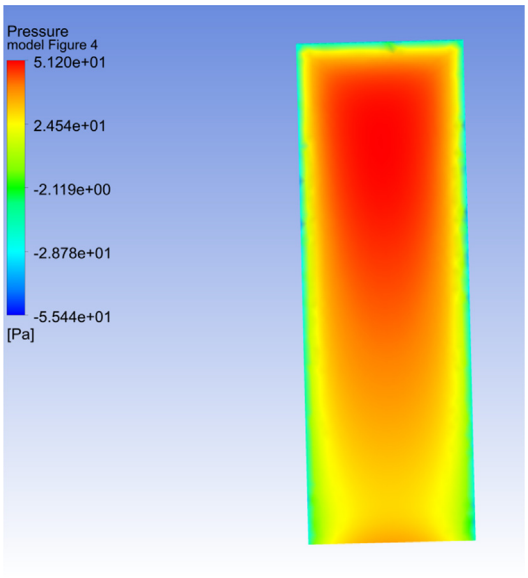


Fig. 18. Front face square model.

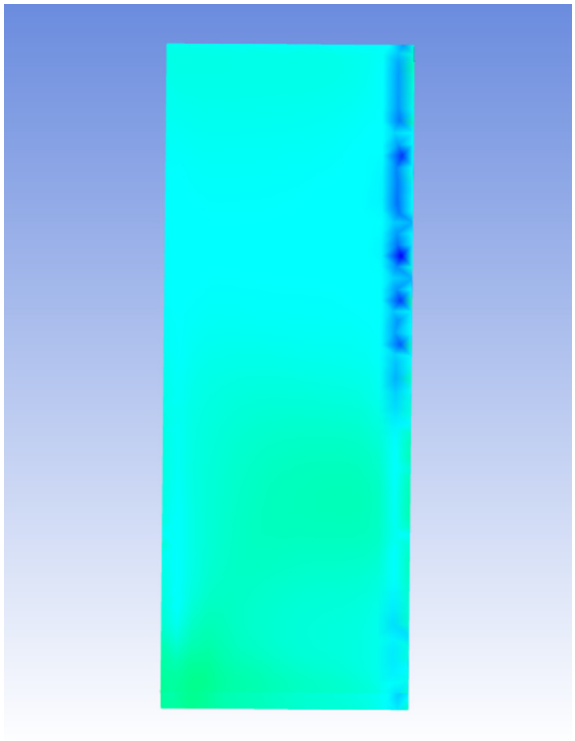


Fig. 19. Side face square model.

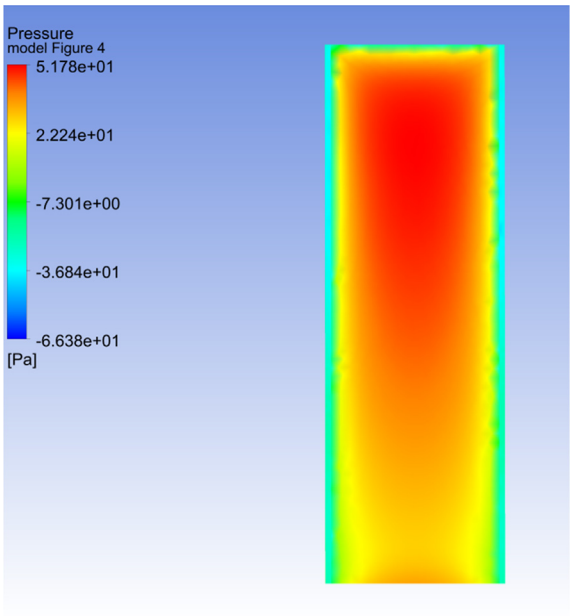


Fig. 20. Front face 6.25 mm corner cut.

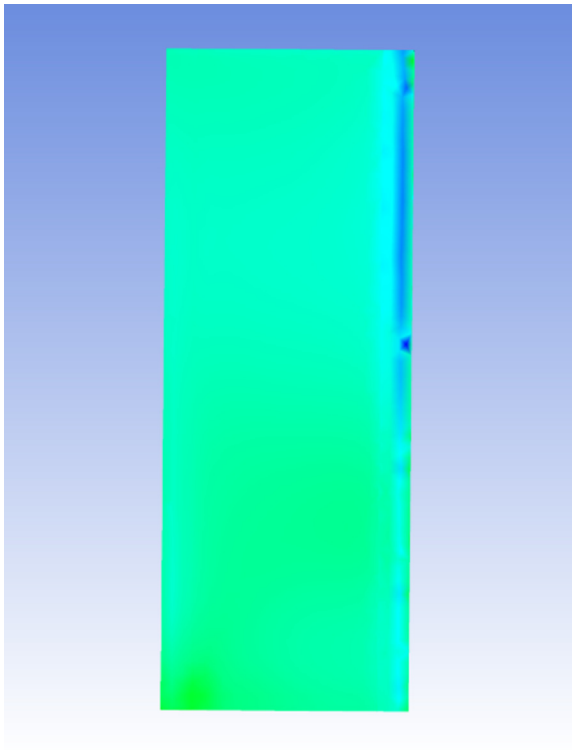


Fig. 21. Side face 6.25 mm corner cut.

The comparison of the power spectral density and coherence can not be presented for this study as the wind tunnel experiment is performed for an aerostatic environment. The base shear, base moments, and twisting moments acting on rigid models are measured by a five component load cell in this study. Hence, the comparison is presented in case of static factors such as base force and moments as presented in further sub-sections.

A closed-form expressions of the statistical errors introduced by interpolation in order to reduce interpolation-based errors in simulation of the multivariate wind field and verified these by a numerical example. The study suggests that the Hermite interpolation is preferred to reduce the errors in a traditional simulation of the multivariate wind field utilizing the spectral representation method (SRM) [37]. A reduced 2D Hermite interpolation-enhanced approach was proposed later by Tao et al. [38] to enhance the efficiency of spectral representation method in simulating fully non-stationary wind fields with respect to the power spectra and correlation functions.

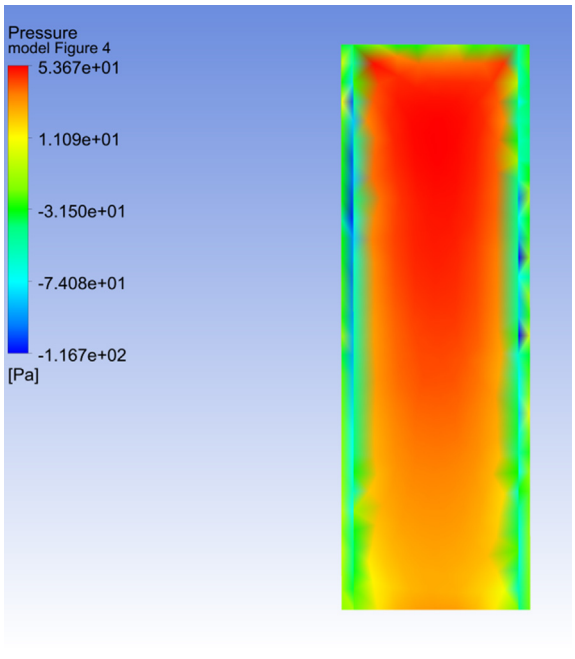


Fig. 22. Front face 12.5 mm corner cut.

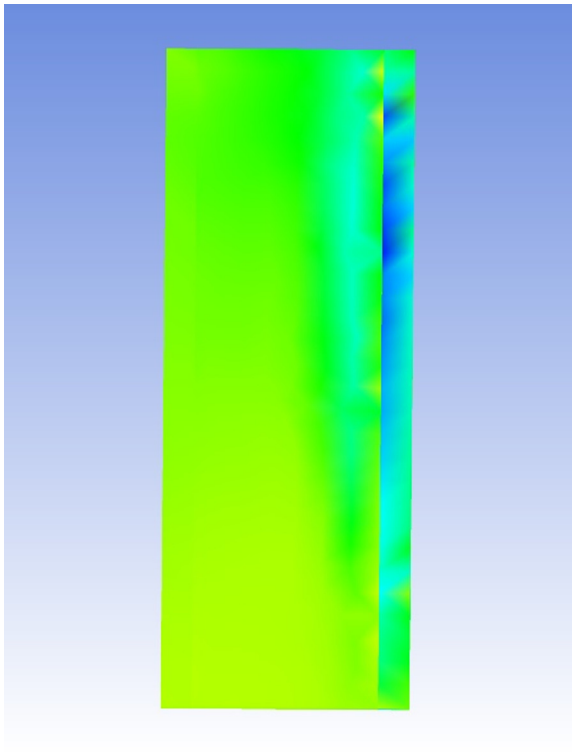


Fig. 23. Side face 12.5 mm corner cut.

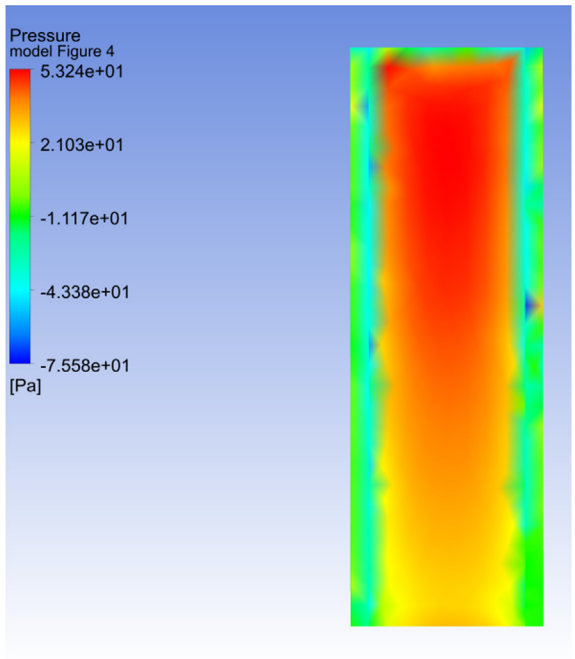


Fig. 24. Front face 18.75 mm corner cut.

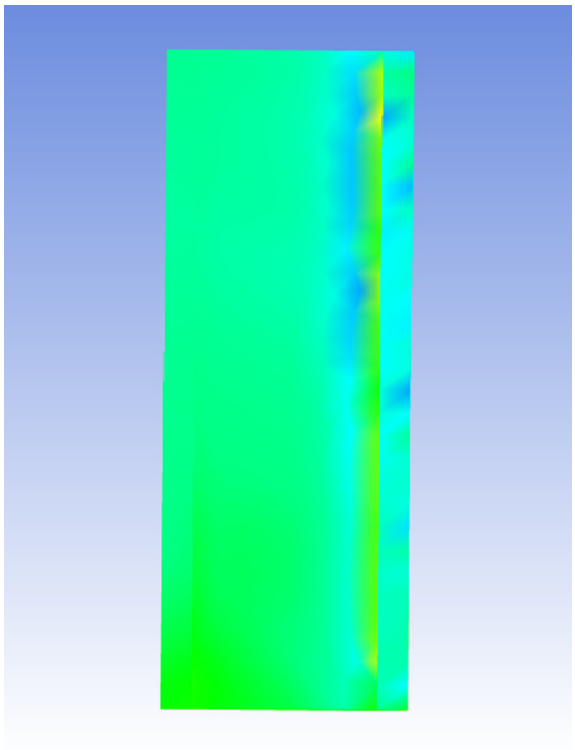


Fig. 25. Side face 18.75 mm corner cut.

Huang et al. [36] suggested that the spectral variations in the non-stationary winds are found to be weak, and the uniformly modulated process could be used to model these winds. Thus, the hannings FFT model is used to plot the power spectral density of lift coefficient (nS(n)/σCfy) versus strouhal number (nD/H) for this simulation.

It is evident from the PSD plot (Fig. 16) for across wind response that the spectra of all corner cut cases flow under the square model spectra, thus producing smaller magnitudes of across wind force than the square model. However, the 50 mm corner cut spectra flows over the square model's spectra, thus producing greater magnitudes of wind response. The gap between PSD responses is least for 6.25 mm corner cut and square model thus, magnitudes of the

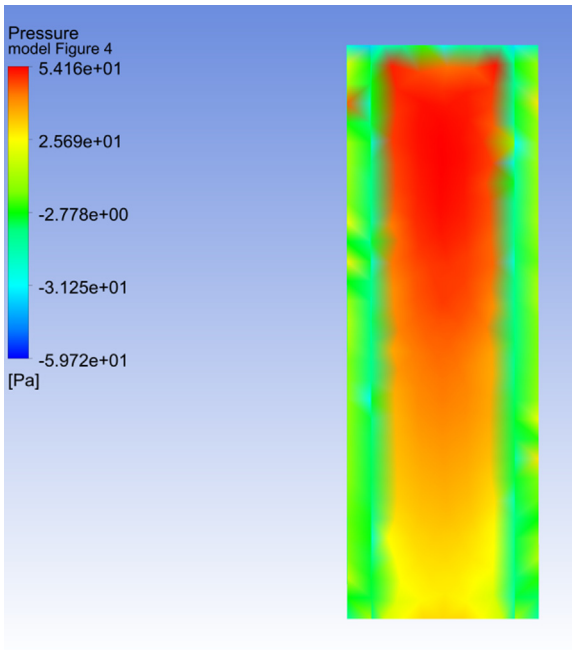


Fig. 26. Front face 25 mm corner cut.

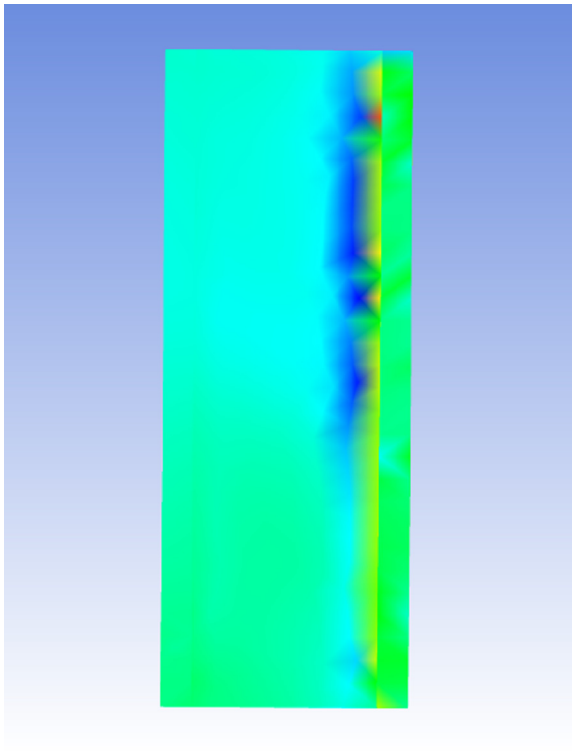


Fig. 27. Side face 25 mm corner cut.

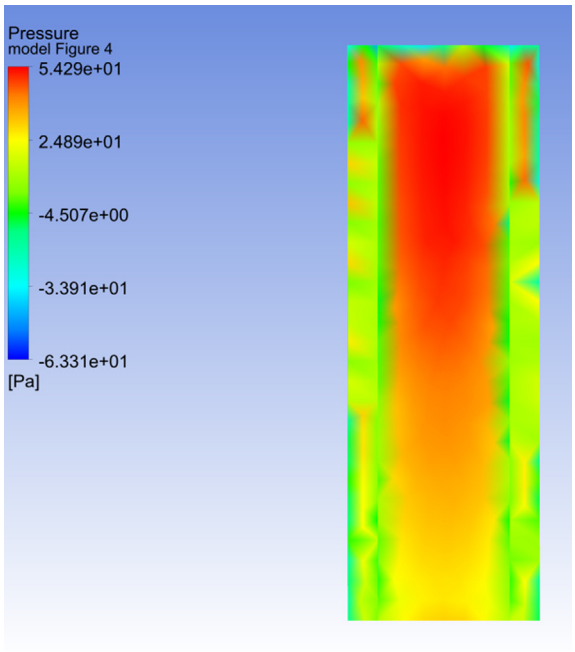


Fig. 28. Front face 31.25 mm corner cut.

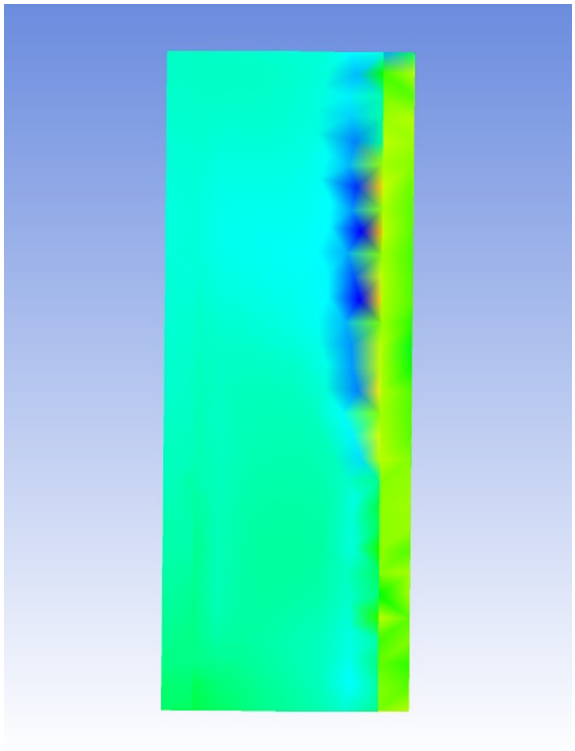


Fig. 29. Side face 31.25 mm corner cut.

force were in closer proximities. The most docile PSD curves were observed for 12.5 mm, and 18.75 mm corner cut cases.

4.2.2. Horizontal pressure coefficient

The pressure coefficient variation due to different corner cut dimensions has been discussed in this section. A perimeter is placed at 0.9H height i.e., 540 mm from the base on the model in the form of a polyline to measure the external pressure coefficient. A series of four probe points are also placed on this pressure belt to monitor the transient pressure variations.

It is evident from the Cp plot (Fig. 17) that for corner cuts of 12.5 mm and 18.75 mm, the front face Cp values turn negative, which is responsible for the reduction in drag force in these cases. The transient pressure study on the front face later suggested suction points on the model's front face. In 12.5 mm corner cut condition, windward projections for front and side faces reverse in

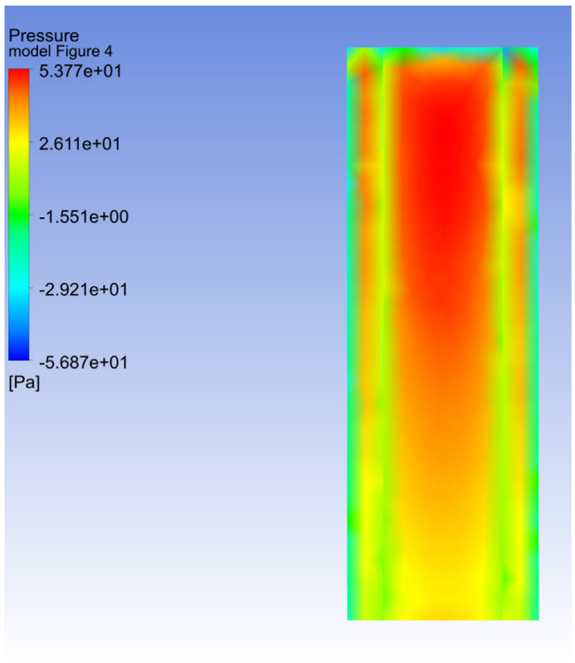


Fig. 30. Front face 37.5 mm corner cut.

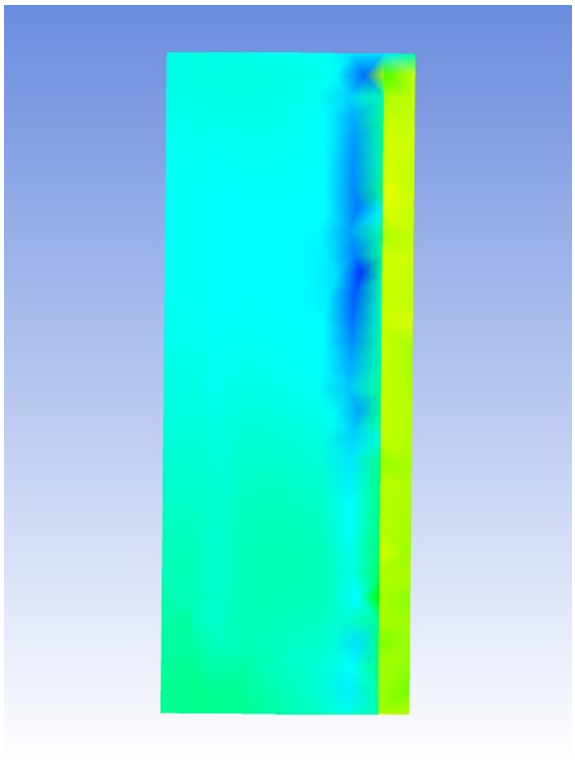


Fig. 31. Side face 37.5 mm corner cut.

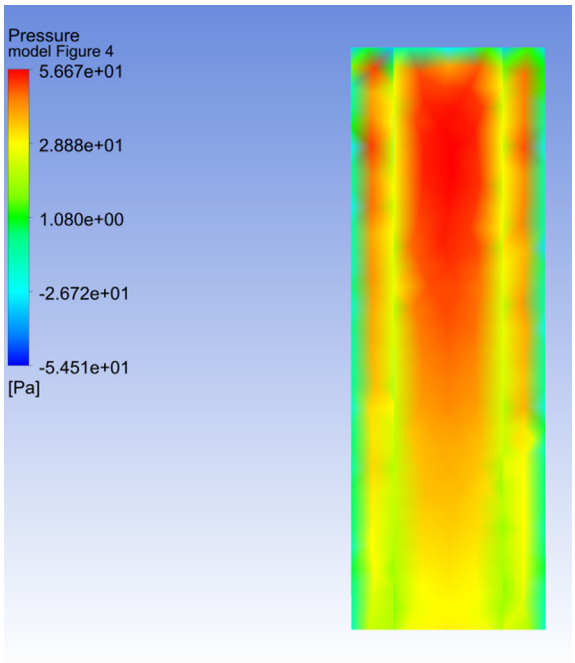


Fig. 32. Front face 43.75 mm corner cut.

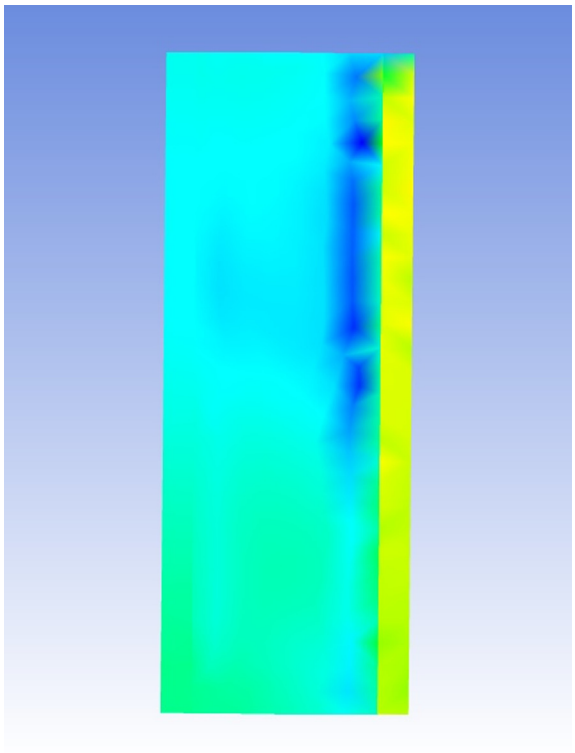


Fig. 33. Side face 43.75 mm corner cut.

magnitudes as corner cut disturbs the flow profile around the model. The maximum variation in Cp values is observed for 12.5 mm and 50 mm corner cut case. The increment in positive Cp in the case of 50 mm cut is observed 24.8%, thus producing higher drag values. The increment is responsible for more turbulence in the wake regions of the model. The maximum negative Cp value doubles in the case of a 12.5 mm corner cut.

The Cp values for the side faces reduce further than the square model values in case of corner cut models as evident from the perimeter set between 20 and 40 cm and and 60–80 cm (Fig. 17). This is further emphasized by the more negative values at the lower ends of the pressure contour legends for corner cut models (Figs. 20–35) than a square model (Fig. 18). This is due to suction around the corner cuts and reattachment of flow in the areas of turbulence in case of corner cut model. This reattachment stabi-

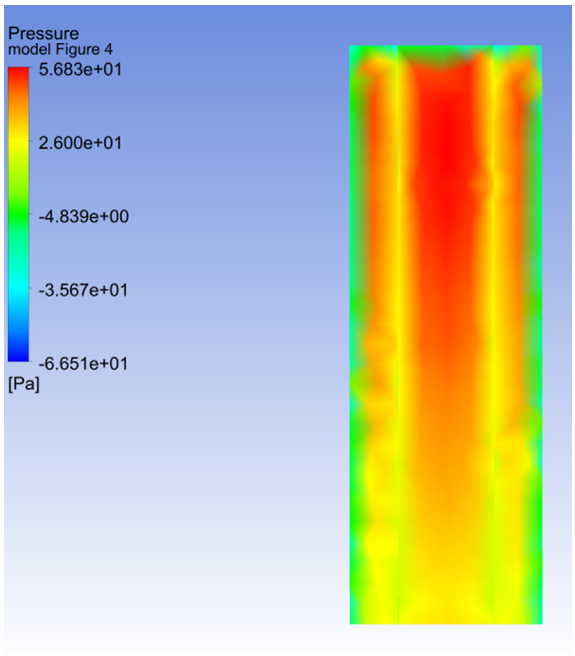


Fig. 34. Front face 50 mm corner.

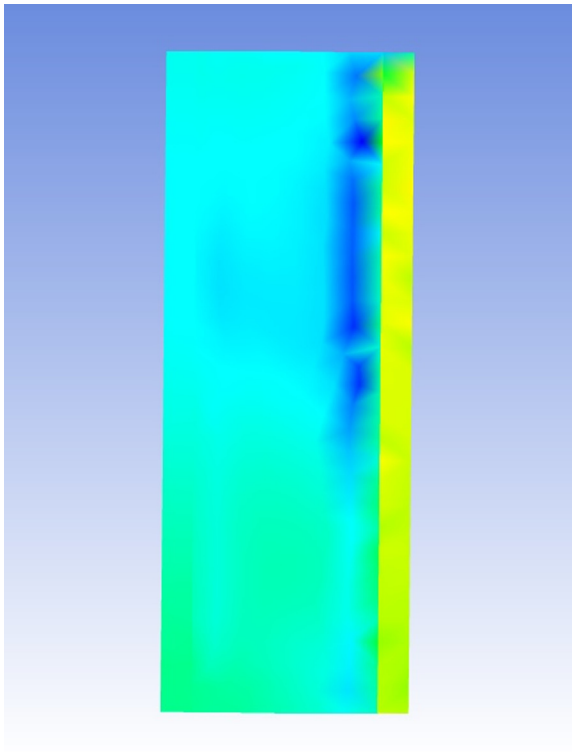


Fig. 35. Side face 50 mm corner cut.

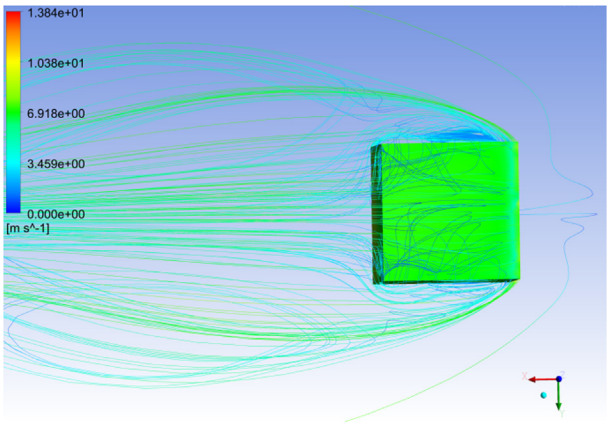


Fig. 36. Velocity streamline for square model.

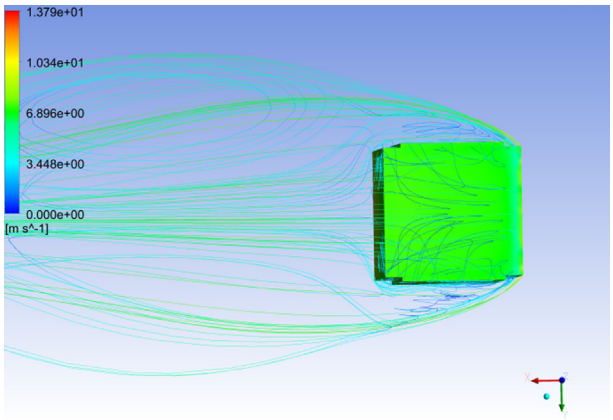


Fig. 37. Velocity streamline for 6.25 mm corner cut.

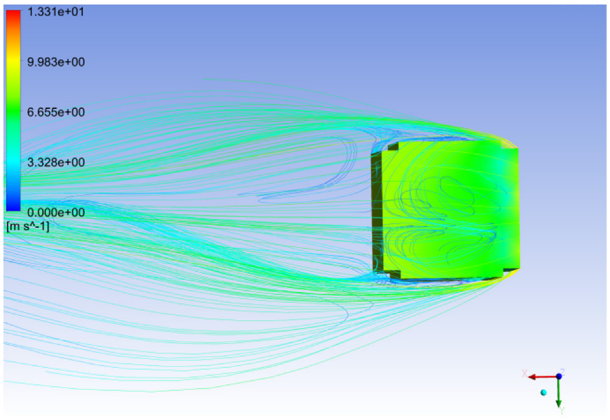


Fig. 38. Velocity streamline for 12.5 mm corner cut.

lizes the wake turbulence, and suction is developed on the model face thus resulting in a reduced across wind response. The flow separation and reattachment are further confirmed based on the velocity streamlines (Figs. 36–44). The pressure contours for front and side face of the models (Figs. 18–35) further suggest that the extent of suctions at the corner cuts enhance with the increased corner cut dimensions.

4.2.3. Kinetic energy fluctuations and turbulence

On monitoring probability density function, many subsequent peaks were observed. These subsequent peaks may suggest vortex shedding and reattachment of the separated turbulent flow. These flow conditions produce turbulence in the wake regions. The vortex shedding and turbulence majorly contribute to the kinetic energy (KE) in the wake region. A section plane was created at the top of the model to analyze the pattern of wake region employing kinetic energy fluctuations and compared them with the square model (Figs. 45–53). The kinetic energy plot of the square model

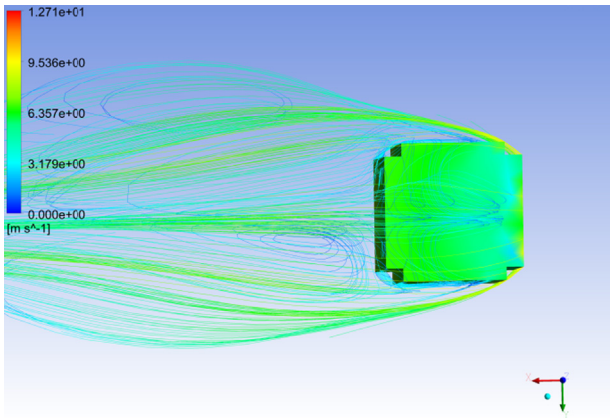


Fig. 39. Velocity streamline for 18.75 mm corner cut.

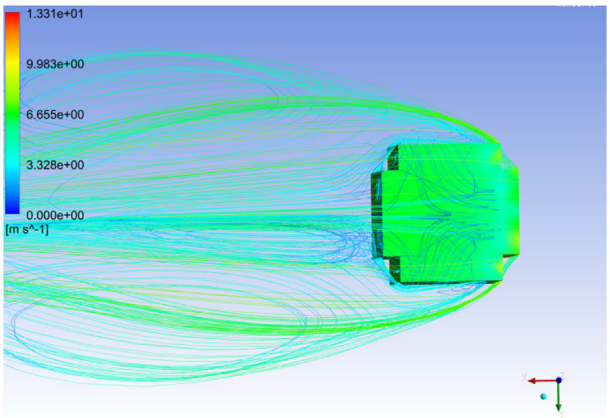


Fig. 42. Velocity streamline for 37.5 mm corner cut.

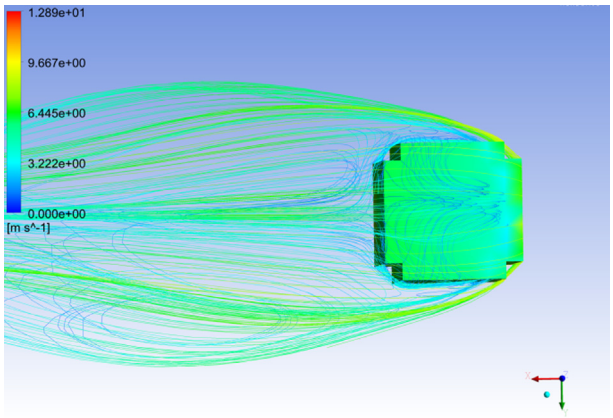


Fig. 40. Velocity streamline for 25 mm corner cut.

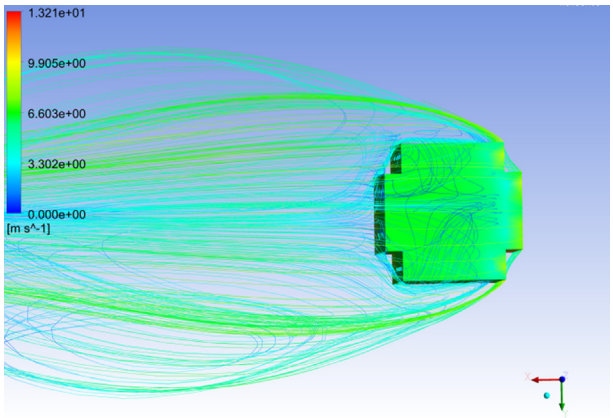


Fig. 43. Velocity streamline for 43.75 mm corner cut.

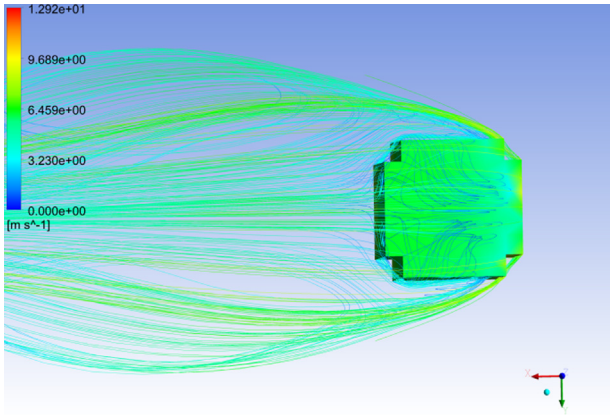


Fig. 41. Velocity streamline for 31.25 mm corner cut.

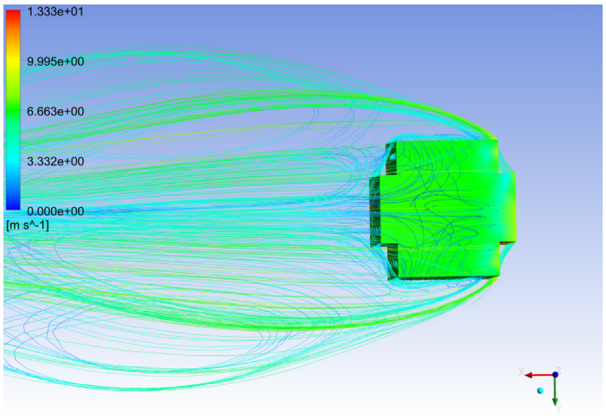


Fig. 44. Velocity streamline for 50 mm corner cut.

suggests flow separation at model corners. The corner cuts help to reattach this separated flow. The reattachment of flow is observed in the wake region in these cases. In the case of 12.5 mm and 18.75 mm corner cuts, the reattached flow was the most stable, and prolonged wake religion was observed. This reattachment helps to stabilize the wake turbulence, and suction is developed on the model face which was evident from the Cp plots on the polyline.

Furthermore, these responses showed prolonged and broader wake religions as magnitudes of corner cuts increase. A higher magnitude of turbulence with a larger wake in kinetic energy plot was observed for 50 mm case. A reduction of 50% in wake region kinetic energy magnitude was observed for all cases except 6.25 mm corner cut.

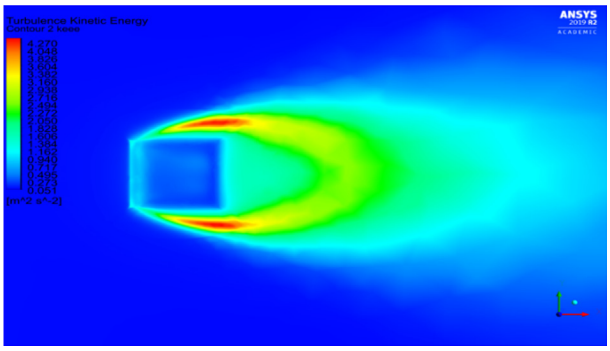


Fig. 45. Kinetic energy spectrum for square mode.

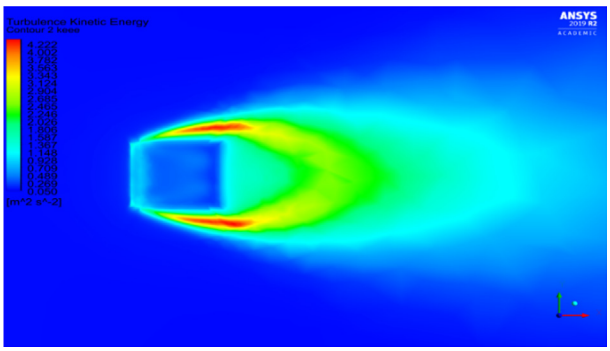


Fig. 46. KE spectrum for 6.25 mm cut section.

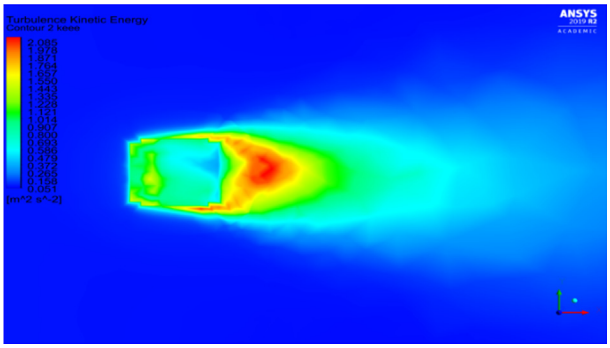


Fig. 47. KE spectrum for 12.5 mm cut section.

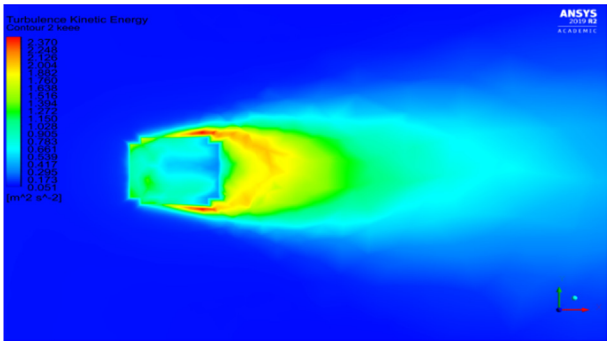


Fig. 48. KE spectrum for 18.75 mm cut section.

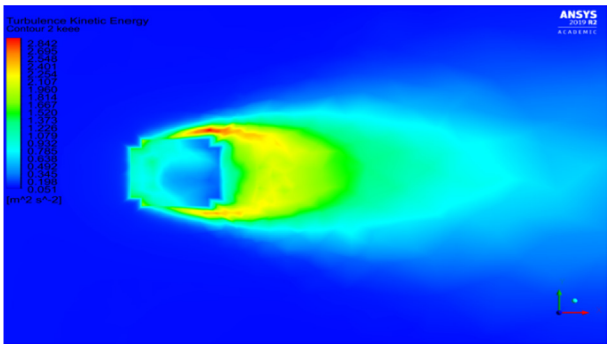


Fig. 49. KE spectrum for 25 mm cut section.

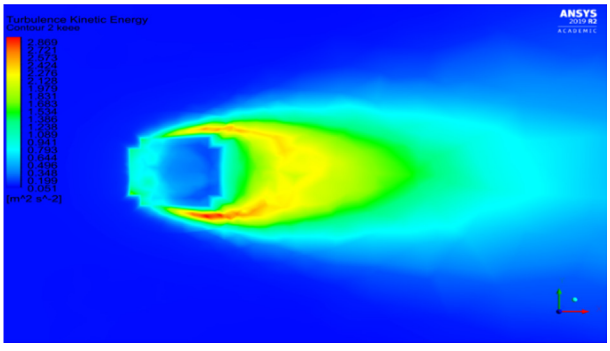


Fig. 50. KE spectrum for 31.25 mm cut section.

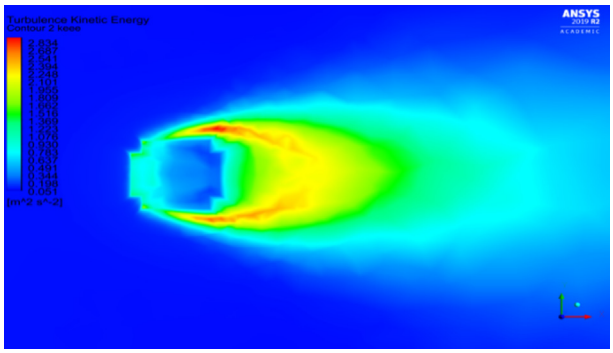


Fig. 51. KE spectrum for 37.5 mm cut section.

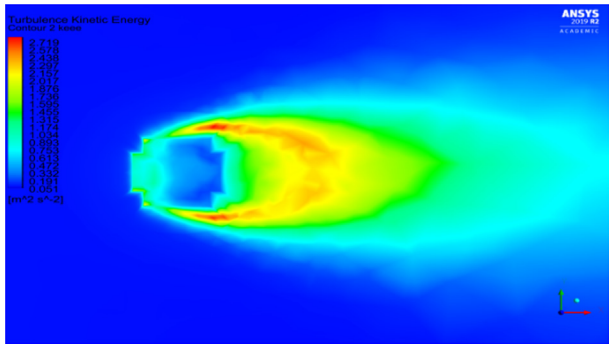


Fig. 52. KE spectrum for 43.75 mm cut section.

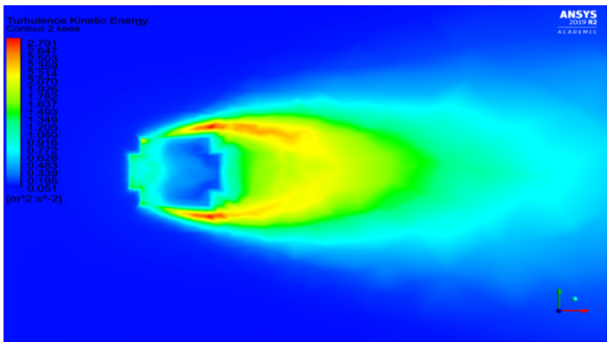


Fig. 53. KE spectrum for 50 mm cut section.

5. Discussion and conclusions

In this study, the prediction of along wind components with numerical simulations has shown great accuracy. The findings of this study have led to a comprehensive understanding of the aerodynamic and response characteristics of super-tall buildings with various cross-sections. It has been observed that the computational fluid dynamics simulations in the numerical model agree with the experimental results. The results are in better agreement on the front face and across wind responses but show a slight discrepancy in lift forces. Some modifications in the fluid domain and meshing have to be performed in more complex models to generate more convergent results. The k-ε turbulence model has shown great performance in numerical simulations. This study suggests that numerical simulations can be reliable for more complex wind simulations. This evaluation suggests the suitability of low-dimensional steady computational fluid dynamics modeling for performance assessment. The case study for corner modifications suggests the following

- Minor modifications like corner-cutting may lead to reductions in drag force and moment by 25% and 20% respectively.
- Corner modification showed better performance upto 12–15% modifications in section. More than 22% reduction leads to aerodynamic instability of the structure.
- Streamline study and pressure field of the models show suction near the corner cut the walls which may be responsible for drag force reductions.
- The reattachment of separated flow in corner cut models may result in a more stable wake region than the square model.
- The introduction of corner-cutting (chamfer and fillet) in building plan can be effective in aerodynamic mitigation but requires a complete understanding of the building aerodynamics and wind responses.

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.

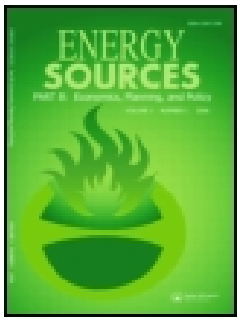
References

[1] Davenport. The response of super tall buildings to wind, second century of the skyscraper. Council of tall buildings and urban habitat; 1988. p. 705–25.
[2] Tamura T, Miyagi T. The effect of turbulence on aerodynamic forces on a square cylinder with various corner shapes. J Wind Eng Ind Aerodyn 1999;83:135–45.
[3] Gu M, Quan Y. Across-wind loads of typical tall buildings. J Wind Eng Ind Aerodyn 2004;92(13):1147–65. doi: https://doi.org/10.1016/j.jweia.2004.06.004.
[4] Tse KT, Hitchcock PA, Kwok KCS, Thepmongkorn S, Chan CM. Economic perspectives of aerodynamic treatments of square tall buildings. J Wind Eng Ind Aerodyn 2009;97:455–67.
[5] Virote Boonyapinyo, Pongsakorn Wangkansirikun. Aerodynamic modifications of high-rise buildings for wind load and response reductions. In: The 2016 international conference on advances in wind and structures (AWAS16).
[6] Kwok KCS, Isyumov N. Aerodynamic measures to reduce the wind-induced response of build-ings and structures. In: Proceedings of structural engineers world congress, San Francisco, CD-ROM: T179-6; 1998.
[7] Carassale L, Freda A, Marrè-Brunenghi M. Experimental investigation on the aerodynamic behavior of square cylinders with rounded corners. J Fluids Struct 2014;44:195–204.
[8] Kwok, Kenny CS, Bailey Peter A. Aerodynamic devices for tall buildings and structures. J Eng Mech 1987;113(3):349–65. doi: https://doi.org/10.1061/(asce)0733-9399(1987)113:3(349).
[9] Kawai H. Effect of corner modifications on aeroelastic instabilities of tall buildings. J Wind Eng Ind Aerodyn 1998;74–76:719–29. doi: https://doi.org/10.1016/s0167-6105(98)00065-8.
[10] Kwok KCS et al. Effect of edge configuration on wind-induced response of tall buildings. Eng. Struct. 1988;10(2):135–40. doi: https://doi.org/10.1016/0141-0296(88)90039-9.
[11] Jamieson NJ et al. Wind induced external pressures on a tall building with various corner configurations. J Wind Eng Ind Aerodyn 1992;15(1–3):2401–12. doi: https://doi.org/10.1016/0167-6105(92)90032-6.
[12] Nakamura Y et al. Galloping of rectangular cylinders in the presence of a splitter plate. J Fluids Struct 1991;5(5):521–49. doi: https://doi.org/10.1016/s0889-9746(05)80004-0.
[13] Li Y, Tian X, Tee KF, Li Q-S, Li Y-G. Aerodynamic treatments for reduction of wind loads on high-rise buildings. J Wind Eng Ind Aerodyn 2018;172:107–15. doi: https://doi.org/10.1016/j.jweia.2017.11.006.
[14] Elshaer A, Bitsuamlak GT, El Damatty A. Wind Load Reductions due to Building Corner Modifications. 22nd annual conference of the CFD Society of Canada. Toronto, Canada, 2014.
[15] Lee BE. The effect of turbulence on the surface pressure field of a square prism. J Fluid Mech 1975;69(2):263–82. doi: https://doi.org/10.1017/s0022112075001437.
[16] Otsuki Y, Fujii K, Washizu K, Ohya A. Wind tunnel experiments on aerodynamic forces and pressure distribution of rectangular cylinders in a uniform flow. In: Fifth symposium on wind effects on structures. p. 169–76.
[17] Ravindra HV, Krishnegowda YT, Gangadara Gowda CJ, Vikram CK, Manu. Numerical simulation of flow past two square cylinders of different perimeters. In: International conference on emerging trends in engineering (ICETE-2011), NMAM Institute of Technology, Nitte; May 4–5, 2011.
[18] David Malott, Zhizhe Yu KPF, Dennis Poon, Torsten Gottlebe, Thornton Tomasetti, Ping an finance center: pioneering china's tallest – efficiencies of form and structures, Sep 2012 – CTBUH 2012 9th world congress, Shanghai.
[19] Franke J, Hirsch C, Jensen AG, Krüs HW, Schatzmann M, Westbury PS, et al. Recommendations on the use of CFD in wind engineering. In: Proceedings of the international conference on Urban wind engineering and building aerodynamics: COST C14: impact of wind and storm on city life and built environment. Rhode, Saint-Genèse; 2004.
[20] Zhai Zhiqiang, Zhang Zhao, Zhang Wei, Chen Qingyan. Evaluation of various turbulence models in predicting airflow and turbulence in enclosed environments by CFD: Part 1—Summary of prevalent turbulence models; 2007.
[21] Shirzadi Mohammadreza, Mirzaei Parham, Naghashzadegan Mohammad. Improvement of k-epsilon turbulence model for CFD simulation of atmospheric boundary layer around a high-rise building using stochastic optimization and Monte Carlo Sampling technique. J Wind Eng Ind Aerodyn 2017;171. doi: https://doi.org/10.1016/j.jweia.2017.10.005.
[22] Bhattacharyya Biswarup, Dalui Sujit, Ahuja Ashok. Wind induced pressure on 'E' plan shaped tall buildings. Jordan J Civil Eng 2014;8:120–34.
[23] Ahirwar Ritu. Experimental investigation of the effect of cross-sectional shape on wind loads on tall buildings; 2012.
[24] IS-875-(part 3)-1987-Wind Loads Bureau of Indian Standards, Manak Bhawan, New Delhi.
[25] Mori Y, Hishida K, Maeda M. Buoyancy effects on the wake behind a heated obstacle 619 immersed in a turbulent boundary layer. Int J Heat Fluid Flow 1995;16:405–16.
[26] Deepakkumar R, Jayavel S, Tiwari S. A comparative study on effect of plain- and wavy-wall confinement on wake characteristics of flow past circular cylinder. Sādhanā 2017;42:963–80. doi: https://doi.org/10.1007/s12046-017-0649-1.
[27] Dey P, Das AKR. A numerical study on effect of corner radius and Reynolds number on fluid flow over a square cylinder. Sādhanā 2017;42:1155–65. doi: https://doi.org/10.1007/s12046-017-0680-2.
[28] Cao Shuyang, Tamura Yukio, Kikuchi Naoshi, Saito Mamoru, Nakayama Ikuo, Matsuzaki Yutaka. Wind characteristics of a strong typhoon. J Wind Eng Ind Aerodyn 2009;97:11–21. doi: https://doi.org/10.1016/j.jweia.2008.10.002.
[29] Carassale Luigi, Solari Giovanni. Monte Carlo simulation of wind velocity fields on complex structures. J Wind Eng Ind Aerodyn 2006;94:323–39. doi: https://doi.org/10.1016/j.jweia.2006.01.004.

[30] Haitham Aboshosha, Ahmed Elshaer, Bitsuamlak G, Ashraf El Damatty. J Wind Eng Ind Aerodyn 142 2015. doi: https://doi.org/10.1016/j.jweia.2015.04.004.
[31] Smirnov, Shi, Stacey, Celik, Ismail. Random flow generation technique for large eddy simulations and particle-dynamics modeling. J Fluids Eng 2001;123:359. doi: https://doi.org/10.1115/1.1369598.
[32] Xu YL, Chen Jun. Characterizing nonstationary wind speed using empirical mode decomposition. J Struct Eng-ASCE 2004;130. doi: https://doi.org/10.1061/(ASCE)0733-9445(2004)130:6(912).
[33] Tao T, Wang H. Modelling of longitudinal evolutionary power spectral density of typhoon winds considering high-frequency subrange. J Wind Eng Ind Aerodyn 2019;193:103957. doi: https://doi.org/10.1016/j.jweia.2019.103957.
[34] Tao Tianyou, Xu You-Lin, Huang Zifeng, Zhan Sheng, Wang Hao. Buffeting analysis of long-span bridges under typhoon winds with time-varying spectra and coherences. J Struct Eng 2020;146:04020255. doi: https://doi.org/10.1061/(ASCE)ST.1943-541X.0002835.
[35] Tao Tianyou, Shi Peng, Wang Hao. Spectral modelling of typhoon winds considering nexus between longitudinal and lateral components. Renewable Energy 2020;162:2019–30. doi: https://doi.org/10.1016/j.renene.2020.09.130.
[36] Huang Guoqing, Zheng Haitao, Xu You-lin, Li Yongle. Spectrum models for nonstationary extreme winds. J Struct Eng 2015;141:04015010. doi: https://doi.org/10.1061/(ASCE)ST.1943-541X.0001257.
[37] Tao Tianyou, Wang Zhi, Hu Liang, Kareem Ahsan. Error analysis of multivariate wind field simulated by interpolation-enhanced spectral representation method. J Eng Mech 2020;146:04020049. doi: https://doi.org/10.1061/(ASCE)EM.1943-7889.0001783.
[38] Tao Tianyou, Wang Zhi, Liang Hu, Zhao Kaiyong. Efficient simulation of fully non-stationary random wind field based on reduced 2D hermite interpolation. Mech Syst Signal Process 2020;150. doi: https://doi.org/10.1016/j.ymssp.2020.107265. ISSN 0888-3270.

Nikhil Gaur is currently working as a design engineer. He has received his B.E from the National Institute of Technology, Srinagar, India in the field of Civil Engineering in 2016 and his M.E in the field of Structural Engineering from Delhi Technological University in 2020. His research interests include Wind effects on tall buildings and CFD. He has published 2 research paper in journals and conferences.

Dr. Ritu Raj is Assistant Professor in Civil Engineering Department, DTU. He has published more than 50 technical papers in reputed journal and International conferences. He has supervised no. of Students at the level of graduates, post-graduates, and doctorates. His research interest includes behavior of wind load on the response of tall building, Aerodynamics Study, aero-elastic model study inside BLWT, Behavior of high strength concrete with and without fiber, earthquake engineering. He has more than 7.5 years of graduate and post-graduate teaching experience. Teaching Areas: Structure Engineering, Strength of Materials, Steel Structures, Design of Tall Buildings.



Energy Sources, Part B: Economics, Planning, and Policy

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uesb20

An efficient Jaya algorithm with Powell's Pattern Search for optimal power flow incorporating distributed generation

Saket Gupta, Narendra Kumar & Laxmi Srivastava

To cite this article: Saket Gupta, Narendra Kumar & Laxmi Srivastava (2021): An efficient Jaya algorithm with Powell's Pattern Search for optimal power flow incorporating distributed generation, Energy Sources, Part B: Economics, Planning, and Policy, DOI: 10.1080/15567249.2021.1942595

To link to this article: https://doi.org/10.1080/15567249.2021.1942595

Published online: 29 Jun 2021.

Submit your article to this journal

Article views: 23

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=uesb20

ENERGY SOURCES, PART B: ECONOMICS, PLANNING, AND POLICY
https://doi.org/10.1080/15567249.2021.1942595

Taylor & Francis
Taylor & Francis Group

Check for updates

An efficient Jaya algorithm with Powell's Pattern Search for optimal power flow incorporating distributed generation

Saket Gupta [a], Narendra Kumar [a], and Laxmi Srivastava [b]

[a]Electrical Engineering Department, Delhi Technological University, Delhi, India; [b]Department of Electrical Engineering, Madhav Institute of Technology and Science, Gwalior, India

ABSTRACT

Optimal power flow solutions are required for the cost-effective operation and control of an existing power system as well as for its future expansion planning. In this paper, a hybrid meta-heuristic Jaya-Powell's Pattern Search method is proposed to solve the optimal power flow problem integrated with distributed generating units. Powell's Pattern Search method has been incorporated into the Jaya algorithm in three different ways, resulting in three variants namely; Jaya-PPS1, Jaya-PPS2, and Jaya-PPS3. Performance of these variants are evaluated by solving optimal power flow in IEEE 30-bus, IEEE 57-bus, and IEEE 118-bus systems. In all the cases, Jaya-PPS1 is found to provide a lower objective function value with a smaller deviation over multiple runs than other algorithms. For example, the combined objective function found by Jaya-PPS1 for IEEE 30-bus system is 937.340, with a reduction of 2.86% of the original system. This benefit increases further with system size.

KEYWORDS

Distributed generation; hybrid Jaya-PPS algorithm; Jaya algorithm; meta-heuristic; optimal power flow; PPS

1. Introduction

At present, the continuously increasing global energy demand is currently unable to be met by centralized generation. Approximately 16% of the world's people still do not have access to electricity (Quadri Bhowmick and Joshi 2018). Therefore, the growing inclination of penetration of distributed generation (DG) units in inter-connected restructured power systems has increased the importance of solving Optimum Power flow (OPF) many times. The efficient planning for a power system to utilize the best possible use of existing capacity is of prime importance and key input to the process of economic development (Chattopadhyay 1994). OPF results are needed for planning, economic operation, and control of existing electrical power systems as well as for its future expansion planning. In a given electrical network, the OPF solution is required to regulate the control or decision variables setting in the feasible region that optimizes pre-defined objective functions (Abido 2002). OPF is a highly non-linear, large-dimensional, multi-modal, and non-convex problem having discrete and continuous control variables (Bouchekara et al. 2014). Carpentier introduced a first-time OPF problem based on economic load dispatch in the early 1960s (Carpentier 1979).

The optimal placement of DG has a major impact on the reliability of power supply, operational cost, voltage profile, power loss, environmental pollution, and voltage stability (Khaled, Eltamaly, and Beroual 2017). The inclusion of different kinds of renewable energy sources into a power system has resulted in improvement of the system voltage profile and uniform distribution of line flows and in making the system robust and more stable (Singh and Parida 2016). With the increase of power injection from DGs into a power network, it is equally important to find out the optimum power generation and optimal setting of other control parameters to minimize fuel cost, emission cost, and real power loss and to improve voltage profile (Elattar and ElSayed 2019). Congestion of transmission

CONTACT Saket Gupta sguptamits@gmail.com Department of Electrical Engineering, Delhi Technological University, Delhi 110042, India

© 2021 Taylor & Francis Group, LLC

system creates obstacles to meet demands. In such cases, integration of an optimum distributed generation unit is required to eliminate congestion on transmission lines of the power system (Peesapati, Yadav, and Kumar 2018). Stavrakas, Papadelis, and Flamos (2019) presented an Agent-based Technology adoption Model (ATOM). The proposed model is used for agent-based modeling policy assessment and can be a valuable tool in an effective policy design. Integrating and absorbing expanded shares of non-dispatchable renewable energy sources without compromising the power system's security and reliability is a primary task of such transformations (Stavrakas. and Flamos 2020). Maximization of the social welfare function under power balance and transmission congestion constraints is part of the congestion management problem formulation. Maximizing the social welfare role would help in maximization of consumers' benefits and minimization of the cost of generation of suppliers and distributed generators (Singh et al. 2014).

Conventional optimization techniques (Momoh, Adapa, and Hawary 1999) such as the non-linear programming, quadratic programming, gradient process, interior point, and so on have been used to solve the OPF problem, but these approaches have disadvantages such as the requirement of an initial point and insecure convergence characteristics (Sakthivel, Suman, and Sathya 2020), etc. Moreover, because of the non-differential, non-linear, multi-modal, and non-convex properties of the OPF problem, most of the conventional techniques sometimes converge to a local optimum (Srivastava and Singh 2015).

Continuous research in the soft computing area has led to the development of many meta-heuristic algorithms, which offer a choice to handle the drawbacks of conventional methods. Some of these are: (Abido 2002) used Tabu Search Algorithm for solving OPF problem. The proposed algorithm is applied IEEE 30-bus test system taking into account practical issue of generator units like valve loading effects. Bouchekara and Abido 2014, proposed Differential Search Algorithm to solve the OPF problem. The conventional IEEE 30-bus and IEEE 118-bus test systems were used to demonstrate the effectiveness of the proposed algorithm to solve OPF problems. (Bouchekara et al. 2014) have applied League Championship Algorithm (LCA) to solve the OPF problem: a case study of the Algerian power system. The Algerian 59-bus system is a practical power system used to demonstrate the validity and scalability of the proposed LCA to solve the OPF problem. (Mohamed et al. 2017) solved the OPF problem using the Moth swarm algorithm (MSA). The proposed algorithm was applied in IEEE 30-bus, IEEE 57-bus, and IEEE 118-bus test systems to solve the OPF problem considering generator constraints such as valve point effect and multi-fuel type of generation units, and many others. In addition to the application of evolutionary computing-based single-objective and multi-objective optimization algorithms, hybrid meta-heuristic algorithms are also being applied for various electrical engineering problems including OPF.

Ghasemi et al. (2014) proposed a multi-objective modified imperialist competitive algorithm for solving multi-objective optimal power flow problems considering four objectives: fuel cost, emission cost, real power loss, and total voltage deviation. Naderi, Kasmaei, and Abdi (2019) presented optimal power flow problems integrated with FACTS devices using fuzzy-based improved comprehensive-learning PSO algorithm. Khelifi, Bentouati, and Chettih (2019) proposed a hybrid firefly krill herd method to solve the OPF problem. Azizivahed et al. (2020) suggested a dynamic economic emission dispatch problem using a hybrid gray wolf optimizer-particle swarm optimization algorithm. Biswas et al. (2020) applied a multi-objective evolutionary algorithm based on decomposition (MOEA/D) to solve the multi-objective OPF problem. Naderi et al. (2021) presented the solution of single- and multi-objective optimal power flow problems using self-adaptive particle swarm optimization (SPSO) and differential evolution algorithms.

Meta-heuristic algorithms are random search algorithms and require parameter tuning to find the optimal solution. Also, they need several trials and long computational time to obtain the best optimal solution. Hence, the motivation behind this paper is to apply an algorithm-specific parameter-less optimization algorithm namely, Jaya algorithm, and to avoid its premature convergence problem (Elattar and ElSayed 2019) by incorporating Powell's Pattern Search method. This resulted in a hybrid

Jaya-PPS algorithm. Hybridization of the Jaya and PPS algorithms was carried out in different ways to find better results for real-world complex, constrained, and high dimensional optimization problems.

The paper's contribution can be summed up as follows:

- The main contribution of this paper is to implement hybridization of two algorithms (Jaya and Powell's Pattern Search) in different manners and at different levels to find the best option for hybridization.
- Powell's Pattern Search method has been incorporated into Jaya algorithm in three different ways, resulting in three variants namely; Jaya-PPS1, Jaya-PPS2, and Jaya-PPS3.
- The proposed hybrid Jaya and Powell's Pattern Search method utilizes the exploration property of the Jaya algorithm and the exploitation quality of Powell's Pattern Search method.
- Effectiveness of these variants is demonstrated by solving optimal power flow problem with and without incorporation of Distributed Generating units for minimization of fuel cost, emission cost, real power loss, and total voltage deviation simultaneously using Jaya-PPS1, Jaya-PPS2, Jaya-PPS3, Dragonfly Algorithm, Grey Wolf Optimization and Jaya algorithms in standard IEEE 30-bus, and IEEE 57-bus test systems.
- These three variants have shown significantly improved performance as compared to the Jaya algorithm. Based on statistical analysis, the proposed hybrid Jaya-PPS1 algorithm, particularly, is found to be the most efficient (in terms of quality of solution) and robust over Dragonfly Algorithm, Grey Wolf Optimization, Jaya, Jaya-PPS2, Jaya-PPS3 algorithms, and the reported results.
- It is observed that Jaya-PPS1 provided the lowest value of the combined objective function consistently in all the four cases related to IEEE 30-bus and IEEE 57-bus systems. Also, this variant provided the least standard deviation over multiple runs as compared to other meta-heuristic algorithms.

This paper is organized as follows: Section 2 presents the formulation of the OPF problem. The proposed hybrid Jaya-PPS algorithm is discussed in Section 3. Section 4 includes the results and discussion, while statistical analysis has been shown in Section 5. Conclusions are presented in Section 6.

2. Problem formulation

The objective functions of the optimal power flow problem and the constraints considered in this work are formulated as follows:

2.1. OPF objective functions

In optimal power flow, four objective functions namely, minimization of fuel cost, minimization of emission cost, minimization of real power losses, and improvement of voltage profile by minimizing the total voltage deviation are considered in this paper.

2.1.1. Fuel cost minimization

The main aim of this objective function is to minimize overall production/fuel cost. Consequently, it may be defined as (Nguyen 2019):

$$Z_{FCM}(u,x) = \sum_{i=1}^{NGN} f(P_{g_i})\left(\$/h\right) = \sum_{i=1}^{NGN} A_i + B_i P_{g_i} + C_i P_{g_i}^2 \left(\$/h\right) \quad (1)$$

where A_i, B_i, and C_i are the quadratic fuel cost coefficients of the i^{th} generating unit and P_{g_i} is the active power output of i^{th} generating unit

The control variables (u) include the generator active power output (P_g) except at slack bus, generator bus voltage (V_g), tap-setting of the transformer (T_{TR}), and shunt *VAR* compensation (Q_c). The state variables (x) consists of slack bus active power output $P_{g_{Slack}}$, Load bus voltage (V_L), generator reactive power output (Q_g), and power flow in transmission lines (S_{tl}). The control variables and state variables vectors can be defined as follows:

$$u = \left[P_{g_2} \ldots P_{g_{NGN}}, V_{g_1} \ldots V_{g_{NGN}}, Q_{C_1} \ldots Q_{C_{NC}}, T_1 \ldots T_{NTR}\right] \tag{2}$$

$$x = \left[P_{g_{Slack}}, V_1 \ldots V_{NLB}, Q_{g1}, \ldots Q_{gNGN}, S_1, \ldots S_{Ntl}\right] \tag{3}$$

where *NGN, NC*, and *NTR* are the number of generating units, number of *VAR* compensation units, and number of regulating transformers, respectively. *NLB* is the number of load buses and *Ntl* is the number of transmission lines.

2.1.2. Total voltage deviation

The aim of the second objective function is the minimization of voltage variation at all the load buses from 1.0 pu and this can be described mathematically as follows (Nguyen 2019):

$$Z_{VPI}(u, x) = \sum_{i \in NLB} |V_i - 1| \tag{4}$$

where V_i is the voltage magnitude at i^{th} load bus.

2.1.3. Emission cost minimization

This objective function can be formulated as follows (Nguyen 2019):

$$Z_{ECM}(u, x) = \sum_{i=1}^{NGN} \alpha_i + \beta_i P_{g_i} + \gamma_i P_{g_i}^2 + \xi_i \exp\left(\lambda_i P_{g_i}\right) \tag{5}$$

where $\alpha_i, \beta_i, \gamma_i, \xi_i, \lambda_i$ are the emission coefficients of i^{th} generating unit.

2.1.4. Real power losses minimization

The real power losses of an electrical network can be computed using the following expression (Nguyen 2019):

$$Z_{Loss}(u, x) = \sum_{i=1}^{NB} P_{gi} - \sum_{i=1}^{NB} P_{di} \tag{6}$$

Where *NB* is the no. of buses, P_{gi} is the active power generation at *the* i^{th} generating unit and P_{di} is the real power load at the i^{th} bus.

2.2. Combined Objective Function (COF)

The multi-objective function which consists of four objective functions; minimization of fuel cost, emission cost, real power losses, and total voltage deviation is transformed into a single-objective function by using weighing factors as given below:

$$Z_{COF}(u, x) = Z_{FCM} + w_{ECM} \times Z_{ECM} + w_{RPLM} \times Z_{RPLM} + w_{TVDM} \times Z_{TVDM} \tag{7}$$

Where, w_{ECM}, w_{RPLM}andw_{TVDM} are weighing factors (Khelifi, Bentouati, and Chettih 2019).

2.3. Constraints

OPF problem consists of two types of constraints namely, equality and inequality. The equality constraints consist of power flow equations. The power flow equations are as follows:

$$\sum_{i=1}^{NB}(P_{gi} + P_{DGi}) = \sum_{i=1}^{NB} P_{di} + P_{Loss} \tag{8}$$

$$\sum_{i=1}^{NB}(Q_{gi} + Q_{DGi}) = \sum_{i=1}^{NB} Q_{di} + Q_{Loss} \tag{9}$$

Where NB is the no. of buses, P_{g_i}, Q_{g_i} are the active and reactive power output of i^{th} generating unit, P_{DGi}, Q_{DGi} are the active and reactive power output of i^{th} DG unit, P_{di}, Q_{di} are active and reactive power load demand at i^{th} bus, P_{Loss}, Q_{Loss} is the total active and reactive power losses occurring in the lines of the power system.

Inequality constraints are the operating constraints of a power system as described below:

- Generator Constraints:

The voltage V_g, active power P_g and reactive power output Q_g should be regulated by their lower and upper limits for all the generators including slack bus generator;

$$V_{g_k}^{min} \le V_{g_k} \le V_{g_k}^{max} k = 1 \ldots..NGN \tag{10}$$

$$P_{g_k}^{min} \le P_{g_k} \le P_{g_k}^{max} k = 1 \ldots.NGN \tag{11}$$

$$Q_{g_k}^{min} \le \ Q_{g_k} \le Q_{g_k}^{max} k = 1 \ldots\ldots NGN \tag{12}$$

- Shunt VAR compensator constraints:

Controllable VAR sources (Q_c) are maintained within the maximum and minimum limits;

$$Q_{C_k}^{min} \le Q_{C_k} \le Q_{C_k}^{max} k = 1 \ldots NC \tag{13}$$

- Transformer Constraints:

Transformer taps-settings (T_k) are regulated to their lowest and highest limits:

$$T_k^{min} \ \le T_k \le T_k^{max} k = 1 \ldots\ldots NTR \tag{14}$$

- Security Constraints:

The voltage of load buses (V_{L_k}) and power flow in transmission lines (S_{l_k}) should vary within their minimum and maximum limits

$$V_{L_k}^{min} \le V_{L_k} \le V_{L_k}^{max} k = 1 \ldots\ldots NLB \tag{15}$$

$$S_{l_k} \le S_{l_k}^{max} k = 1 \ldots\ldots Ntl \tag{16}$$

- Distributed Generation Constraints

The active (P_{DGk}) and reactive powers (Q_{DGk}) of DG units should lie in between their minimum and maximum limits.

$$P_{DGk}^{min} \leq P_{DGk} \leq P_{DGk}^{max} k = 1 \ldots . NG \tag{17}$$

$$Q_{DGk}^{min} \leq Q_{DGk} \leq Q_{DGk}^{max} k = 1 \ldots . NG \tag{18}$$

2.4. Constraints handling

The working limits of operating constraints, such as shunt *VAR* constraints, bus voltage constraints, line flows, transformer constraints, and generator constraints are managed effectively in OPF problem by implementing the Penalty Factor approach. In the Penalty Factor approach, any violation is penalized heavily by multiplying the violation by a large number, so that infeasible solutions are discarded and only the feasible solutions are considered. The inequalities are incorporated into the objective function to find a feasible solution and thus, the extended objective function obtained can be defined as:

$$\begin{aligned} Z_{aug} &= Z_{COF}(u, x) + R_1 \left(P_{g_{Slack}} - P_{g_{Slack}}^{lim} \right)^2 + R_2 \sum_{i=1}^{NGN} \left(Q_{g_k} - Q_{g_k}^{lim} \right)^2 + R_3 \sum_{i=1}^{NLB} \left(V_{L_k} - V_{L_k}^{lim} \right)^2 \\ &\quad + R_4 \sum_{i=1}^{Ntl} \left(S_{l_i} - S_{l_i}^{lim} \right)^2 \end{aligned} \tag{19}$$

Where R_1,R_2, R_3, and R_4 represent penalty factors (10^5) corresponding to limit violations.

3. Jaya algorithm

Jaya algorithm, recently developed by Rao (Rao 2016) is an efficient meta-heuristic optimization algorithm. The Jaya algorithm is an algorithm-specific parameter-less algorithm like the TLBO algorithm. The main advantage of the Jaya algorithm over most of the other meta-heuristic algorithms is that algorithm operations do not require algorithm-specific parameters tuning other than population size and maximum iterations. Hence, it is simple to implement this algorithm for solving various kinds of optimization problems.

3.1. Maximized (or minimized) value of an objective function

Initial population 'p' is randomly generated within the upper and lower limits of the control variables and is updated as per Eq. (20). The best and worst solutions are determined based on the fitness values of the objective function.

Let, the number of design variable is 'm' (i.e., j = 1, 2, 3 . . ., m) and 'n' is the population size (k = 1, 2 . . ., n). Let $J_{i,j,k}$ represents the value of the j^{th} variable for k^{th} candidate during the i^{th} iteration, then this value is modified as (20)

$$J_{i+1,j,k} = J_{i,j,k} + \gamma_{i,j,1} \left(J_{i,j,B} - abs(J_{i,j,k}) \right) - \gamma_{i,j,2} \left(J_{i,j,W} - abs(J_{i,j,k}) \right) \tag{20}$$

Where, $J_{i,j,W}$ and $J_{i,j,B}$ are the worst candidate and best candidate value of variable j^{th} respectively. The updated value of $J_{i,j,k}$ is $J_{i+1,j,k}$, and throughout the i^{th} iteration, $\gamma_{i,j,1}$and $\gamma_{i,j,2}$ are the two random numbers between 0 and 1, for the j^{th} variable.

3.2. Powell's pattern search

Powell's pattern search (PPS) was proposed by Powell in 1962. This method is an expansion of the basic Pattern Search method. It is based on the conjugate direction method and is a derivative-free

optimization technique. PPS with meta-heuristic algorithm offers a flexible, balanced operator to enhance local search capability in contrast to other meta-heuristic algorithms. The following is the summary of the PPS algorithm underlying mechanism (Narang, Dhillon, and Kothari 2012; Rao 2019).

The search direction for l^{th} coordinate for g^{th} dimension of the 'n' dimension search space can be defined as:

$$S_g^l = \begin{cases} 1; g = l \\ 0; g \neq l \end{cases} (g = 1,\ 2, ..n; l = 1,\ 2, ..n) \tag{21}$$

The step length λ_g^* for g^{th} decision variable can be determined as:

$$\lambda_g^* = \lambda_g^{\min} + \text{rand} \times \left(\lambda_g^{\max} - \lambda_g^{\min}\right) (g = 1,\ 2, \ldots n) \tag{22}$$

Here, λ_g^{min}, λ_g^{max} is the minimum and maximum step length for g^{th} decision variable, respectively.

The decision variable (X_g) is modified once along the coordinate direction (l) as:

$$X_g = X_g + \lambda_g^* \times S_g^l (g = 1,\ 2, \ldots n) \tag{23}$$

The vector of control variables is modified on the basis of minimum objective function value. For all 'n' coordinates, this process is continued. The pattern search direction is obtained for the next optimization cycle as:

$$S_g^l = X_g - Z_g (g = 1,\ 2, ..n; l = n + 1) \tag{24}$$

where Z_g is the initial value of the decision variable X_g.

Additionally, one of the coordinate's directions is discarded in the direction of pattern 'm' as:

$$S_g^m = S_g^l\ (g = 1,\ 2, ..n; l = n + 1) \tag{25}$$

The process is continued till the entire direction of the coordinate is discarded and the entire operations restart in one of the coordinate directions again. Finally, until Powell method has been reached maximum iterations, the process of updating continues.

3.3. Proposed hybrid Jaya-PPS algorithm

Jaya algorithm has a strong capability to explore the search space globally, but sometimes it suffers from the issue of premature convergence and it can be stuck simply in local optima (Elattar and ElSayed 2019). To overcome this problem and to make this algorithm more efficient, a hybrid Jaya algorithm that combines Jaya algorithm and PPS algorithm is proposed in this paper. The aim to incorporate PPS with Jaya is to combine the benefits of both algorithms. For the implementation of PPS algorithm, the standard Matlab toolbox was used taking the initial point as obtained by the Jaya algorithm at different levels in the three hybrid variants.

The integration of local search procedure (Powell's Pattern Search) into the classical Jaya algorithm has been carried out in three different ways, which resulted in three variants termed Jaya-PPS1, Jaya-PPS2, and Jaya-PPS3. To evaluate the performance of these variants, the common controlling parameters and the total number of function evaluations (NFE) used in the three variants are set to be the same as that of the classical Jaya algorithm. The NFE has been used as a reference to the check efficiency of various algorithms in this paper.

In the first strategy, that is in Jaya-PPS1, the Jaya algorithm and PPS algorithm have been applied sequentially in each iteration. The optimal setting of control variables as provided by the Jaya algorithm is used as the initial point setting for the PPS technique to acquire a better solution around this solution by applying the PPS algorithm for a pre-defined number of function evaluations. This better solution as provided by the PPS technique replaces the earlier one in the population of the Jaya

algorithm and is used for further modification in the next iteration by the Jaya algorithm. This procedure is repeated for a pre-specified number of iterations. The flow chart of the proposed hybrid Jaya-PPS algorithm is shown in Figure 1.

In the second strategy (Jaya-PPS2), the Jaya algorithm and PPS algorithm have been applied sequentially after exploiting the 80% problem-solving capability of the Jaya algorithm, that is, on the solution achieved by applying the Jaya algorithm for 80% iterations. In other words, the optimization process has been divided into two steps. In step 1 (for 80% of $Iter_{max}$), only the Jaya algorithm has been applied, while in step 2 (for the remaining 20% of iterations), Jaya and PPS algorithms both have been applied sequentially as in the case of Jaya-PPS1.

In the third strategy (Jaya-PPS3), the PPS algorithm has been applied considering its initial point as the solution offered by the Jaya algorithm after applying it for 90% iterations. In this case, also, the optimization process is a two-step process. In step 1 (first 90% of $Iter_{max}$), only the Jaya algorithm has been applied. However, in step 2 (remaining 10% of iterations), Jaya and PPS algorithms both have been applied sequentially with the optimal setting of control variables offered by the Jaya algorithm as the initial point for it.

3.3.1. Computational steps for hybrid Jaya-PPS algorithm

The computational steps of this algorithm are summarized as follows:

(i) Initialize the population having control variables and set maximum iteration count $Iter_{max}$ and number of function evaluations *PSFE* for PPS method.
(ii) Set iteration $Iter = 0$.
(iii) Identify the solutions which are worst and best in the population by observing the value of the extended objective function (19).
(iv) Modify the solutions based on the worst and the best solutions using (20).
(v) If the modified solution is found to be better than the previous solution, then move to step *vi*, otherwise move to step *vii*.
(vi) Replace the previous solution with the modified solution. Move to step *viii*.
(vii) Keep the previous solution.
(viii) Select the best solution found by the Jaya algorithm so far as the initial point for the PPS method and apply the PPS method for *PSFE* to attain a better solution.
(ix) Modify the population of the Jaya algorithm by replacing the best solution of the Jaya algorithm by that obtained by the PPS method.
(x) Increase the iteration number by 1, that is, $Iter = Iter + 1$.
(xi) If $Iter < Iter_{max}$, then go to step *iii*, otherwise go to step *xii*.
(xii) Stop. The optimal solution is achieved.

4. OPF results and discussion

Three standard test systems, IEEE 30-bus, IEEE 57-bus, and IEEE 118-bus systems are used to evaluate the effectiveness of the three versions of hybrid Jaya-PPS algorithm, Jaya algorithm, Dragonfly Algorithm, and Gray wolf optimization algorithm for solving OPF problems with and without integration of DG. The control variables' limits, line data, bus data along with their initial settings for the IEEE 30-bus system are taken from (Lee, Park, and Ortiz 1985), while the emission and fuel cost coefficients are taken from (Niknam et al. 2011). The combined objective function, COF is obtained by considering the weighting factors w_{ECM}, w_{RPLM}, and w_{TVDM} as 19, 22, and 21, respectively (Khelifi, Bentouati, and Chettih 2019).

The IEEE 30-bus system is modified by including DG based on Renewable Energy Technologies. The optimal location of the DG is selected based on the sensitivity of active power loss and generation cost to each active and reactive power (Elattar and ElSayed 2019) and it is bus no. 30. At this bus, the capacity selected for the type 1 DG unit is 5 MW, in this paper.

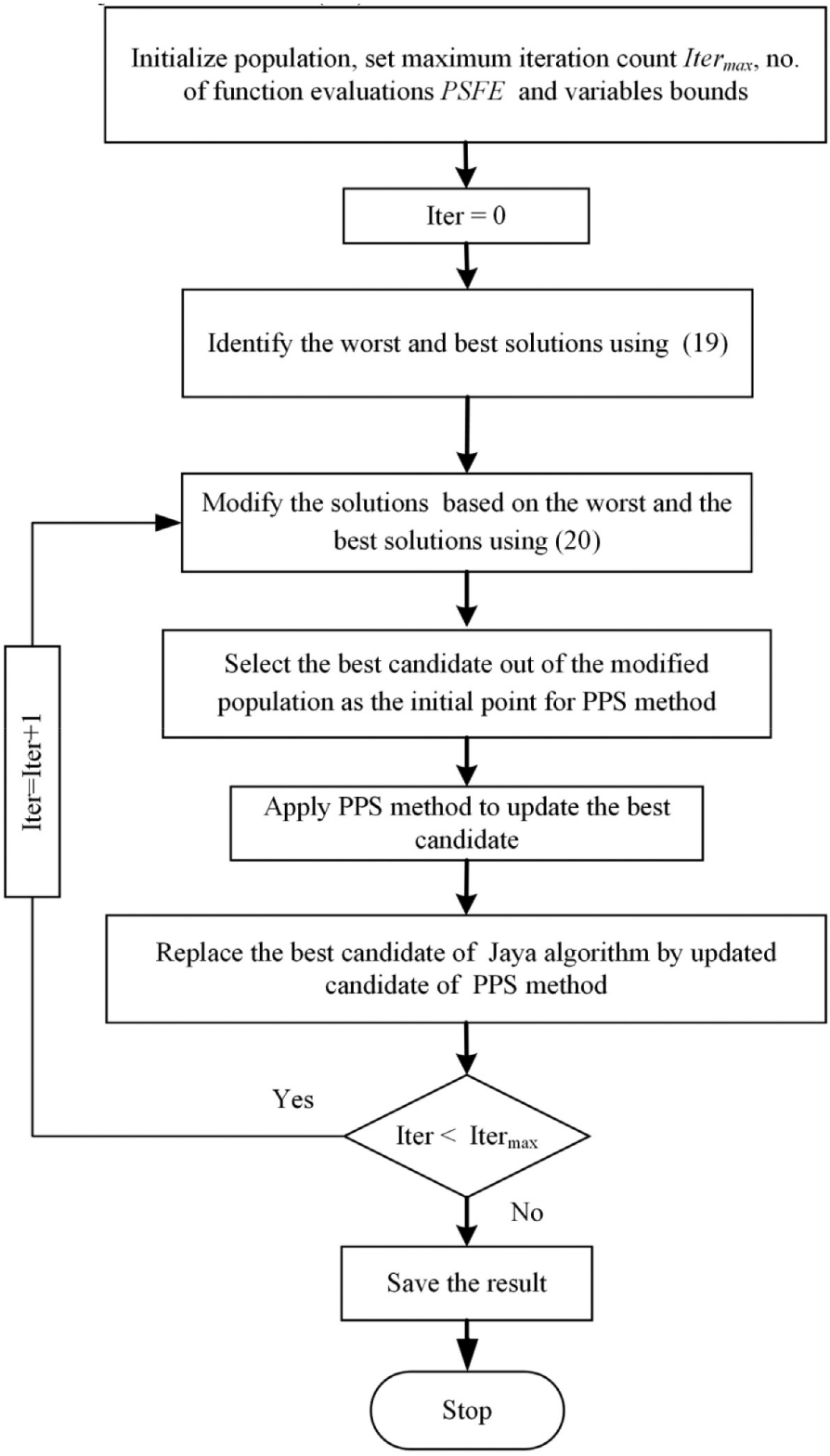


Figure 1. Flowchart of the proposed hybrid Jaya-PPS algorithm.

The IEEE 57-bus test system has 7 generating units and 80 branches. The limits for voltage magnitude at all the buses of the system are considered to be 0.94 pu and 1.06 pu. The limits for the tap changing transformers' tap settings are taken as 0.9 pu and 1.1pu. The generator coefficients, lower and upper limits of all the 33 control variables, and system data (bus data, line data) along with their initial settings are taken from (Freris. and Sasson 1968). The active and reactive power demands of this system are 12.508 pu and 3.364 pu, respectively, at 100 MVA base. In the case of the IEEE 57-bus

system, the combined objective function, COF is obtained by considering the weighting factors w_{ECM}, w_{RPLM}, and w_{TVDM} as 300, 30, and 600, respectively, in this paper. This system is modified by integrating two type 1 DGs optimally at bus nos. 35 and 36 with the capacities of 47.9067 MW and 47.2636 MW, respectively (Charles and Odero 2013).

To evaluate the scalability of proposed algorithms and to prove their efficacy to solve large-scale problems, all the three variants of Jaya-PPS algorithms, GWO, and DA algorithm were applied to solve the OPF problem in the IEEE 118-bus system. The system data, generator coefficients, lower and upper limits of all the 130 control variables along with their initial settings are taken from (Power system test cases, 0000). The active and reactive power demands of this test system are 42.42 pu and 14.38 pu, respectively, at 100 MVA base.

To demonstrate the effectiveness of the proposed algorithm, five cases are considered as given below:

Case 1: OPF no DG in IEEE 30-bus system
Case 2: OPF with DG in IEEE 30-bus system
Case 3: OPF no DG in IEEE 57-bus system
Case 4: OPF with DG in IEEE 57-bus system
Case 5: OPF no DG in IEEE 118-bus system

Various trials were carried out with different population sizes and no. of iterations. The best results thus achieved and reported in this paper are for Population size (pop) = 30 and Maximum no. of iterations ($Iter_{max}$) = 200 for IEEE 30-bus system and pop = 40 and $Iter_{max}$ = 300 for IEEE 57-bus and IEEE 118-bus systems. All these algorithms were developed using Matlab 13a version in a 3.6 GHz Intel Processor, 8GB RAM Core i7 and 64-bit operating personal computer.

To compare the performance of various algorithms, all the algorithms were run for the same number of function evaluations (NFE) which is equal to 6000 in the case of the IEEE 30-bus system and 12,000 in the case of IEEE 57-bus and IEEE 118-bus systems. Details of the implementation of various algorithms and inclusion of PPS in the three variants of hybrid Jaya-PPS algorithms are given in Table 1.

For various cases, 50 runs were taken using DA, GWO, Jaya, Jaya-PPS1, Jaya-PPS2, and Jaya-PPS3 algorithms to solve the OPF problem with the above-mentioned objective functions. The best results obtained out of the 50 runs are given here.

4.1. Case 1: OPF no DG in IEEE 30-bus system

The proposed hybrid Jaya-PPS algorithms, GWO algorithm (Mirjalili, Mirjalili, and Lewis 2014), Dragonfly algorithm (Meraihi et al. 2020), and Jaya algorithm (Rao 2016) were applied to solve the OPF problem for the combined objective function consisting of the fuel cost, emission cost, real power loss, and total voltage deviation. The results obtained using these methods along with optimal control variable settings are shown in Table 2. The simulation outcomes demonstrate that Jaya-PPS1 is superior to other techniques. Its combined objective function (964.962) is less than those achieved using other approaches without violation of the pre-specified constraints. Results of the hybrid Jaya-PPS algorithms are compared with DA, GWO, Jaya, and also with the reported results available in recent literature in Table 3, and it is found that the proposed Jaya-PPS1 algorithm provided the minimum value of the COF.

Convergence characteristics of DA, GWO, Jaya, Jaya-PPS1, Jaya-PPS2, and Jaya-PPS3 algorithms are shown in Figure 2a. As can be observed from Figure 2a, the Jaya-PPS1 algorithm has offered the best convergence characteristic. During the iterative process for these six algorithms, the trajectories of all the components of the COF (e.g., fuel cost, emission cost, real power loss, and total voltage deviation) are shown in Figure 2(b-g). As these figures show, the Jaya-PPS1 algorithm has given smooth convergence curves with no oscillations and fast convergence speed in comparison with other methods. Figure 3 shows the voltage profile provided by the proposed Jaya-

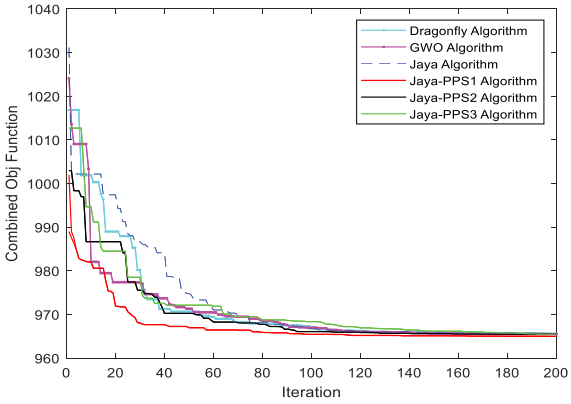
Table 1. Details of the DA, GWO, Jaya, Jaya-PPS1, Jaya-PPS2, and Jaya-PPS3 algorithms.

IEEE-30 bus system			
Algorithm	**Population**	**Iterations**	**Total NFE = 6000**
Dragonfly Algorithm	30	200	30 × 200
GWO Algorithm	30	200	30 × 200
Jaya Algorithm	30	200	30 × 200
Jaya-PPS1	30	200	(30JFE+30PSFE)×100
Jaya-PPS2	30	200	30JFE×160+(30JFE+30PSFE)×20
Jaya-PPS3	30	200	30JFE×180+(30JFE+30PSFE)×10
IEEE-57 bus & IEEE 118 bus system			
Algorithm	**Population**	**Iterations**	**Total NFE = 12000**
Dragonfly Algorithm	40	300	40 × 300
GWO Algorithm	40	300	40 × 300
Jaya Algorithm	40	300	40 × 300
Jaya-PPS1	40	300	(40JFE+40PSFE)×150
Jaya-PPS2	40	300	40JFE×240+(40JFE+40PSFE)×30
Jaya-PPS3	40	300	40JFE×270+(40JFE+40PSFE)×15

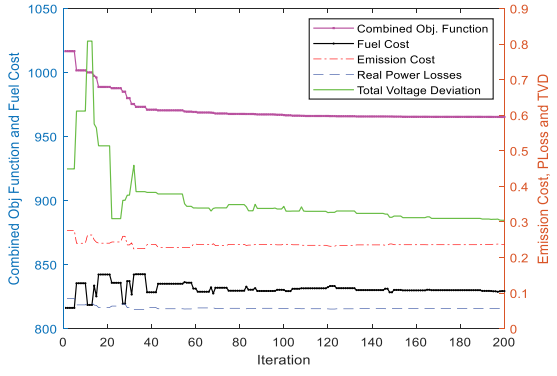
NFE = Number of function evaluations, JFE = Jaya Function Evaluations, PSFE = Pattern Search Function Evaluations.

Table 2. OPF results with optimum values of control variables for IEEE 30-bus system (Case 1).

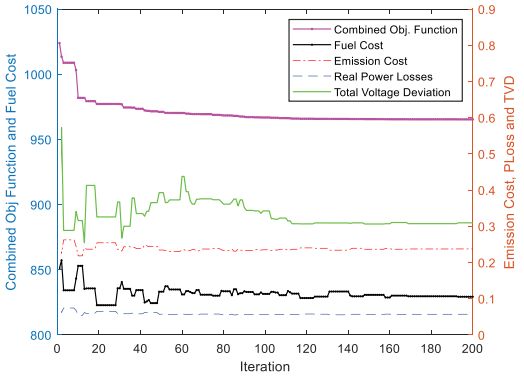
S. No.	Control Variable	DA	GWO	Jaya	Jaya-PPS1	Jaya-PPS2	Jaya-PPS3
Generator Real power output							
1	Pg_2	0.5265	0.5255	0.5167	0.5255	0.5279	0.5210
2	Pg_5	0.3114	0.3106	0.3221	0.3143	0.3146	0.3096
3	Pg_8	0.3500	0.3500	0.3497	0.3499	0.3500	0.3499
4	Pg_{11}	0.2577	0.2625	0.2726	0.2698	0.2750	0.2701
5	Pg_{13}	0.2167	0.2118	0.2071	0.2085	0.2035	0.2172
Generator voltage setting							
6	Vg_1	1.0742	1.0745	1.0728	1.0736	1.07302	1.0710
7	Vg_2	1.0597	1.0603	1.0590	1.0599	1.05934	1.0586
8	Vg_5	1.0312	1.0347	1.0337	1.0317	1.03193	1.0318
9	Vg_8	1.0414	1.0423	1.0415	1.0416	1.04188	1.0419
10	Vg_{11}	1.0545	1.0534	1.0501	1.0381	1.04865	1.0404
11	Vg_{13}	1.0160	1.0193	1.0273	1.0150	1.02283	1.0225
Transformer tap setting							
12	$T_{6\text{-}9}$	1.0677	1.0890	1.100	1.0998	1.0997	1.100
13	$T_{6\text{-}10}$	1.0140	0.9811	0.9483	0.9557	0.9459	0.9310
14	$T_{4\text{-}12}$	1.0216	1.0123	1.0258	1.0253	1.0272	1.0334
15	$T_{28\text{-}27}$	1.0018	1.0072	1.0034	1.0021	1.0046	1.0035
Shunt VAR source setting							
16	Qc_{10}	0.0496	0.0486	0.0000	0.0474	0.0038	0.0002
17	Qc_{12}	0.0002	0.0009	0.0005	0.0486	0.0445	0.0413
18	Qc_{15}	0.0363	0.0186	0.0496	0.0355	0.0003	0.0475
19	Qc_{17}	0.0487	0.0318	0.0500	0.0500	0.0499	0.0448
20	Qc_{20}	0.0499	0.0482	0.0498	0.0500	0.0464	0.0446
21	Qc_{21}	0.0500	0.0500	0.0499	0.0500	0.0492	0.0494
22	Qc_{23}	0.0489	0.0462	0.0173	0.0393	0.0498	0.0399
23	Qc_{24}	0.0497	0.0500	0.0498	0.0500	0.0470	0.0472
24	Qc_{29}	0.0253	0.0322	0.0303	0.0251	0.0290	0.0292
COF		965.351	965.302	965.286	**964.962**	965.267	965.255
Fuel Cost (FC)		829.358	829.239	831.549	**830.467**	830.850	830.290
Emission Cost (EC)		0.237	0.237	0.235	**0.236**	0.235	0.235
Real Power Loss (RPL)		5.685	5.684	5.578	**5.625**	5.616	5.642
Total Voltage Deviation (TVD)		0.304	0.309	0.311	**0.297**	0.304	0.302
Pgslack		122.838	123.021	122.147	122.183	121.891	122.250
L-Index (LI)		0.138	0.138	0.139	0.138	0.139	0.139



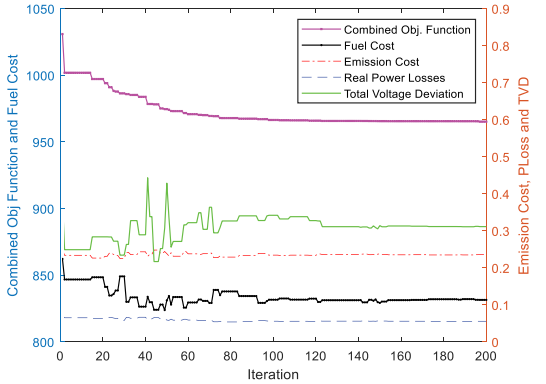
(a) Convergence Characteristics for various algorithms for Case 1



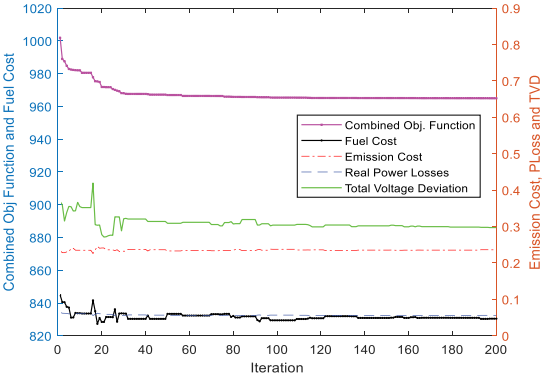
(b) Dragonfly Algorithm



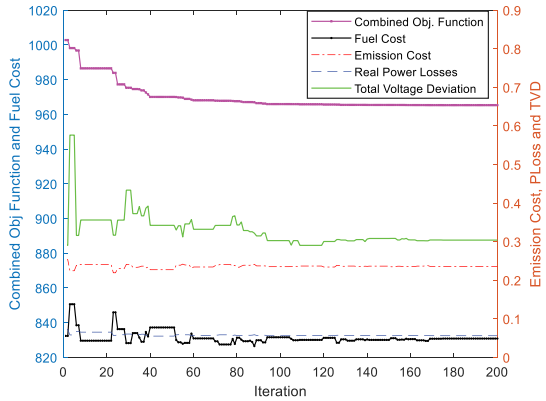
(c) GWO algorithm



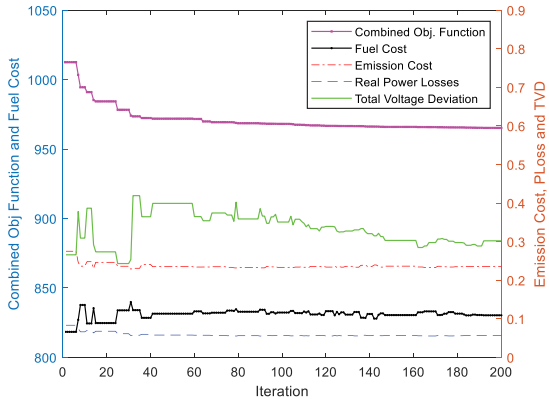
(d) Jaya Algorithm



(e) Jaya-PPS1 algorithm



(f) Jaya-PPS2 algorithm



(g) Jaya-PPS3 algorithm

Figure 2. Convergence and variation of different objective functions during the iterative process for Case 1 (a)Convergence Characteristics for various algorithms for Case 1 (b) Dragonfly Algorithm (c) GWO algorithm (d) Jaya Algorithm (e) Jaya-PPS1 algorithm (f) Jaya-PPS2 algorithm (g) Jaya-PPS3 algorithm.

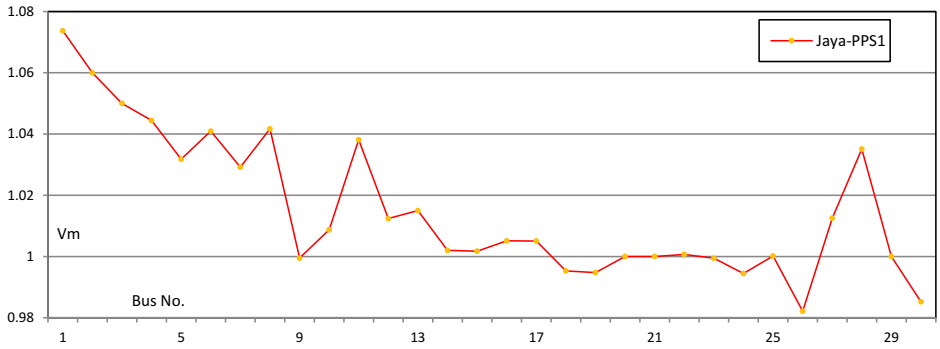


Figure 3. Voltage profile obtained using Jaya-PPS1 for Case 1.

Table 3. Results of the proposed methods and other methods for case 1.

Algorithm	Combine Objective Function (COF)	Fuel Cost	Emission	Real Power Loss	Total Voltage Deviation
Base Case	1336.645	902.004	0.222	5.842	1.160
DA	965.351	829.358	0.237	5.685	0.304
GWO	965.302	829.239	0.237	5.684	0.309
Jaya	965.286	831.549	0.235	5.578	0.311
Jaya-PPS1	**964.962**	**830.467**	**0.236**	**5.625**	**0.297**
Jaya-PPS2	965.267	830.850	0.235	5.616	0.303
Jaya-PPS3	965.255	830.290	0.235	5.642	0.302
MSA (Mohamed et al. 2017)	965.290	830.639	0.252	5.621	0.293
MPSO (Mohamed et al. 2017)	986.006	833.680	0.252	6.524	0.189
MDE (Mohamed et al. 2017)	973.611	829.094	0.257	6.056	0.303
MFO (Mohamed et al. 2017)	965.807	830.913	0.252	5.597	0.331
FPA (Mohamed et al. 2017)	971.907	835.369	0.247	5.515	0.496
MSA (Ehab and ElSayed 2019)	a	838.923	0.211	5.614	0.153
ABC (Ehab and ElSayed 2019)	a	835.523	0.207	5.394	0.138
CSA (Ehab and ElSayed 2019)	a	834.512	0.209	5.425	0.137
GWO (Ehab and ElSayed 2019)	a	851.049	0.205	4.892	0.201
BSOA (Ehab and ElSayed 2019)	a	830.711	0.225	5.744	0.183
MJAYA(Ehab and ElSayed 2019)	a	833.341	0.206	5.177	0.119
FKH (Khelifi, Bentouati, and Chettih 2019)	-	828.327	0.254	5.382	0.492
KH (Khelifi, Bentouati, and Chettih 2019)	-	827.705	0.252	5.497	0.493
FA (Khelifi, Bentouati, and Chettih 2019)	-	829.577	0.252	5.510	0.566
MOMICA (Ghasemi et al. 2014)	-	830.188	0.252	5.585	0.297
MOICA (Ghasemi et al. 2014)	-	831.225	0.267	6.022	0.404
MNSGA-II (Ghasemi et al. 2014)	-	834.561	0.252	5.660	0.430
BB-MOPSO (Ghasemi et al. 2014)	-	833.034	0.247	5.650	0.394
NKEA (Ghasemi et al. 2014)	-	834.643	0.249	5.893	0.444
MOEA/D-SF (Biswas et al. 2020)	-	883.322	0.218	4.452	0.132
GA (Taher et al. 2018)	-	830.580	0.252	5.577	0.308
MGOA (Taher et al. 2019)	-	829.963	0.252	5.634	0.291
GOA (Taher et al. 2019)	-	831.412	0.251	5.571	0.332
NLP (Alsac and Stott 1974)	-	813.74	-	7.11	-
Bisection method (Reddy. and Bijwe 2018)	-	799.455	-	8.838	-
GPM (Lee, Park, and Ortiz 1985)		804.853	-	10.486	-

aDifferent weighting factors; MSA = Moth swarm algorithm, MPSO = Modified PSO; MDE = Modified DE; MFO = Moth-Flame Optimization; FPA = Flower Pollination Algorithm; ABC = Artificial bee colony; CSA = cuckoo search algorithm; BSOA = backtracking search optimization algorithm; FFA = Firefly algorithm; KH = krill herd; MOICA = Multi-Objective Imperialist Competitive Algorithm; MNSGA-II = modified NSGA-II, BB-MOPSO = bare bones multi-objective PSO; NKEA = Neighborhood knowledge-based evolutionary algorithm; GA = genetic algorithm; MGOA = modified grasshopper optimization; NLP = Non-linear programming; GPM = Gradient Projection method.

Table 4. OPF results with optimum values of control variables for IEEE 30-bus system (Case 2).

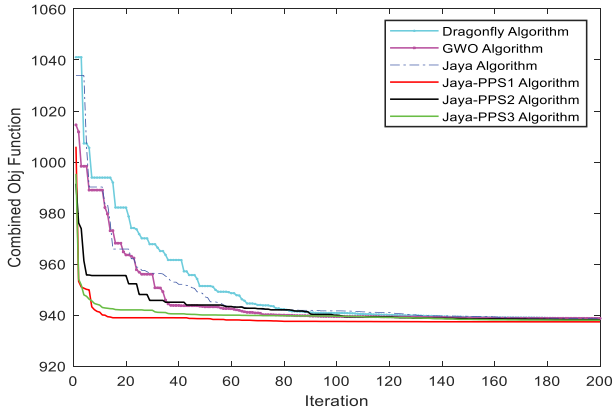
S. No.	Control Variable	DA	GWO	Jaya	Jaya-PPS1	Jaya-PPS2	Jaya-PPS3
Generator Real power output							
1	Pg_2	0.5190	0.5155	0.5157	0.5192	0.5371	0.5202
2	Pg_5	0.3118	0.3116	0.3114	0.3108	0.3112	0.3171
3	Pg_8	0.3500	0.3499	0.3500	0.3500	0.3500	0.3496
4	Pg_{11}	0.2588	0.2574	0.2625	0.2605	0.2588	0.2592
5	Pg_{13}	0.2056	0.2019	0.2056	0.2037	0.2087	0.2057
Generator voltage setting							
6	Vg_1	1.0710	1.0742	1.0637	1.0729	1.0730	1.0698
7	Vg_2	1.0574	1.0601	1.0492	1.0595	1.0602	1.0568
8	Vg_5	1.0315	1.0331	1.0223	1.0328	1.0331	1.0305
9	Vg_8	1.0400	1.0420	1.0321	1.0420	1.0415	1.0394
10	Vg_{11}	1.0993	1.0415	1.0478	1.0396	1.0394	1.0347
11	Vg_{13}	1.0243	1.0159	1.0299	1.0127	1.0190	1.0306
Transformer tap setting							
12	$T_{6\text{-}9}$	1.0101	1.0990	1.0995	1.0997	1.100	1.0919
13	$T_{6\text{-}10}$	1.1000	0.9244	0.9252	0.9592	0.9334	0.9384
14	$T_{4\text{-}12}$	1.0343	1.0231	1.0334	1.0234	1.0255	1.0520
15	$T_{28\text{-}27}$	1.0054	1.0213	1.0016	1.0076	1.0202	1.0149
Shunt VAR source setting							
16	Qc_{10}	0.0000	0.0013	0.0039	0.0499	0.0162	0.0220
17	Qc_{12}	0.0141	0.0499	0.0000	0.0500	0.0163	0.0346
18	Qc_{15}	0.0498	0.0368	0.0407	0.0476	0.0465	0.0310
19	Qc_{17}	0.0393	0.0500	0.0499	0.0495	0.0306	0.0383
20	Qc_{20}	0.0212	0.0001	0.0495	0.0497	0.0491	0.0499
21	Qc_{21}	0.0491	0.0500	0.0486	0.0500	0.0496	0.0500
22	Qc_{23}	0.0364	0.0500	0.0389	0.0360	0.0377	0.0496
23	Qc_{24}	0.0500	0.0500	0.0485	0.0499	0.0479	0.0459
24	Qc_{29}	0.0149	0.0500	0.0244	0.0194	0.0500	0.034
COF		938.581	938.498	938.378	**937.340**	937.958	937.792
Fuel Cost (FC)		811.947	811.210	812.334	**811.786**	813.336	813.036
Emission Cost (EC)		0.232	0.234	0.232	**0.232**	0.230	0.231
Real Power Loss (RPL)		5.231	5.283	5.287	**5.225**	5.182	5.193
Total Voltage Deviation (TVD)		0.338	0.314	0.252	**0.294**	0.296	0.289
Pg_{slack}		119.099	120.033	119.147	119.178	116.984	118.398
L-Index (LI)		0.104	0.101	0.103	0.101	0.101	0.102

PPS1 algorithm, which indicates that all the bus voltages are within the specified upper and lower limits.

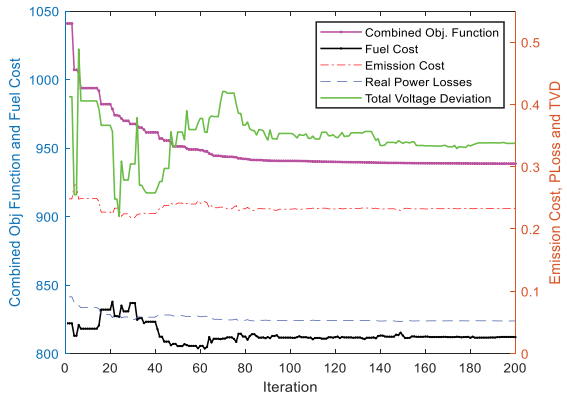
4.2. Case 2: OPF with DG in IEEE 30-bus system

In this case, the various algorithms were applied to solve the OPF problem incorporating one type 1 DG with the same COF. In Table 4, the results of this case are shown for all six algorithms along with optimal control variable settings. Table 4 shows that the proposed Jaya-PPS1 algorithm is more efficient in solving the OPF problem as compared to other methods. The combined objective function value obtained using the Jaya-PPS1 algorithm is 937.340, which is a minimum than those of DA, GWO, Jaya, Jaya-PPS2, and Jaya-PPS3.

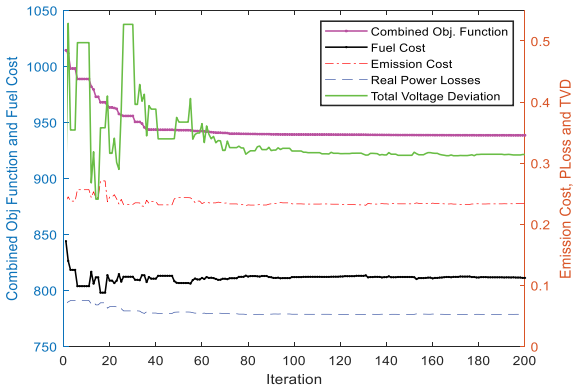
It can also be noted that after placing the DG as anticipated, the combined objective function of the proposed Jaya-PPS1 is reduced by 2.86% from 964.962 (Case 1) to 937.340 (Case 2). The convergence characteristics of various algorithms are depicted in Figure 4(a), which shows that the proposed Jaya-PPS1 algorithm has provided a smooth convergence curve with no oscillations and high convergence speed in comparison with other methods. Variation of the various objective functions during the iterative process for all the six algorithms are shown in Figure 4(b-g). Figure 5 displays the voltage



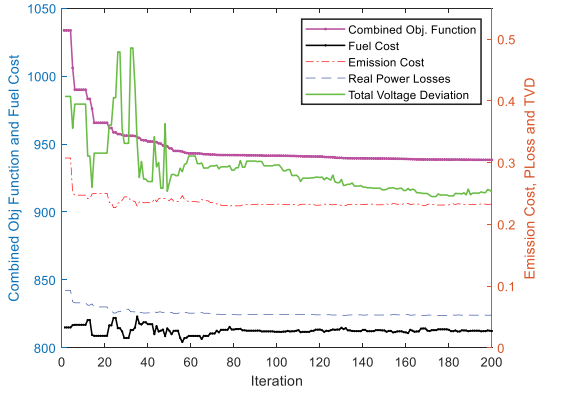
(a) Convergence Characteristics for various algorithms for Case 2



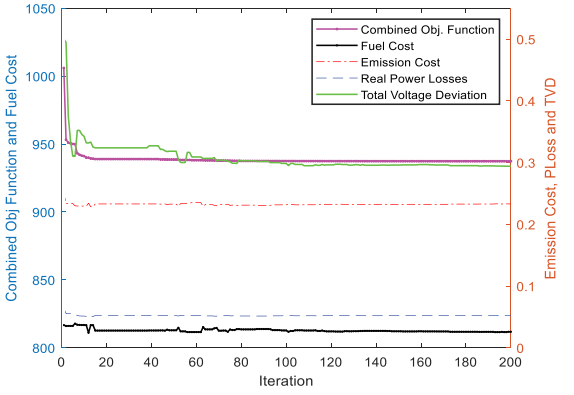
(b) Dragonfly Algorithm



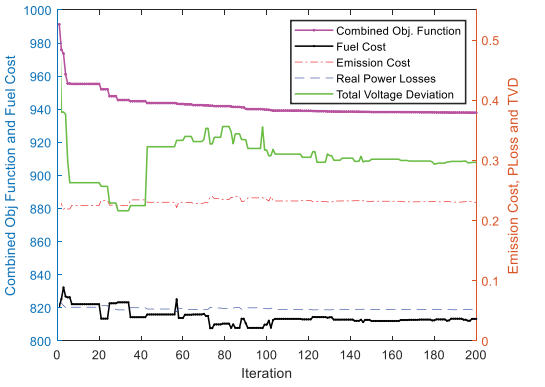
(c) GWO algorithm



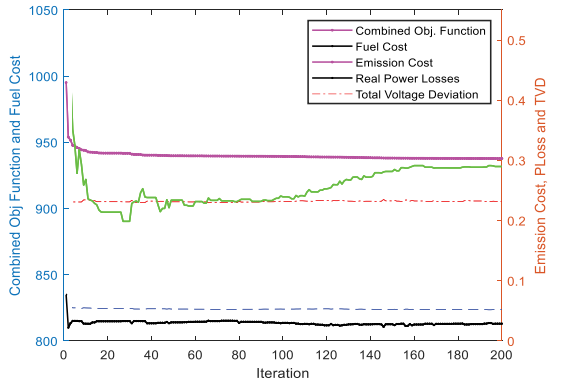
(d) Jaya algorithm



(e) Jaya-PPS1 algorithm



(f) Jaya-PPS2 algorithm



(g) Jaya-PPS3 algorithm

Figure 4. Convergence and variation of different objective functions during the iterative process for Case 2 (a) Convergence Characteristics for various algorithms for Case 2 (b) Dragonfly Algorithm (c) GWO algorithm (d) Jaya algorithm (e) Jaya-PPS1 algorithm (f) Jaya-PPS2 algorithm (g) Jaya-PPS3 algorithm.

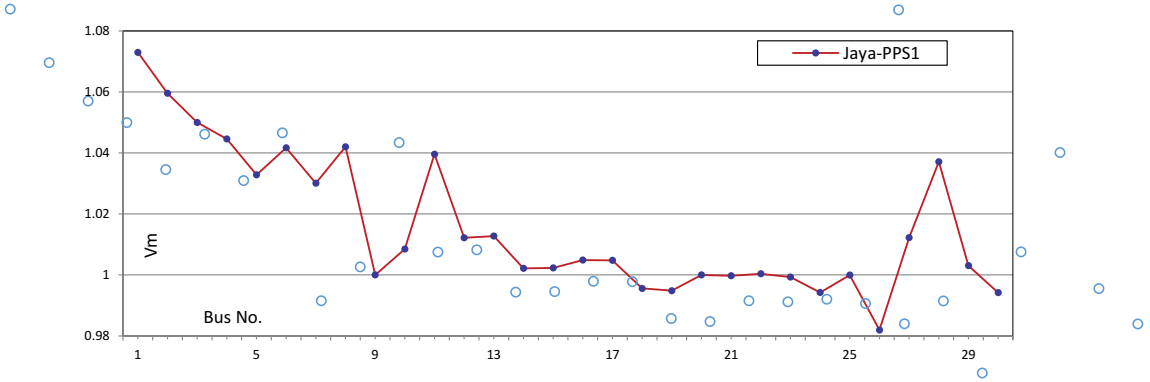


Figure 5. Voltage profile of Jaya-PPS1 for Case 2.

profile provided by the proposed Jaya-PPS1 algorithm. As can be noticed from Figure 5, the voltage magnitudes at all the buses are within the specified limits.

4.3. Case 3: OPF no DG in IEEE 57-bus system

In this case, all six algorithms were applied to solve the OPF problem in the IEEE 57-bus system without DG for the same COF. The OPF results along with the optimal control variable settings of the Jaya-PPS1 algorithm are compared with DA, GWO, Jaya, Jaya-PPS2, and Jaya-PPS3 in Table 5 and with the reported results in Table 6.

The results in Table 6 prove the dominance of the hybrid Jaya-PPS1 algorithm over other Evolutionary Computing (EC) based and Jaya, Jaya-PPS2, Jaya-PPS3 algorithms in solving the OPF problem. The proposed Jaya-PPS1 algorithm provided the COF value as 43763.10361, which is better than the COF offered by other algorithms. Also, the Jaya-PPS1 algorithm provided fast and smooth convergence characteristics in comparison with other algorithms as can be noted from Figure 6(a-g). The bus voltage profile obtained using the Jaya-PPS1 algorithm is within the specified limits as can be observed from Figure 7.

4.4. Case 4: OPF with DG in IEEE 57-bus system

In this case, IEEE 57-bus test system with two DGs is considered to determine the efficacy of the Jaya-PPS1 algorithm for solving the OPF problem. The OPF results along with optimum control variable settings obtained by DA, GWO, Jaya, Jaya-PPS1, Jaya-PPS2, and Jaya-PPS3 algorithm are listed in Table 7.

The results in Table 7 demonstrate the superiority of the proposed hybrid Jaya-PPS1 algorithm over other EC-based and hybrid Jaya-PPS algorithms in handling OPF problems in this case also. The proposed Jaya-PPS1 algorithm provided the COF value as 39108.172, which is less than that offered by other algorithms. Further, after implanting the two DGs, the COF of the proposed Jaya-PPS1 algorithm is decreased (by 10.64%) from 43763.103 (Case 3) to 39108.172 (Case 4).

As can be seen from Figure 8(a-g), the proposed Jaya-PPS1 algorithm has offered fast and smooth convergence characteristics as compared to other algorithms. The bus voltages profile obtained by the Jaya-PPS1 algorithm is also within the limits as can be observed from Figure 9.

4.5. Case 5: OPF no DG in IEEE 118-bus system

In Case 5, fuel cost is selected as the main objective. The minimum fuel cost obtained by the Jaya-PPS1 algorithm is $129221.889/h, while the minimum fuel costs obtained by Jaya-PPS2 and Jaya-PPS3

Table 5. Optimum values of control variables for IEEE 57-bus system without DG (Case 3).

S. No	Control variable	DA	GWO	Jaya	Jaya-PPS1	Jaya-PPS2	Jaya-PPS3
Generator active power output							
1	Pg2	0.9998	1.0000	1.0000	1.0000	0.9944	0.9999
2	Pg3	0.5282	0.6353	0.5709	0.5891	0.5899	0.5701
3	Pg6	0.9934	0.9236	0.8796	0.8831	0.8956	0.9564
4	Pg8	3.1544	3.1370	3.2140	3.1814	3.2025	3.1335
5	Pg9	0.9997	1.0000	0.9999	1.0000	1.0000	0.9977
6	Pg12	4.0998	4.0985	4.0992	4.1000	4.0998	4.0995
Generator voltage setting							
7	Vg1	1.0389	1.0478	1.0333	1.0308	1.0154	1.0359
8	Vg2	0.9512	1.0982	1.0998	1.0478	1.0566	1.0991
9	Vg3	1.0799	0.9754	1.0897	1.0170	1.0970	1.0567
10	Vg6	0.9500	1.0200	0.9704	0.9505	1.0027	1.0030
11	Vg8	0.9911	0.9954	1.0074	1.0139	0.9997	1.0107
12	Vg9	0.9515	1.0303	0.9734	1.0683	1.0502	1.0583
13	Vg12	1.0052	1.0168	1.0160	1.0162	1.0031	1.0141
Transformer tap setting							
14	T4-18	1.0680	1.0991	0.9802	1.0211	1.0985	1.0990
15	T4-18	0.9015	0.9000	0.9169	0.9000	0.9000	0.9482
16	T21-20	1.0028	0.9716	1.0977	1.0054	0.9722	1.0132
17	T24-25	0.9508	1.0332	1.0930	1.0527	1.0895	0.9002
18	T24-25	1.0109	1.0663	0.9002	1.0419	1.0583	1.0501
19	T24-26	1.0455	1.0357	1.0347	1.0396	1.0602	1.0314
20	T7-29	0.9257	0.9528	0.9419	0.9678	0.9257	0.9654
21	T34-32	0.9266	0.9371	0.9398	0.9278	0.9135	0.9403
22	T11-41	0.9036	0.9000	0.9001	0.9000	0.9000	0.9025
23	T15-45	0.9482	0.9631	0.9464	0.9791	0.9687	0.9667
24	T14-46	0.9462	0.9643	0.9713	0.9592	0.9543	0.9644
25	T10-51	0.9818	1.0261	0.9940	1.0025	0.9646	0.9790
26	T13-49	0.9229	0.9000	0.9309	0.9014	0.9001	0.9070
27	T11-43	0.9131	0.9141	0.9255	0.9699	0.9363	0.9729
28	T40-56	1.0981	1.0306	1.0658	0.9713	0.9991	0.9843
29	T39-57	0.9008	0.9542	0.9183	0.9001	0.9344	0.9054
30	T9-55	0.9794	0.9463	0.9924	0.9896	0.9777	0.9810
Shunt VAR source setting							
31	Qc18	0.0584	0.0326	0.0011	0.0033	0.0424	0.1997
32	Qc25	0.0844	0.1924	0.1283	0.2000	0.1692	0.1240
33	Qc53	0.1475	0.1004	0.1480	0.1870	0.1095	0.1815
COF		43887.417	43864.841	43833.642	**43763.103**	43804.936	43790.825
Fuel Cost (FC)		42584.455	42587.965	42547.094	**42573.898**	42542.989	42571.028
Emission Cost (EC)		1.357	1.344	1.370	**1.363**	1.364	1.350

(*Continued*)

Table 5. (Continued).

S. No	Control variable	DA	GWO	Jaya	Jaya-PPS1	Jaya-PPS2	Jaya-PPS3
Real Power Loss (RPL)		13.606	13.272	12.772	**12.496**	12.912	12.274
Total Voltage Deviation		0.812	0.792	0.820	**0.675**	0.775	0.743
Pg_{slack}		186.848	184.621	187.197	187.925	185.465	187.339
L-Index (LI)		0.263	0.242	0.251	0.240	0.248	0.250

Table 6. OPF Results of IEEE 57-Bus System without DG (Case 3).

Algorithm	COB	FCost ($/h)	Emission (ton/h)	PLoss (MW)	TVD (pu)
Base Case	53828.143	51395.570	2.761	28.365	1.255
DA	43887.437	42584.469	1.357	13.606	0.812
GWO	43864.841	42587.972	1.344	13.272	0.792
Jaya	43833.629	42547.092	1.370	12.771	0.820
Jaya-PPS1	**43763.103**	**42573.898**	**1.363**	**12.496**	**0.675**
Jaya-PPS2	43804.936	42542.989	1.364	12.912	0.775
Jaya-PPS3	43790.825	42571.028	1.350	12.274	0.743
MOMICA (Ghasemi et al. 2014)	-	41983.058	1.496	13.696	0.797
MOICA (Ghasemi et al. 2014)	-	41998.566	1.760	13.335	0.874
MNSGA-II (Ghasemi et al. 2014)	-	42070.824	1.496	14.455	0.889
BB-MOPSO (Ghasemi et al. 2014)	-	41994.019	1.533	12.609	1.074
NKEA (Ghasemi et al. 2014)	-	42065.996	1.517	13.976	1.042
MOEA/D-SF (Biswas et al. 2020)	-	42648.69	1.343	11.886	0.671

algorithms are $129227.810/h and $129231.178/h, respectively. The minimum fuel cost obtained using hybrid Jaya-PPS algorithms and other meta-heuristic algorithms are depicted in Table 8. From Table 8, it is clear that the fuel cost obtained from the Jaya-PPS1 algorithm is the least as compared to those of other methods. This demonstrates the effectiveness of the proposed Jaya-PPS1 algorithm as compared to Jaya-PPS2, Jaya-PPS3, DA, GWO algorithm, and other competitors in handling OPF problems in large size power system. The fuel cost characteristics of various meta-heuristic algorithms for Case 5 are shown in Figure 10.

In this paper, Powell's Pattern Search method for the local search with the classical Jaya algorithm has been incorporated in three different ways resulting in three variants namely; Jaya-PPS1, Jaya-PPS2, and Jaya-PPS3. To demonstrate the efficacy of the proposed algorithm and its potential to solve the OPF problem, it is tested on the standard IEEE 30-bus system with 24 control variables, on IEEE 57-bus system with 33 control variables, and on IEEE 118-bus system with 130 control variables.

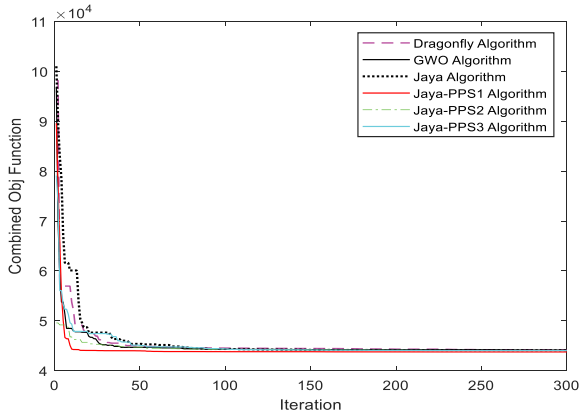
Considering the three IEEE test systems without and with DG, five cases have been solved using various EC-based computing techniques and their results are compared among themselves and with those of the recently developed algorithm available in the literature. This has been observed and can also be noted from the results shown in Table 2–8, the obtained OPF results using various algorithms are well aligned with those available in the recent literature.

For the IEEE 30-bus system, after implanting one DG unit, the combined objective function using the proposed Jaya-PPS1 algorithm is decreased by 2.86% (from 964.962, Case 1, to 937.340, Case 2). For IEEE 57-bus with two Distributed Generating units, the Jaya-PPS1 algorithm provided the fuel cost reduction by 10.64% (from 43763.103, Case 3 to 39108.172, Case 4). In the case of the IEEE 118-bus system, the Jaya-PPS1 algorithm provided fuel cost reduction by 1.52% (from 131220.020, base case to 129221.889, Case 5).

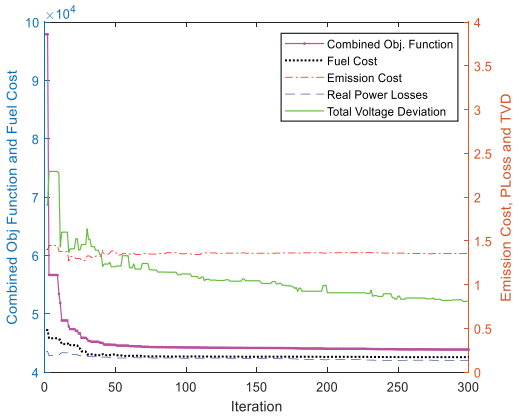
Power engineers and policymakers should perform more research in this area of study, which has the potential to yield both economic and environmental benefits. It's worth noting that the proposed Jaya-PPS1 algorithm not only minimizes the combined objective function for solving the OPF problem but also reduces fuel cost, emission, total voltage deviation, and power losses. To provide the importance to the four objectives, minimization of cost, total voltage deviation, emission, and real power losses differently, the weighing factors are to be modified accordingly. The proposed work would be helpful for power engineers for various policy analyses related to operational planning, economic operation, and future expansion of an existing system.

Type 1 DG has been modeled as a negative load and therefore its integration reduces the total load demand and thus the fuel cost and the combined objective function. The same can be observed from Table 4 (Case 2) and Table 7 (Case 4).

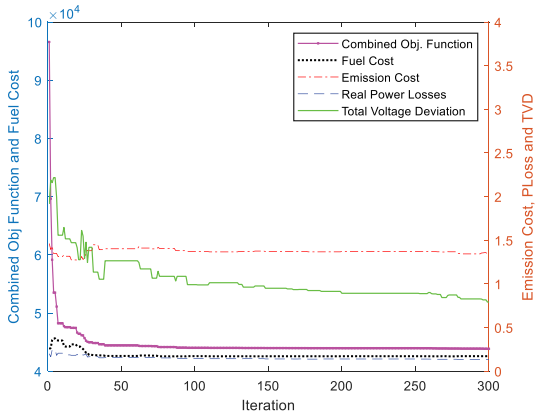
In terms of convergence characteristics, Figs. 2, 4, 6, 8, and 10 represent the convergence characteristics of all methods for cases 1, 2, 3, 4, and 5, respectively. In all these cases, the Jaya-PPS1



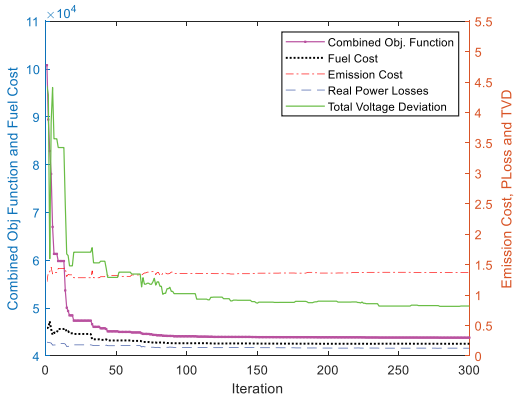
(a) Convergence Characteristics for all the six algorithms, Case 3



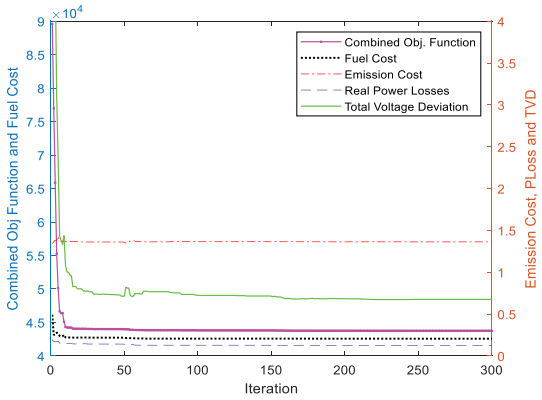
(b) Dragonfly algorithm



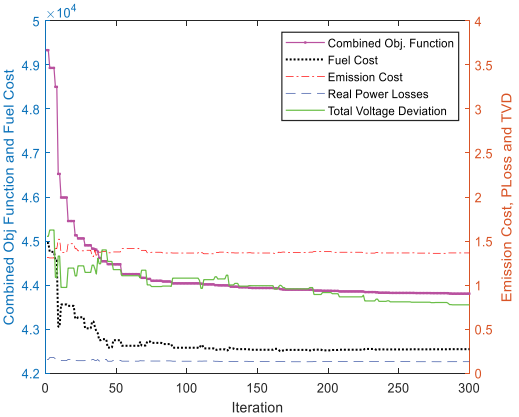
(c) GWO algorithm



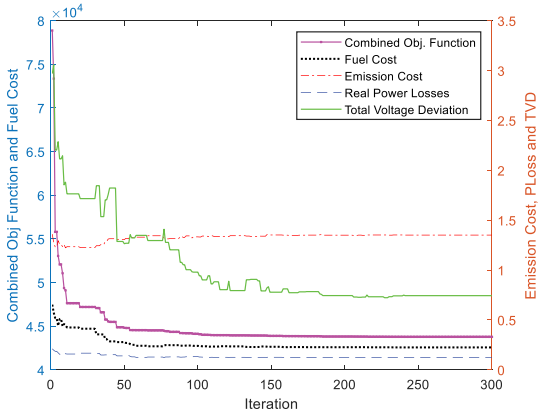
(d) Jaya algorithm



(e) Jaya-PPS1 algorithm



(f) Jaya-PPS2 algorithm



(g) Jaya-PPS3 algorithm

Figure 6. Convergence and variation of different objective functions during iterative process for Case 3 (a) Convergence Characteristics for all the six algorithms, Case 3 (b) Dragonfly algorithm (c) GWO algorithm (d) Jaya algorithm (e) Jaya-PPS1 algorithm (f) Jaya-PPS2 algorithm (g) Jaya-PPS3 algorithm.

Table 7. OPF Results for IEEE 57-bus system with two DGs (Case 4).

S. No.	Control variable	DA	GWO	Jaya	Jaya-PPS1	Jaya-PPS2	Jaya-PPS3
Generator active power output							
1	Pg2	0.9098	0.9991	0.8986	0.9891	0.9826	0.9750
2	Pg3	0.4757	0.4920	0.4745	0.4775	0.4881	0.4776
3	Pg6	0.5954	0.3695	0.4621	0.4710	0.3423	0.3086
4	Pg8	3.3367	3.4992	3.4752	3.3728	3.4902	3.5101
5	Pg9	0.9935	0.9999	0.9997	0.9963	1.0000	0.9985
6	Pg12	3.9401	3.8602	3.9434	3.9204	3.9601	3.9447
Generator voltage setting							
7	Vg1	1.0183	1.0183	1.0155	1.0199	1.0195	1.0146
8	Vg2	1.1000	1.0957	1.0989	1.0817	1.0464	1.0664
9	Vg3	1.0608	1.0541	1.0698	1.0974	1.0879	1.0996
10	Vg6	0.9500	1.0855	1.0955	0.9821	1.0610	1.0344
11	Vg8	1.0129	1.0101	1.0034	1.0069	0.9964	1.0084
12	Vg9	1.0113	0.9503	0.9863	1.0445	1.0022	1.0426
13	Vg12	1.0118	1.0202	1.0042	1.0017	1.0040	1.0059
Transformer tap setting							
14	T4-18	0.9003	1.0942	0.9000	0.9472	0.9512	1.1000
15	T4-18	1.0999	0.9000	1.0984	1.0091	0.9554	0.9449
16	T21-20	1.0443	0.9856	0.9864	0.9794	0.9992	0.9855
17	T24-25	0.9000	1.1000	1.034	1.0881	1.0357	0.9800
18	T24-25	1.0793	0.9000	0.9788	1.0669	1.0255	1.0663
19	T24-26	1.0369	1.0029	1.0099	1.0178	1.0105	1.0175
20	T7-29	0.9720	0.9722	0.9516	0.9830	0.9670	0.9709
21	T34-32	0.9435	1.0062	0.9959	0.9886	0.9845	0.9939
22	T11-41	0.9784	0.9660	0.9568	0.9746	0.9309	0.9472
23	T15-45	0.9786	0.9790	0.9809	0.9860	0.9913	0.9841
24	T14-46	0.9785	0.9769	0.9689	0.9748	0.9764	0.9801
25	T10-51	0.9885	0.9930	0.9800	0.9795	0.9781	0.9858
26	T13-49	0.9243	0.9381	0.9309	0.9015	0.9197	0.9198
27	T11-43	0.9903	1.0037	0.9842	0.9813	0.9853	0.9949
28	T40-56	0.9192	0.9008	0.9364	0.9000	0.9955	0.9146
29	T39-57	0.9807	0.9982	0.9877	1.0018	0.9268	0.9999
30	T9-55	0.9510	0.9891	0.9815	0.9278	0.9219	0.9496
Shunt VAR source setting							
31	Qc18	0.1862	0.0000	0.1289	0.0113	0.0041	0.1751
32	Qc25	0.0617	0.1467	0.1143	0.1945	0.1359	0.1334
33	Qc53	0.1296	0.1999	0.1662	0.1464	0.0910	0.1557
COF		39200.178	39173.097	39162.889	**39108.172**	39122.922	39119.337
Fuel Cost (FC)		38120.833	38114.735	38105.956	**38069.932**	38015.998	38056.723
Emission Cost (EC)		1.275	1.309	1.321	**1.284**	1.333	1.341
Real Power Loss (RPL)		12.318	13.170	12.570	**12.716**	12.816	12.852
Total Voltage Deviation (TVD)		0.545	0.450	0.472	**0.452**	0.537	0.457
Pg_{slack}		142.798	146.780	142.825	145.614	142.09	147.004
L-Index (LI)		0.139	0.124	0.129	0.129	0.127	0.127

algorithm provided smooth convergence characteristics with no oscillations and high convergence speed in comparison with other methods.

The feasibility of the results obtained in all the cases was validated by running the NR load flow program with the attained control variables settings and checking the various constraints in load flow results.

A comparison of the OPF results proves that the three variants of Jaya-PPS have provided significantly improved performance over other meta-heuristic algorithms. Out of the three variants, the proposed Jaya-PPS1 algorithm dominates other approaches for solving the OPF problem.

To assess the robustness of the DA, GWO, Jaya, Jaya-PPS1, Jaya-PPS2, and Jaya-PPS3 algorithms to solve the OPF problem with and without DG, statistical analysis was performed and shown in Tables 9 and 10. It can be seen from Tables 9 and 10, Jaya-PPS1 produces a lesser standard deviation than the DA, GWO, Jaya, Jaya-PPS1, Jaya-PPS2, and Jaya-PPS3. The small value of standard deviation reveals

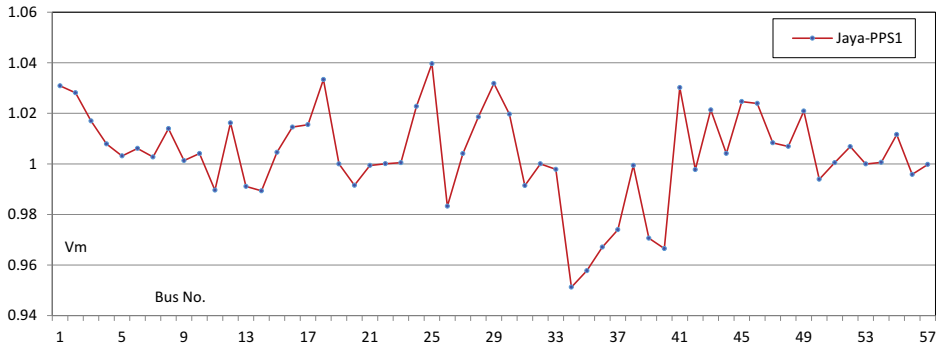


Figure 7. Voltage profile of Jaya-PPS1 for Case 3.

Table 8. Case 5 (Fuel cost minimization) results in IEEE 118-bus system.

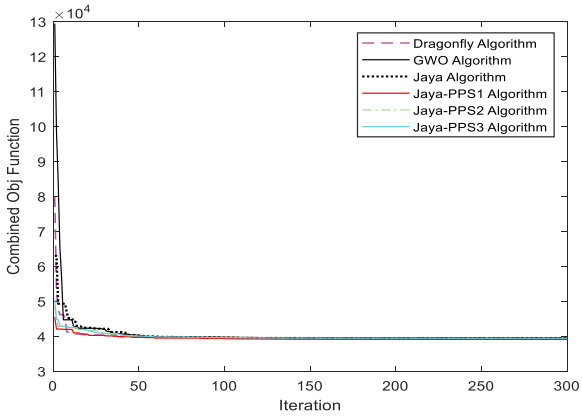
				Power Loss	
Algorithm	Fuel Cost ($/h)	TVD (pu)	Slack Bus Power (Pg_{slack})	MW	MVAr
Base Case	131220.020	1.438	513.810	132.810	782.607
Jaya-PPS1	**129221.889**	**1.535**	**471.970**	**109.336**	**746.692**
Jaya-PPS2	129227.810	1.280	468.832	102.181	647.686
Jaya-PPS3	129231.178	1.290	442.370	100.928	697.991
Jaya	129241.129	1.304	467.125	109.336	751.179
DA	129248.356	1.270	462.121	102.181	636.267
GWO	129257.145	1.401	441.035	100.928	695.114
FAHSPSO-DE (Naderi et al. 2021)	129519.38	N. A	N. A	N. A	N. A
IMFO (Taher et al. 2018)	131820.000	1.5944	407.192	77.652	-910.020
ALC-PSO (Singh, Mukherjee, and Ghoshal 2016)	129,546.084	N. A	N. A	N. A	N. A
PSOGSA (Radosavljević et al. 2015)	129, 733.58	N. A	N. A	73.21	N. A
GPU-PSO (Roberge, Tarbouchi, and Okou 2016)	129,627.03	N. A	N. A	76.984	N. A
Interior point (Jiang et al. 2010)	129,720.70	N. A	N. A	N. A	N. A
CC-ACOPF (Brust and Anitescu 2021)	129662.0	N. A	N. A	N. A	N. A
NLP (Fortenbacher. and Demiray 2019)	129700	N. A	N. A	N. A	N. A
QP(Fortenbacher. and Demiray 2019)	129600	N. A	N. A	N. A	N. A
MIQP (Fortenbacher. and Demiray 2019)	129600	N. A	N. A	N. A	N. A

IMFO = improved moth-flame optimization; ALC-PSO = particle swarm optimization with an aging leader and challengers; PSOGSA = Hybrid Particle Swarm Optimization and Gravitational Search Algorithm; GPU-PSO = Partial swarm optimization based graphics processing units; CC-ACOPF = Chance Constrained Optimal Power Flow; QP = quadratic programming; MIQP = Mixed Integer quadratic programming.

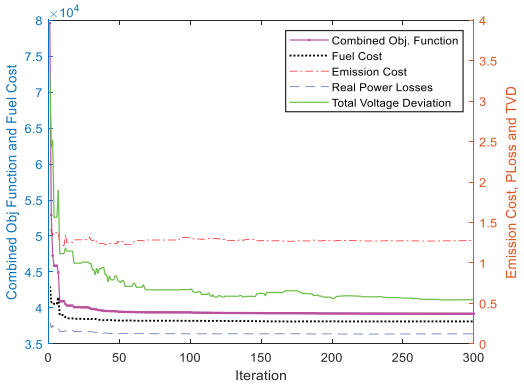
that the proposed Jaya-PPS1 approach produces statistically meaningful results as compared to other versions of the Jaya, DA, and GWO algorithm.

5. Statistical analysis

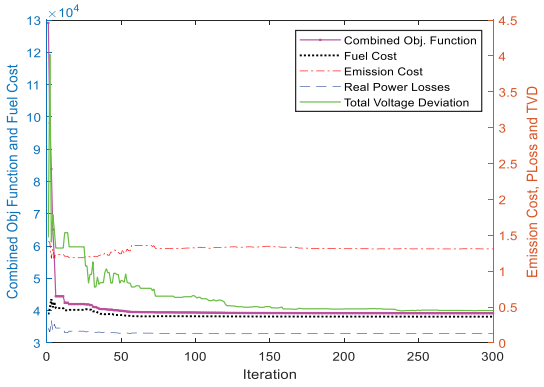
To assess the robustness of the DA, GWO, Jaya, Jaya-PPS1, Jaya-PPS2, and Jaya-PPS3 algorithms to solve the OPF problem with and without DG, statistical analysis was performed. For each case of the IEEE 30-bus and IEEE 57-bus systems, 50 independent trials with the same population size and the same number of function evaluations were carried out. Results of these 50 trials used to calculate the best, the worst, the average (mean), and the standard deviation (SD) are shown in Tables 9 and 10. Out of all the cases, the lowest values of the best, worst, average, and standard deviation provided by the proposed Jaya-PPS1 algorithm reveal that the proposed Jaya-PPS1 approach produces statistically meaningful results. This confirms the robustness of the proposed Jaya-PPS1 algorithm.



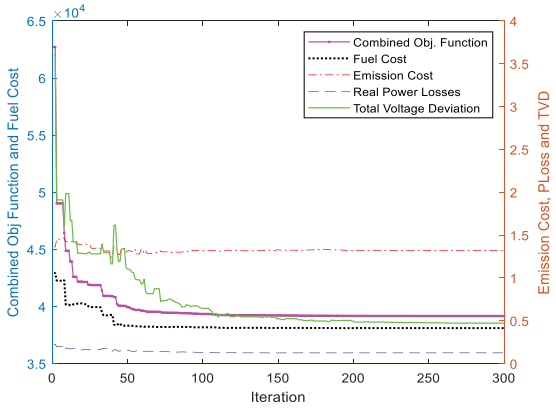
(a) Convergence Characteristics for all the algorithms for Case 4



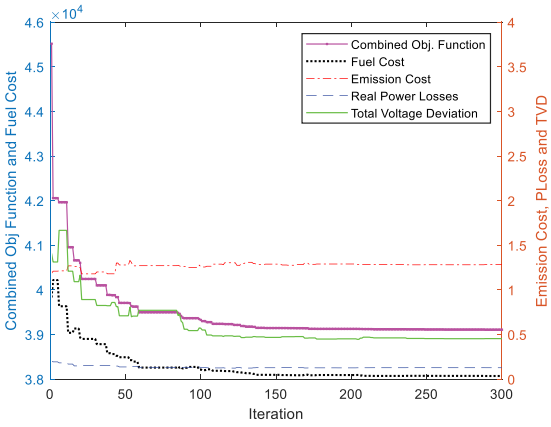
(b) Dragonfly Algorithm



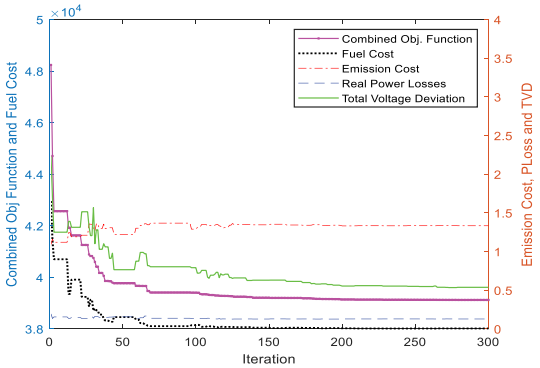
(c) GWO algorithm



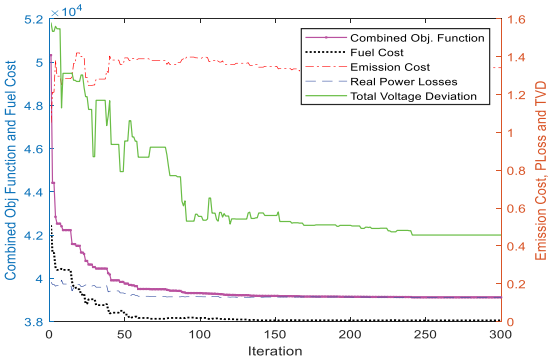
(d) Jaya algorithm



(e) Jaya-PPS1 algorithm



(f) Jaya-PPS2 algorithm



(g) Jaya-PPS3 algorithm

Figure 8. Convergence and variation of different objective functions during the iterative process for Case 4 (a) Convergence Characteristics for all the algorithms for Case 4 (b) Dragonfly Algorithm (c) GWO algorithm (d) Jaya algorithm (e) Jaya-PPS1 algorithm (f) Jaya-PPS2 algorithm (g) Jaya-PPS3 algorithm.

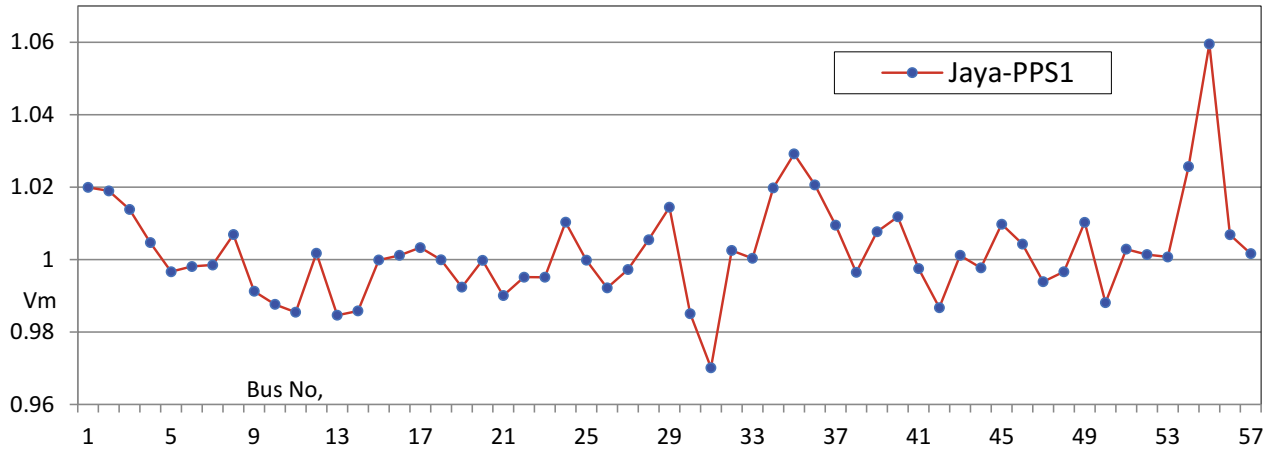


Figure 9. Voltage profile obtained using Jaya-PPS1 for Case 4.

6. Conclusion

With the inclusion of different kinds of power electronic appliances and renewable energy sources in the modern inter-connected restructured power system, the importance of solving the OPF problem is increasing many folds. In this paper, a hybrid Jaya-PPS algorithm combining Jaya algorithm with Powell's Pattern Search algorithm has been proposed to solve the OPF problem in a power system with and without DG. To demonstrate the efficacy of the proposed algorithm and its potential to solve OPF problems, it is tested on the standard IEEE 30-bus, IEEE 57-bus, and IEEE 118-bus systems for minimization of fuel cost, emission cost, real power loss, and total voltage deviation simultaneously by combining these objective functions.

Over several runs, Jaya-PPS1 consistently provided a lower objective function value in all the case studies regardless of the complexities and size of the power system. For example, the IEEE 30-bus and IEEE 57-bus system results in savings of fuel cost approximately 2.86% and 10.64%, respectively, in the original system, which is equivalent to the savings of 241,968.72 USD and 40,777,195.6 USD annually. In the case of standard IEEE 118-bus system, the Jaya-PPS1 algorithm provided a fuel cost reduction by 1.52% in comparison to the base case, which is equivalent to cost saving of 17,503,627.60 USD yearly. The policymakers should consider these findings for future planning, control, and cost-effective operation of power system.

To provide the importance to the four objectives, minimization of cost, total voltage deviation, emission, and real power losses differently, the weighing factors are to be modified accordingly. The OPF results of all the cases obtained using various meta-heuristic algorithms were compared with the reported results. A comparison of the OPF results and statistical analysis proves that the Jaya-PPS1 algorithm is the best option for solving OPF problems and maybe an efficient hybrid methodology for solving other optimization problems.

Table 9. Performance Measures of various algorithms for IEEE 30-Bus System.

	Without DG				Incorporating DG			
Algorithm	Best	Worst	Mean	Std. Deviation	Best	Worst	Mean	Std. Deviation
DA	965.351	966.435	965.873	0.0252	938.581	939.176	938.755	0.0261
GWO	965.302	966.733	965.756	0.0215	938.498	939.246	938.867	0.0238
Jaya	965.286	966.815	965.897	0.0198	938.378	939.254	938.978	0.0195
Jaya-PPS1	**964.962**	**965.127**	**965.018**	**0.0114**	**937.340**	**938.012**	**937.832**	**0.0102**
Jaya-PPS2	965.267	966.395	965.787	0.0195	937.958	938.878	938.138	0.0161
Jaya-PPS3	965.255	966.385	965.987	0.0189	937.792	938.437	938.198	0.0158

Table 10. Performance Measures of various algorithms for IEEE 57-Bus System.

	Without DG				Incorporating DG			
Algorithm	Best	Worst	Mean	Std. Deviation	Best	Worst	Mean	Std. Deviation
DA	43887.437	43973.873	43893.893	0.0298	39200.178	39218.879	39207.656	0.0288
GWO	43864.841	43896.887	43871.698	0.0291	39173.097	39181.365	39178.432	0.0282
Jaya	43833.629	43845.953	43839.894	0.0282	39162.889	39175.542	39168.764	0.0281
Jaya-PPS1	**43763.103**	**43774.391**	**43767.692**	**0.0133**	**39108.172**	**39112.265**	**39110.356**	**0.0132**
Jaya-PPS2	43804.936	43815.873	43811.348	0.0271	39122.922	39131.654	39126.554	0.0273
Jaya-PPS3	43790.825	43804.768	43798.562	0.0269	39119.337	39128.278	39125.456	0.0271

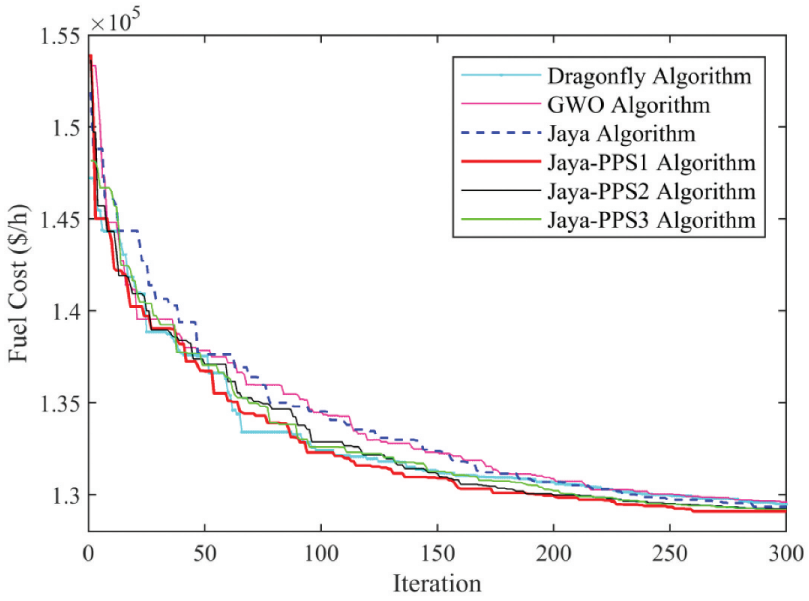


Figure 10. Convergence Characteristics for various algorithms for Case 5.

The aim to incorporate PPS with Jaya is to combine the benefits of both algorithms. Three variants of the Jaya-PPS algorithm, Jaya-PPS1, Jaya-PPS2, and Jaya-PPS3 were developed by incorporating hybridization in different manners. In comparison with other meta-heuristic algorithms, the proposed hybrid Jaya-PPS1 approach has the benefits of the simplicity to adopt, fast, and smooth convergence characteristic, and a guaranteed near global optimal solution. As the hybrid Jaya-PPS1 algorithm has good exploration and exploitation property, it can be employed to solve OPF problems in practical power systems and many real-world optimization problems. When any hybrid algorithm is developed for solving an optimization problem, some options to incorporate hybridization should also be tried to obtain the best one.

As a future scope of the work, the OPF problem can be considered with discrete control variables like transformer taps-setting and shunt compensators. The Jaya-PPS1 hybrid algorithm may be applied for solving OPF problems in practical power systems under normal and contingency conditions. In this paper, the multi-objective OPF problem with various objective functions was converted into a single-objective OPF problem and was solved using a single-objective optimization algorithm. The same multi-objective OPF problem can be solved using any multi-objective optimization algorithm like the multi-objective Jaya algorithm or so.

Nomenclature

x	dependent variable	$Q_{g_k}^{min}$ and $Q_{g_k}^{max}$	minimum and maximum limits of the reactive power output of the *kth* generator
u	control variable	$P_{g_k}^{min}$ and $P_{g_k}^{max}$	minimum and maximum active power limits of *kth* generating unit
Z (x, u)	objective function	T_k^{max} and T_k^{min}	maximum and minimum tap setting of *kth* transformers
NGN	numbers of generators	$V_{L_k}^{max}$ and $V_{L_k}^{min}$	maximum and minimum voltage limit of *kth* load bus
NB	number of buses	$S_{l_k}^{max}$	maximum MVA flow in *kth* transmission line
NLB	number of load buses	R_1, R_2, R_3 and R_4	penalty factors corresponding to limit violations
Ntl	number of transmission lines	A_i, B_i, and C_i	fuel cost coefficients of the *ith* generating unit
P_i and Q_i	active and reactive power of *ith* bus	P_{g_1}	slack bus generator's active power output
NC	number of VAR compensation units	$\alpha_i, \beta_i, \gamma_i, \xi_i, \lambda_i$	emission coefficients of *ith* generating unit
NTR	number of regulating transformers	pop	Population size
P_{di} and Q_{di}	active and reactive load	$Iter_{max}$	Maximum No. of iterations
P_{gi} and Q_{gi}	Active & reactive power generations	NFE	Number of function evaluations
P_{Loss} and Q_{Loss}	real and reactive power loss	JFE	Number of Jaya Function Evaluations
$V_{g_k}^{min}$ and $V_{g_k}^{max}$	minimum & maximum voltage limits of *kth* generator bus	PSFE	Number of PPS Function Evaluation

ORCID

Saket Gupta http://orcid.org/0000-0003-4541-5798
Narendra Kumar http://orcid.org/0000-0003-1083-2129
Laxmi Srivastava http://orcid.org/0000-0002-5922-6835

References

Abido, M. A. 2002. Optimal power flow using tabu search algorithm. *Electric Power Components and Systems* 30 (5):469–83. doi:10.1080/15325000252888425.

Alsac., O., and B. Stott. 1974. Optimal load flow with steady-state security. *IEEE Transactions on Power Apparatus and Systems* 93 (3):745–51. doi:10.1109/TPAS.1974.293972.

Azizivahed, A., A. Arefi, E. Naderi, H. Narimani, M. Fathi, and M. R. Narimani. 2020. An efficient hybrid approach to solve bi-objective multi-area dynamic economic emission dispatch problem. *Electric Power Components and Systems* 48 (4):485–500. doi:10.1080/15325008.2020.1793830.

Biswas, P. P., P. N. Suganthan, R. Mallipeddi, and G. A. J. Amaratunga. 2020. Multi-objective optimal power flow solutions using a constraint handling technique of evolutionary algorithms. *Soft Computing* 24 (4):2999–3023. doi:10.1007/s00500-019-04077-1.

Bouchekara, H. R. E. H., and M. A. Abido. 2014. Optimal power flow using differential search algorithm. *Electric Power Components and System* 42 (15):1683–99. doi:10.1080/15325008.2014.949912.

Bouchekara, H. R. E. H., M. A. Abido, A. Chaib, and R. Mehasni. 2014. Optimal power flow using the league championship algorithm: A case study of the Algerian power system. *Energy Conversion and Management* 87:58–70. doi:10.1016/j.enconman.2014.06.088.

Brust, J. J., and M. Anitescu. 2021. Convergence analysis of fixed point chance constrained optimal power flow problems. *arXiv Preprint*. arXiv:2101.11740.

Carpentier, J. 1979. Optimal power flows. *International Journal of Electrical Power and Energy Systems* 1 (1):3–15. doi:10.1016/0142-0615(79)90026-7.

Charles, J. K., and N. A. Odero. 2013. Effects of distributed generation penetration on system power losses and voltage profiles. *International Journal of Scientific and Research Publications* 3:12. ijsrp.org/research-paper-1213/ijsrp-p2458.pdf.

Chattopadhyay, D. 1994. Operational planning of power system: An integrated approach. *Energy Sources* 16 (1):59–73. doi:10.1080/0090831940890906.

Elattar, E. E., and S. K. ElSayed. 2019. Modified JAYA algorithm for optimal power flow incorporating renewable energy sources considering the cost, emission, power loss and voltage profile improvement. Energy 178:598–609. doi:10.1016/j.energy.2019.04.159

Elattar, E. E., and S. K. ElSayed. 2019. Modified JAYA algorithm for optimal power flow incorporating renewable energy sources considering the cost, emission, power loss and voltage profile improvement. *Energy* 178:598–609. doi:10.1016/j.energy.2019.04.159.

Fortenbacher., P., and T. Demiray. 2019. Linear/quadratic programming-based optimal power flow using linear power flow and absolute loss approximations. *Electrical Power and Energy Systems* 107:680–89. doi:10.1016/j.ijepes.2018.12.008.

Freris., L. L., and A. M. Sasson. 1968. Investigation of the load-flow problem. *Proceeding IEE (London)* 115 (10):1459–70. doi:10.1049/piee.1968.0260.

Ghasemi, M., S. Ghavidel, M. M. Ghanbarian, M. Gharibzadeh, and A. A. Vahed. 2014. Multi-objective optimal power flow considering the cost, emission, voltage deviation and power losses using multi-objective modified imperialist competitive algorithm. *Energy* 78:276–89. doi:10.1016/j.energy.2014.10.007.

Jiang, Q., G. Geng, C. Guo, and Y. Cao. 2010. An efficient implementation of automatic differentiation in interior point optimal power flow. *IEEE Transactions on Power Systems* 25 (1):147–55. doi:10.1109/TPWRS.2009.2030286.

Khaled, U., A. M. Eltamaly, and A. Beroual. 2017. Optimal power flow using particle swarm optimization of renewable hybrid distributed generation. *Energies* 10 (7):1013. doi:10.3390/en10071013.

Khelifi, A., B. Bentouati, and S. Chettih. 2019. Optimal power flow problem solution based on hybrid Firefly Krill Herd method. *International Journal of Engineering Research in Africa* 44:213–28. doi:10.4028.

Lee, K. Y., Y. M. Park, and J. L. Ortiz. 1985. A united approach to optimal real and reactive power dispatch. *IEEE Transactions on Power Apparatus and Systems* 104 (5):1147–53. doi:10.1109/TPAS.1985.323466.

Meraihi, Y., A. R. Cherif, D. Acheli, and M. Mahseur. 2020. Dragonfly algorithm: A comprehensive review and applications. *Neural Computing and Applications* 32 (21):16625–46. doi:10.1007/s00521-020-04866-y.

Mirjalili, S., S. M. Mirjalili, and A. Lewis. 2014. Grey Wolf Optimizer. *Advances in Engineering Software* 69:46–61. doi:10.1016/j.advengsoft.2013.12.007.

Mohamed, A.-A. A., Y. S. Mohamed, A. A. M. Gaafary, and A. M. Hemeida. 2017. Optimal power flow using moth swarm algorithm. *Electric Power Systems Research* 142:190–206. doi:10.1016/j.epsr.2016.09.025.

Momoh, J. A., R. Adapa, and M. E. E. Hawary. 1999. A review of selected optimal power flow literature to 1993. I. Nonlinear and quadratic programming approaches. *IEEE Transactions on Power Systems* 14 (1):96–104. doi:10.1109/59.744492.

Naderi, E., M. P. Kasmaei, and H. Abdi. 2019. An efficient particle swarm optimization algorithm to solve optimal power flow problem integrated with FACTS devices. *Applied Soft Computing Journal* 80 (243):262. doi:10.1016/j.asoc.2019.04.012.

Naderi, E., M. P. Kasmaei, M. F. V. Cerna, and M. Lehtonen. 2021. A novel hybrid self-adaptive heuristic algorithm to handle single- and multi-objective optimal power flow problems. *International Journal of Electrical Power and Energy Systems* 125:106492. doi:10.1016/j.ijepes.2020.106492.

Narang, N., J. S. Dhillon, and D. P. Kothari. 2012. Multi-objective fixed head hydrothermal scheduling using integrated predator-prey optimization and Powell search method. *Energy* 47 (1):237–52. doi:10.1016/j.energy.2012.09.004.

Nguyen, T. T. 2019. A high performance social spider optimization algorithm for optimal power flow solution with single objective optimization. *Energy* 171:218–40. doi:10.1016/j.energy.2019.01.021.

Niknam, T., M. R. Narimani, M. Jabbari, and A. R. Malekpour. 2011. A modified shuffle frog leaping algorithm for multi-objective optimal power flow. *Energy* 36 (11):6420–32. doi:10.1016/j.energy.2011.09.027.

Peesapati, R., V. K. Yadav, and N. Kumar. 2018. Flower pollination algorithm based multi-objective congestion management considering optimal capacities of distributed generations. *Energy* 147:980–94. doi:10.1016/j.energy.2018.01.077.

Power system test cases. [online]. Accessed <http://www.ee.washington.edu/research/pstca >

Quadri, I. A., S. Bhowmick, and D. Joshi. 2018. A comprehensive technique for optimal allocation of distributed energy resources in radial distribution systems. *Applied Energy* 211:1245–60. doi:10.1016/j.apenergy.2017.11.108

Quadri, I. A., S. Bhowmick, and D. Joshi. 2018. A comprehensive technique for optimal allocation of distributed energy resources in radial distribution systems. *Applied Energy* 211:1245–60. doi:10.1016/j.apenergy.2017.11.108.

Radosavljević, J., D. Klimenta, M. Jevtić, and N. Arsić. 2015. Optimal power flow using a hybrid optimization algorithm of particle swarm optimization and gravitational search algorithm. *Electric Power Components and Systems* 43 (17):1958–70. doi:10.1080/15325008.2015.1061620.

Rao, R. V. 2016. Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems. *International Journal of Industrial and Engineering Computing* 7 (1):19–34. doi:10.5267/j.ijiec.2015.8.004.

Rao, S. S. 2019. *Engineering optimization: Theory and practice*. 5th. New York: John Wiely & Sons. doi:10.1002/9781119454816.

Reddy., S. S., and P. R. Bijwe. 2018. An efficient optimal power flow using bisection method. *Electrical Engineering* 100 (4):2217–29. doi:10.1007/s00202-018-0698-4.

Roberge, V., M. Tarbouchi, and F. Okou. 2016. Optimal power flow based on parallel metaheuristics for graphics processing units. *Electric Power Systems Research* 140:344–53. doi:10.1016/j.epsr.2016.06.006.

Sakthivel, V. P., M. Suman, and P. D. Sathya. 2020. Squirrel search algorithm for economic dispatch with valve-point effects and multiple fuels. *Energy Sources, Part B: Economics, Planning, and Policy* 15 (6):351–82. doi:10.1080/15567249.2020.180345.

Singh, A. K., and S. K. Parida. 2016. Impact analysis of RES-based DGs with economical considerations in restructured electricity market. *Energy Sources, Part B: Economics, Planning, and Policy* 11 (5):472–79. doi:10.1080/15567249.2011.649824.

Singh, K., V. K. Yadav, N. P. Padhy, and J. Sharma. 2014. Congestion management considering optimal placement of distributed generator in deregulated power system networks. *Electric Power Components and Systems* 42 (1):13–22. doi:10.1080/15325008.2013.843218.

Singh, R. P., V. Mukherjee, and S. P. Ghoshal. 2016. Particle swarm optimization with an aging leader and challengers algorithm for the solution of optimal power flow problem. *Applied Soft Computing Journal* 40:161–77. doi:10.1016/j.asoc.2015.11.027.

Srivastava, L., and H. Singh. 2015. Hybrid multi-swarm particle swarm optimisation based multi-objective reactive power dispatch. *IET Generation, Transmission & Distribution* 9 (8):727–39. doi:10.1049/iet-gtd.2014.0469.

Stavrakas., V., and A. Flamos. 2020. A modular high-resolution demand-side management model to quantify benefits of demand-flexibility in the residential sector. *Energy Conversion and Management* 205:112339. doi:10.1016/j.enconman.2019.112339.

Stavrakas, V., S. Papadelis, and A. Flamos. 2019. An agent-based model to simulate technology adoption quantifying behavioural uncertainty of consumers. *Applied Energy* 255. doi:10.1016/j.apenergy.2019.113795.

Taher, M. A., S. Kamel, F. Jurado, and M. Ebeed. 2018. An improved moth-flame optimization algorithm for solving optimal power flow problem. *International Transactions on Electrical Energy Systems*. doi:10.1002/etep.2743.

Taher, M. A., S. Kamel, F. Jurado, and M. Ebeed. 2019. Modified grasshopper optimization framework for optimal power flow solution. *Electrical Engineering* 101: (1):121–48. doi:10.1007/s00202-019-00762-4.

ACHSM Asia-Pacific Journal of Health Management

RESEARCH ARTICLE

AN INTEGRATIVE APPROACH FOR WELLNESS: AN ASSESSMENT OF POTENTIAL OF AYURVEDA DURING COVID-19 PANDEMIC

Richa Chauhan[1], Nidhi Maheshwari[2]

1. Banasthali Vidyapith, Rajasthan, India
2. Delhi Technological University, Delhi, India

Correspondence: richachauhan1986@gmail.com

ABSTRACT

The integrative approach for wellness has become an arena where integrating the indigenous practice with the dominant bio-medical system has activated a transformation process.Further, a debate characterized by fundamental dichotomies of health/wellness is there to find an approach in agreement with holistic wellness. The limitations of protocol-driven medical science and avoiding the patient behind the patient have again compelled us during this pandemic to explore holistic health.

This paper presents an analysis and explores the feasibility of Ayurveda as an integrative system to wellness and tests the Ayurveda concept of holistic wellness against three such dichotomies: 'health' vs wellness, 'body' vs 'mind,' and 'curative vs. preventive,' where the current medical model can be said to represent the first elements in the three dichotomies.

Keeping the current Covid-19 pandemic as a critical incident, researchers tried to justify that the synergistic confluence will act as a complementary/supportive knowledge in humanity's struggle against the pandemic. For synergistic confluence of Ayurveda and medical Science, the scientist's temperament and evidence-based medical science approach should be applied to validate the principles and therapies of Ayurveda.

KEYWORDS

Holistic Wellness, Ayurveda, Synergistic Confluence, Integrative approach of Wellness, Covid-19 Pandemic

INTRODUCTION

Covid-19, pandemic compelled us to ponder over the essence of life. The crisis has prompted the foundations and settled ways of living life and remedies for managing the situation. Further, it has aggravated the need to relook, reassess, and reappraise the prevalent customs and systems, as specific ways of crisis management are no longer appropriate for sustaining our very existence.

In this context, highly qualified practitioners from bio-medical science are trying to address the pandemic through an evidence-based medical approach. A systematic approach wherein the best available research evidence with clinical expertise and patient values are reviewed for problem-solving. As a result, the knowledge generated by the bio-medical system is reliable as well as reproducible, which provides the substantial groundwork for identifying new health technologiestreatments that have led to longevity and quality of human life.

The scientific community's efforts to fight against COVID are mechanical as they are protocol-driven and algorithmic.[1] However, to manage a crisis of this magnitude, we need to develop a holistic approach involving social, psychological, medical, physiological, administrative, and even integrative intelligence for the human victory over the virus. [2] Here the researchers are not defying the truth that medical science has not collected a comprehensive repository of knowledge about the impact of COVID-19 on human life. Still, the efforts are embedded in physiological nourishment, i.e., body only. A holistic view that stimulates body, mind, and spirit is not yet emphasized while evolving measures to manage this crisis. At this juncture, when the high intensity of psychological distress directed by the pandemic is all-pervasive, [3] we need to recalibrate our ways to manage the crisis. The psychological distress generated during the crisis can potentially enhance the vulnerability to viral upper respiratory infections. [4] This requires a rational approach to find ways to minimize pain to improve immunity so that chances of infection can be reduced. [5]

Medical science is an evidence-based curative approach and more inclined towards the elimination of diseases. Medical science focuses on diseases, diagnostics, and medicines rather than the patient as a person in the treatment process. The body is merely appreciated in science rather than the complete perspective necessarily being comprised of body, mind, and spirit. It is appropriate to mention here that the absence of disease is no guarantee of health. Human helplessness has created a need for an integrative approach where along with evidence-based functioning, experience-based systems like Ayurveda, Yoga can also be adopted.

The underlying meaning of Ayurveda is "knowledge (Veda) of life (ayur)."It is an Indian system of life sciences that is based on predictive and personalized medicine.

While advocating this health and wellness wisdom of ancient India, we have no intention to degrade reductive medical science; instead, we promote the synergistic confluence of these two sciences. Integrative intelligence is helpful for disease prevention and curative care. The COVID pandemic is an opportunity to explore this integrative approach to fulfil our unmet needs and unresolved challenges. Further, the integrative approach postulates the active involvement of individuals in maintaining health which is always required to minimize the passive dependence on medical treatments. [7]

THE OBJECTIVE OF THE STUDY:

1. To strengthen the advocacy that Ayurveda is based on holistic wellness principles.
2. To explore the possibility of synergistic confluence of medical science and indigenous Ayurvedic practices.
3. To reinforce the perspective that the indigenous system can support the effort of medical science against the Covid-19 pandemic.

Building on this background, the current research reviewed the existing literature to understand the possibility of synergistic confluence of medical science and indigenous Ayurvedic practices. Though both medical science and indigenous Ayurveda differ epistemologically, their philosophies are not significantly different.

Further, keeping the current pandemic as a critical incident, researchers tried to justify that the synergistic confluence can act as a complementary/supportive knowledge to the medical science. Finally, to provide evidence to our advocacy that Ayurveda is based on holistic wellness principles, the impact of Ayurveda's interventions, including immunity boosting measures for self-care during the COVID-19 crisis was reviewed.

The approach of Ayurveda comes from the indigenous practices adopted by Indian communities, and it is ongoing, creating a solid premise of experience-based healing therapy. We begin by arguing that the curative approach is not sufficient for wellbeing and then tried to elaborate that Ayurveda has tremendous potential for synergistic confluence.

METHODOLOGY

Ayurveda's philosophy and theories are advocated through a literature review as a holistic approach to wellness. In our advocacy, we do not make any extreme position, either bio-medical or Ayurveda; instead, we suggested that synergistic confluence of Ayurveda and biomedical systems may be possible through integrative approaches for holistic wellness. Individually, they are not the complete package for health, but their synergistic confluence can help the world suffering from incurable diseases and illness.

To strengthen our perspective regarding the potential impact of Ayurveda on psycho and physical immunity, taking COVID as a critical incident, different initiatives

taken by state governments and practitioners who have a firm conviction for Ayurveda and followed the guidelines of the Indian government during the pandemic was reviewed. Finally, we discuss ways to progress for the synergistic confluence of Ayurveda with bio-medical science by collecting wisdom from this pandemic. The presented data is collected from several published reports, Ayurveda's classical scriptures, and books on Ayurveda, one of the indigenous sciences of life. The methodological approach combines the concepts and philosophy of Ayurveda with as many different sources as possible: written sources, scriptures, and discussions.

Holistic wellness and the dichotomies of health-wellness, body-mind and cure vs care as constructed

Wellness is considered the holistic integration of the energies of body, mind, and spirit where the body is stimulated, the reason is engaged, and nature is cherished rather than physical health.

Nurturing physical, intellectual, emotional, vocational, financial, social, spiritual, and environmental mutually supporting dimensions is a must for holistic wellness, as neglect of any of these can affect individuals' wellbeing.
In contrast to holistic wellness, the therapeutic approach of the healthy body is defined only in mechanistic terms. [8] Therefore, it prompts that mechanized and engineered interventions can sort out health issues. [9] Thus, illness is caused by specific conditions and organisms, for which remedial action is proposed bothregulatory and eliminatingthe illnessin a particular manner. [10] This way, it is postulated the health and disease are directed by physical, chemical, and biological laws. So, the cause-effect is established in the linear relationship between the body as a system and its interaction with discrete entities. [11]

In a contrary sense holistic wellness recognizesthe integrated impact of body, mind, and spirit on health, which in principle differs from the dichotomies of health-illness and body-mind as constructed in medical science. Under holistic wellness and the predictive and preventive approach, it is essential to follow personalized and participatory medicine for complete wellness. The systems approach can help demystify diseases and democratizing health care [12], but holistic wellness is a proactive approach to eliminating illness.

In the contemporary context, when medical science struggles to sort out the puzzle of Covid-19, the alternative/complementary/integrative systems needs to be explored. In this context, keeping the principle of holistic wellness in premise, this paper explores the feasibility of Ayurveda as an integrative system for welfare.

This paper tests the Ayurveda concept of holistic wellness against three such dichotomies: health vs wellness, body vs mind, and curative vs. preventive, where the current medical model can be said to represent the first elements in the three dichotomies.

Health versus wellness dichotomy

In biomedical science, health is accepted as the absence of illness, injury, disease, and health care, ensuring to regain the state where there is no illness. On the contrary, in indigenous studies, wellbeing is not just physical health, but preventative health carewhich is more important. Further, in indigenous discussions 'substantial knowledge often has been trivialized in favor of more scientific, objective ways of knowing'. [13]

World Health Organization in 1946 states that –health is a state of complete physical, mental and social wellbeing and not merely the absence of disease or infirmity – lendingitself well to the focus of holistic wellness (WHO 1946). This definition favoursthe view that the indigenous definition and measurement of the health of communities can be referred to while focusing on holistic wellness.

Body versus mind dichotomy

Under biomedical philosophy, the body is taken as a machine. That is why it is defined as a functional, mechanical, physical, and systematic thing. Elaborating on the body as a machine, [14] identified three critical dimensions, being the body is the sum of its parts means it can be reducible to organ, tissue, and cells. Like machines, it can be understood from the outside without emphasizing emotions and mood. Third, thebody is not guided by values. From this narrative, the disinterest of biomedical science is apparent towards the spirit or mind.

Indigenous management defines the mind as a controller over the five sense organs and works as a motor organ-enabling activities through speech and other organs. The body is represented as a realm of molecules, while the mind is a field of thoughts. That's why indigenous management postulates that whatever affects the brain has its effect on memory and vice-versa. It implies that

wellness can't be achieved just by adopting the functional/bodily measures rather, a holistic approach assimilating body, mind, and spirit is needed. [15]

Curative versus preventive dichotomy

Curative medicine is a branch of medical science aiming to restore and maintain health once individuals fall ill. In contrast, Preventative medicine is a proactive approach that ensures that people do not fall sick in the first place. Therefore, preventive medicine targets the disease-free state, a proactive approach, while curative medicine targets the diseased state, a reactive one.

Most of the prevalent disease of this century, like lifestyle disorders, HIV/AIDS, Ebola, Cancer, Corona, take a longer duration to aggravate before changing into life-threatening conditions. Therefore, preventive strategies make better sense for them.[16] It is not advocated here that treatments should be avoided, or preventive strategies are enough for eradicating disease; instead, early diagnosis and prevention are needed, which can be accessed through indigenous methods.

Holistic wellness, according to Ayurveda.

The underlying meaning of Ayurveda is "knowledge (Veda) of life (ayur)."It is an Indian system of life sciences that is based on predictive and personalized medicine. Based on the severity, stage, and subtype of the disease, pathogenesis, etiological factors, clinical features, therapeutic modalities can be described under Ayurveda. [17] This is different from medical science as an individualized approach and is followed with a thorough examination of the deceased individual. In addition, Ayurveda proliferates the gifts of nature in maintaining healthy and happy living. [18]

Healthcare-related quality standards, as well as all regulatory procedures, are well documented in Ayurveda. Ayurveda's extensive knowledge base on preventive care derives from the concepts of 'Dinacharya'(daily regimes) and 'Ritucharya' (seasonal systems) to maintain a healthy life. This refers to seasonal changes rendering the body profoundly vulnerable to one or different disorders if the body is not synchronized with the weather, which makes up the external environment. As adjustments are the key to endurance, the information on Ritucharya (routine for different seasons) is, therefore, significant. Further, Ayurveda emphasizes 'Samadosha' (balances of psychological elements), 'Samadhatu' (balances of tissues), 'Sama-indriya' (balance of sensory organs) 'Prasannatma' (nurturing soul),' Samagnischa' (proper production of energies), 'Malakriya' (proper elimination of waste). [19]

Integrative approach considering body, mind, and spirit for managing wellness

The logical and reasonable synthesis of Ayurveda (rooted in scientific and rational concepts of the Indian philosophy) defines health as fulfilling continuous physical, mental, and spiritual happiness by striking a balance between psyche, spirit, and physical system. It elaborates those relationships need to be nurtured with fellow creatures and nature as a whole, including the universe and ourselves. Charaka Samhita describes wellbeing as a 'disease-free state' to be followed for the accomplishment of 'virtue, wealth and gratification'. [20]

Ayurvedic principle of three-body humours (TriDosas) viz.- Vata (the energy of movement) Pitta (the energy of digestion and metabolism) Kapha (the energy of lubrication) are the bases on which treatment is defined combining diet, herbal remedies, detoxification, whenever disequilibrium exists between the physical, mental and spiritual energy than this disequilibrium generates imbalanced Agni which brings illness [19]. Ayurveda advocates that personal indulgence for monitoring, managing, and maintenance is required for nutritional status.

Effective interventions that are natural and less invasive

The fundamental precept of Ayurveda talks about the relationship between the microcosm (Pind) and the macrocosm (Brahmand). It elaborates that both the 'being' (Pind) and cosmos/outside world (Brahmand) interacts, and the cosmos influences the being. In simple words, that is a unity between macrocosm and microcosm. This dictum is extended on the ontological basis of the five-element theory (Panchamahabhuta). [21] On these lines, Ayurvedic approach emphasizes health promotion by interconnecting the environment and personal's characteristics. It can be postulated that everybody has the potential of self-healing by knowing the state of his or her body, mind, and spirit. [19]

Ayurveda promotes the usage of materials as medicines that are readily available in Indian households. Simple, self-driven, but effective health practices- such as lukewarm water, drinking Kadha (Decoction) during sickness, scraping the tongue, gazing at a flame to improve

eyesight, gargling for maintaining the health of the throat - are natural and less invasive interventions.

Further, to withstand the agents causing diseases, the most natural healing therapy, 'Yoga' and Dhyana(meditation), has been prescribedto enhance physical strength, emotional resilience and control the nervous system.

The TrayodosaAgni - Thirteen types of digestive fires like- Jatharagni (gastric fire) are narrated in Ayurveda. Ayurveda highlights that the malfunctioning of these fire (Agni) causes most of the diseases. This is how in a natural way, body functioning is explained, which stresses that several diseases can be controlled by systematically managing over 'Ahar' (Food),'Achar'(Routines)'Vichar' (Thoughts) and 'Vihar' (Recreation)' [22]

Individualized approach: addresses the person's unique conditions, needs and circumstances.

Ayurveda differentiates every individual from others based on Prakriti (Constitution). That is why each patient will not get the same treatment in Ayurveda though having a similar disease. Ayurveda accepts that each personality consists of TriGunas (three fundamental universal energies):–that is Satva (power of balance), Rajas (power of fire or motivation), and Tamas (energy of rest or slothfulness), and all individuals have mixed amounts of the three, the predominant Guna(energy) determines an individual's Mansa Prakriti (Mental Constitution). In equilibrium, the three Gunas (energies) preserve the mind (indirectly the body) and different types of mental disorders. [23]

Knowing this classification of personality helps to understand the responses of patients in pathological and stressful situations. Dietary prescription to the patients is also natural to prescribe based on Satvik (foodwhich is usually fresh, seasonal and locally grown like fruits, vegetables, leaves, grains, cereals, milk, honey) Rajsik, and Tamsik Prakriti.

Even the Ayurvedic classification of body tissues, which is done under the principle of SaptaDhatu(Seven types of body tissues), is very helpful in understanding the anatomy of the body and the related causes of disease. [24]

Prevention reduces the need for treatment:

Ayurveda propounded the theory of PancaMahabhutas (Five essential elements) viz.-Akasha (space), Vayu (air), Teja or Agni(fire), Jala(water) and Prithvi (earth). This theory helps the physicians understand the body's essential elements to detect illness, which further helps to identify the real cause of the disease. [25]. Ayurveda stresses mind, body, and spirit wellness, and all these are interconnected, but the spirit has the power to govern the other. The knowledge of Ayurveda empowers individuals to understand a person and generates energy to establish the equilibrium of body, mind, and spirit as per the individual's Prakriti (constitution). The prescribed non-pharmacological approaches from yoga, meditation, and lifestyle play a significant role in treating mental and physiological disease and suggest ways to prevent diseases and protect health. [26]

Inquiry-driven: patient and practitioner are partners in the healing process:

This philosophy of 'our health in our hands' is practiced in Ayurveda, which advocates that all the diseases can be considered as the result of 'mithya,'ahara,' and 'vihara' follow. These three vitiates Agni, are from which conditions are derived. If an individual practices yogic posture and incorporates meditation in their daily routine, it will help strengthen inner strength to handle health challenges. [27] Ayurveda also counts the psyche of the patients and allows them to participate actively in the healing process.[28]

Ayurveda does not consider individuals a system consisting of organs but a well-coordinated network of relationships that define the functions. Eventually, the roles become impaired; that is why the Ayurvedic concepts are more appropriate for wellness. The individualized approach isemphasized in Ayurveda, enabling individuals to recognize their bodily strength, limitations and continuous self-assessment to give them an accurate picture of their immunity strength. Further, every purpose results from the mutual efforts of several parameters in the body ranging from structures biochemistry, electrical and magnetic, mental and emotional, and physiological parameters. This shows that the functional model of Ayurveda is inclusive.

Assessment of Ayurveda Principles during COVID-19 Pandemic

COVID-19 could be a rapidly changing and progressing situation. In this context, Ayurveda has enough potential and possibilities for prevention and an adjunct treatment option for COVID-19.

Based on the Ancient Ayurvedic literature and proved indigenous practices, the Ministry of Ayush, the government of India, recommends certain practices for holistic wellness

(immunity boosting). Similarly, Patanjali research institute, known for its Ayurvedic medicines in collaboration with the National Institute of Medical Science, Jaipur invented the Corona Virus medicineCoronil (a combination of Giloy, Tulsi, and Aswagandha).

Segregating an ailing member of a family as a household chore in a member's demise are still followed across India. These efforts are similar to those we have started taking up as self-isolation measures to prevent the spread of the flu to other family members or within the community if we are afflicted with the COVID-19 virus. [29]

In infectious disease, it is imperative to prevent further progress, which helps control the disease, and various methods can be used for immunity building and recovery. [30] Through Ayurvedic medicines, multiple symptoms associated with COVID-19 like high fever, severe cough, and body pain can be cured, and further progression of these symptoms can be stopped. [30], During COVID-19, un-precedent usage of Ayurveda's medication is an opportunity to explore the possibility of synergistic confluence of Ayurveda and medical science.

Synergistic confluence of Ayurveda and medicalscience

The current pandemic has created a need to understand health and wellness from different lenses rather than just the reductionist health approach. We need to negate the silos approach and find a more objective, integrative, comprehensive, and holistic approach to wellness.

The limitations of protocol-driven medical science and the avoidance of the person behind the patient have again compelled us during this pandemic to explore new ways for wellness and avoid such pandemics in the future. At the same time, Ayurveda cannot just rest on the glory of ancient indigenous methods but must understand the role of the evidence-based scientific temperament of health and resolve the queries/objections of medical science. The flexible approach of both systems can pave the way for confluence, a massive success for humankind.

Integration through thecoherence of philosophies, concepts, and scientific and professional rigours of both systems must be reassessed. The scientist's temperament and evidence-based medical science approach should be applied to validate the principles and therapies of Ayurveda to generate an integrative approach, which will be the complete health solution.

This integration should be unbiased and should be an effort to do the rebalancing of health philosophy. The approach needs to move from reduction to holistic, from prevention to protection and from body to body-mind-spirit. The definition of health determinants needs to be extended from drug to food, lifestyle, and attitude towards life. Environment, Prakriti' and concepts of holistic wellness need to explore for the integrative health approach. To alter the pluralist approach, it is essential that the insights and perceptions contributed by holistic wellness concepts need to be referred to. We need to focus our attention on the attributes of individuals and their context so that a person-centric health approach can be developed.

CONCLUDING REMARKS

The current pandemic and the limitations of medical science have instigated the need to explore an integrative approach. The discussion under the three paradoxes showed that Ayurveda has enough potential and possibilities to prevent and cure the diseases. [30] In contrast to the medical approach, where the body is considered a composition of cells, tissues, and organs, Ayurveda's approach is more holistic. Ayurveda's principles were compared with the holistic wellness principles, and it was found that they agree with the laws of a holistic wellness approach.

Consequently, we suggested an integrative approach where Ayurveda and evidence-based medical science will help to shift from generalized to person-centric treatment and illness to wellness.

References:

1. Naudé W. Artificial intelligence vs COVID-19: limitations, constraints and pitfalls". Ai & Society. 2020;1.
2. Lakhotia SC. Book review- integrative approaches for health: Biomedical research, Ayurveda and yoga. Proc Indian Natl Sci Acad (A Phys Sci) [Internet]. 2015;81(3). Available from: http://dx.doi.org/10.16943/ptinsa/2015/v81i3/48227
3. Norris FH, Friedman MJ, Watson PJ. 60,000 disaster victims speak: Part II. Summary and implications of the disaster mental health research. Psychiatry. 2002 Autumn;65(3):240–60.

4. Pedersen, A., Zachariae, R., &Bovbjerg, D. H. (2010). Influence of psychological stress on upper respiratory infection—a meta-analysis of prospective studies. Psychosomatic medicine, 72(8), 823-832.
5. Sanderson WC, Arunagiri V, Funk AP, Ginsburg KL, Krychiw JK, Limowski AR, et al. The Nature and Treatment of Pandemic-Related Psychological Distress". Journal of Contemporary Psychotherapy. 2020;1–13.
6. Nandram SS. Understanding Integrative Intelligence: Embodied in S-model. 2017.PraanUitgeverij. Publication.
7. Hammond-Tooke, D. (1989). Rituals and medicines: Indigenous healing in Africa. Johannesburg: AD. Donker (Pty) Ltd.
8. Bhatt, S., &Nimkar, S. (2015). Clinical research in Ayurveda: a preliminary review of 225 papers published in Indian Ayurveda journals. Journal of Dental and Medical Sciences, 14(2), 43-50.
9. Social stratification and self-management of health". Journal of Health and Social Behavior. 1986;1–14.
10. Hewa S. The coming revolution in Western medicine: a biopsychosocial model for medical practice". International review of modern sociology. 1994;117–125.
11. Hood L, Price ND. Demystifying disease, democratizing health care". 2014.
12. Tangenberg KM, Kemp S. Embodied practice: Claiming the body's experience, agency, and knowledge for social work". Social work. 2002;47(1): 9–18.
13. The machine metaphor in medicine". Social contexts of health, illness, and patient care. 1981. p. 218–249.
14. Luskin F. Transformative practices for integrating mind–body–spirit". Journal of Alternative & Complementary Medicine. 2004;10(Supplement 1):15.
15. Velanovich V. Preoperative laboratory screening based on age, gender, and concomitant medical diseases". Surgery. 1994;115(1):56–61.
16. Kessler C, Wischnewsky M, Michalsen A, Eisenmann C, Melzer J. Ayurveda: between religion, spirituality, and medicine". Evidence-Based Complementary and Alternative Medicine. 2013;
17. Traditional knowledge-based medicine: a review of history, principles, and relevance in the present context of P4 systems medicine". Progress in preventive medicine. 2017;2(7):0011.
18. Exploring Ayurvedic knowledge on food and health for providing innovative solutions to contemporary healthcare". Frontiers in Public Health. 2016;4:57.
19. Therapeutic potentials of metals in ancient India: A review through Charaka Samhita". Journal of Ayurveda and integrative medicine. 2011;2(2):55.
20. Nandram S. Integrative Spirituality in the Fourth Industrial Revolution: From How We Do Things To Why We Exist. 2019.
21. Vani G, Prasad JSRA. Concept of agni and aharapaka (metabolic transformation) in Ayurveda". International Ayurvedic Medicine Journal. 2013;1:1–7.
22. Shilpa S, Murthy CV. Understanding personality from Ayurvedic perspective for psychological assessment: A case". Ayu. 2011;32(1):12.
23. Sharma V, Chaudhary AK. Concepts of DhatuSiddhanta (theory of tissues formation and differentiation) and Rasayana; probable predecessor of stem cell therapy". Ayu. 2014;35(3):231.
24. Malik K, Mishra B. Panchamahabhuta-AadharbhutSiddhant and their application in Chikitsa".. Journal of Ayurveda and Integrated Medical Sciences (ISSN. 2018;3(5):146–150.
25. Preventive and therapeutic measures of Ayurveda for mental health care". Pharma Science Monitor. 2014;5(1).
26. Pyles L. Healing justice: Holistic self-care for change makers. Oxford University Press; 2018.
27. Girija PLT, Sivan N. Ayurvedic Treatment of COVID-19/SARS-CoV-2: a case report". Journal of Ayurveda and Integrative Medicine. 2020;
28. Nussbaumer-Streit B, Mayr V, Dobrescu AI, Chapman A, Persad E, Klerings I, et al. Quarantine alone or in combination with other public health measures to control COVID-19: a rapid review. Cochrane Database Syst Rev. 2020;9:CD013574.
29. Panda AK, Gangaplara A, Buszko M, Natarajan K, Boyd LF, Sharma S, et al. Cutting edge: Inhibition of the interaction of NK inhibitory receptors with MHC class I augments antiviral and antitumor immunity. J Immunol. 2020;205(3):567–72.
30. Ayurveda and anti-microbial resistance". Journal of Ayurveda and Integrative Medicine. 2018;9(2):85.

GLOSSARY

S.NO.		
1.	The Tri Doshas	Three Body Humours
2.	The Trayodosa Agni	Thirteen types of digestive fires
3.	The Tri Gunas	Three Fundamental universal energies
4.	The SaptaDhatus	Seven types of body tissues rasadhatu (plasma), raktadhatu(blood), mamsadhatu (muscle), medadhatu (fat), asthidhatu (bone), majjadhatu (nervous tissue), and shukradhatu (male reproductive tissue) or artavadhatu (female reproductive tissue).
5.	PanchaMahabhutas	Five basic elements Space (Aakash), Air(Vaayu),Fire(Agni),Water(Jal),Earth(Prithvi)
6.	The Tri Malas	Three types of Body Wastes Faeces (Purisa), Urine (Mutra), Sweat (Sveda)
7.	Sattva	A substance, experience, or mental state infused with the qualities of light, clarity, intelligence, compassion, and wisdom.
8.	Rajas	A substance, experience, or mental state infused with the qualities of kinetic energy, movement, passion, and action.
9.	Tamas	A substance, experience, or mental state infused with the qualities of inertia, darkness, heaviness, slowness, sleepiness, and decay.
10.	Rasayana	A substance that nourishes and tones the entire body; the Ayurvedic practice of rejuvenation therapy (also known as rasayanachikitsa)—a specific process of offering deep nourishment to the cells, tissues, and organs of the body in support of their healing, renewal, and regeneration; this practice is indicated in a number of different situations (e.g., after a deep cleanse like panchakarma) and is believed to enhance immunity, stamina, and longevity.
11.	Vata	It is predominated by the ether and air elements and governs movement and communication; it is light, cold, dry, rough, mobile, subtle, and clear.
12.	Pitta	It is predominated by the fire and water elements, and it governs transformation; it is light, sharp (or penetrating), hot, oily, liquid, and spreading.
13.	Kapha	It is predominated by the earth and water elements and governs structure and cohesiveness; it is heavy, slow, cool, oily, smooth, dense, soft, stable, gross, and cloudy.
14.	Kadha (Decoction)	Decoction is a method of extraction by boiling herbal or plant material to dissolve the chemicals of the material, which may include stems, roots, bark and rhizomes.

15.	ManasPrakriti	The mental constitution; each individual's unique proportion of sattva, rajas, and tamas in the mind; manasprakriti is established at conception, but can change over time, reflecting our capacity to develop more (or less) evolved states of consciousness over the course of our lives.
16.	Prakriti	Constitution: the unique ratio of vata, pitta and kapha established at conception and resulting in a personally unique set of physical, emotional, and mental tendencies, strengths, and vulnerabilities.
17.	Dinacharya	A daily routine: an important part of an Ayurvedic lifestyle that helps to align our bodies with the daily rhythms of nature; the traditional dinacharya includes a wide variety of daily self-care practices including a rich personal hygiene routine, exercise, spiritual practice, meals, and sleep.
18.	Ritucharya (rtucharya)	A seasonal routine: similar to the concept of dinacharya, but also accounting for the cycle of the seasons; rtucharya encourages us to adapt our personal routines to align more closely with the rhythms of the natural world, introducing practices and qualities that naturally promote balance all year long

Ad Hoc Networks 121 (2021) 102603

Contents lists available at ScienceDirect

Ad Hoc Networks

journal homepage: www.elsevier.com/locate/adhoc

Anomaly detection framework to prevent DDoS attack in fog empowered IoT networks

Deepak Kumar Sharma [a,*], Tarun Dhankhar [a], Gaurav Agrawal [a], Satish Kumar Singh [a], Deepak Gupta [b], Jamel Nebhen [c], Imran Razzak [d]

[a] Department of Information Technology, Netaji Subhas University of Technology, Delhi, India
[b] Department of Computer Science and Engineering, Maharaja Agrasen Institute of Technology, Delhi, India
[c] Prince Sattam bin Abdulaziz University, College of Computer Engineering and Sciences, Saudi Arabia
[d] School of Information Technology, Deakin University, Geelong, Australia

ARTICLE INFO

Keywords:
Anomaly detection
Continuous ranked probability score
DARPA-99
DDoS attack
Dimensionality reduction
Fog computing
ICMP attack
Internet of things
Kernel density estimation
SYN attack

ABSTRACT

Internet of things or in short IoT is a network of interconnected entities such as computing devices, mechanical machines, digital gadgets etc. Cloud based IoT infrastructures are susceptible to Distributed Denial of Service (DDoS) attacks. A DDoS attack may render the server useless for a long period of time causing the services to crash due to extensive load. In this project we will try to introduce the concept of fog computing and try to explain its importance in a 3-tier architecture. We have proposed an anomaly detection architecture for IoT networks where the detection actually happens on the fog layer. The algorithm is based on the CRPS metric which is a single variable algorithm which is the case in most statistical algorithms. Therefore, we have proposed a way to use multiple variables and shown why it is required in a heterogeneous network like IoT. For detection purposes(testing data) we have used Week 5 Day 1 data of DARPA 99 as it contains a TCP SYN attack initiated once for a duration of 6 min 51 s and for ICMP Week 4 Day 1 data of DARPA 99 is used it has 2 attacks for 1s each. The algorithm is able to identify these attacks correctly.

1. Introduction

Internet of things or in short IoT is a network of interconnected entities such as computing devices, mechanical machines, digital gadgets etc. These entities work with each other by either collecting or producing some data or by processing some data in order to provide a smart adaptable environment for the users. In smart home environments, all the home appliances communicate with each other to provide better living experience for e.g. a thermostat can inform the air conditioner to regulate the room's temperature. There are many such use cases of IoT networks such as smart traffic systems, smart security networks etc. It is estimated that there will be more than 21 billion IoT devices by 2025. We will be using apps to control the LEDs, door locks and other home appliances in our houses. Our cars will be able to communicate with each other to avoid road accidents. Cities will have a huge network of interconnected security cameras, alert systems etc.

But with this level of connectivity comes a great risk of security. Devices connected to an IoT network are very prone to cyberattacks. Therefore, there is a need to develop security mechanisms, attack detection systems, attack mitigation systems, secure IoT devices and connections etc. in order to make next generation IoT networks more secure and reliable. Many researchers have proposed such attack detection and mitigation techniques working at different layers of the network architecture and preventing the systems from various cyber attacks. A cyberattack over a network creates anomalous behaviour in the network traffic which means the flow of packets is different than usual and this fact can be exploited to detect attacks using various anomaly detection algorithms. Many anomaly detection techniques based on artificial intelligence, neural networks, statistics etc. have been used and these techniques have shown promising results against many attacks such as phishing attacks, DDoS attacks, sinkhole attacks etc. In this work we will propose a similar but efficient anomaly detection technique to prevent DDoS (Distributed Denial of Service) attacks in fog-empowered IoT networks.

In the past, many researchers have proposed various DDoS detection and mitigation techniques [1–4]. Machine learning algorithms and Artificial neural networks are very popular and useful choices for detecting such protocol based attacks in wireless networks. In [5], Wang et al. used a semi supervised clustering algorithm to detect DDoS attacks. In

* Corresponding author.
E-mail addresses: dk.sharma1982@yahoo.com (D.K. Sharma), tdhankhar@gmail.com (T. Dhankhar), gauravagrawal.2626@gmail.com (G. Agrawal), satishsingh23@gmail.com (S.K. Singh), deepakgupta@mait.ac.in (D. Gupta), j.nebhen@psau.edu.sa (J. Nebhen), imran.razzak@deakin.edu.au (I. Razzak).

https://doi.org/10.1016/j.adhoc.2021.102603
Received 15 April 2021; Received in revised form 21 June 2021; Accepted 6 July 2021
Available online 10 July 2021
1570-8705/© 2021 Elsevier B.V. All rights reserved.

this method, 3 features are selected by analysing the characteristics of the attack to form a detection feature vector. Then a multi-feature based constrained k-means algorithm is used which improves the convergence speed and the accuracy of the detection mechanism. In [6], Sumathi et al. proposed a DDoS detection mechanism using neural networks. Their model evaluates the network traffic using a deep learning classifier based on cost minimization strategy for publicly available datasets. They used detection accuracy, cost per sample, packet loss,average delay, packet delivery ratio, overhead and throughput as evaluation metrics for the performance analysis. The proposed algorithm works really well on the accuracy metrics but use of deep learning based methods can be computationally heavy for small edge devices. In [7], Kadri et al. explains a special statistical technique that can be used to detect DDoS attack based on a metric called CRPS. The paper demonstrates how CRPS can be used for probabilistic forecasting of DDoS attack. The paper also demonstrates how it can be used for real time detection in time series based environments. The proposed algorithm uses a single variable for detection of DDoS attack which might not always be a correct method of detecting DDoS attack in such heterogeneous networks like IoT.

We will try to propose an anomaly detection model based on CRPS for Fog-empowered networks with an attempt to show the importance of multiple features and demonstrate how we can use them to detect anomaly in a given traffic.

IoT works on two layer architecture or a three layer architecture and is vulnerable to various cyber security attacks. One such attack is Distributed denial of services(DDoS). DDoS attack is an attempt to disrupt targeted servers in a malicious way. DDoS attack takes help of some compromised nodes and floods the targeted server with requests ultimately leading to the breakdown of the server. At a higher standard, DDoS attacks clog the destination server with unexpected traffic. DDoS attacks are a little less complicated than any other cyber Security attacks. There are 3 categories to DDoS attack (a) Volume based attacks (b) Protocol attacks (c) Application based attacks. Common types of DDoS attack include SYN flood attack, UDP attack, ICMP attack etc.

Recently, it has been seen that IoT devices have been a major force to drive DDoS attacks. This is a threat that has not been diminished. Security in IoT has been a major talk as it opens up many avenues for attacks to take place. It is anticipated that around 20.4 million devices are due to be deployed by 2020, and it is safe to say that the scale of DDoS attack utilizing this vulnerability could have serious consequences. So, it is important that researchers work on a security technology that is well baked from the start. Some further observation that motivated us to work on DDoS is the use of Multivariate statistical algorithms in detection of DDoS attack. Considering multiple variables rather than single variables can bring out some interesting results.

With respect to the interest and motivation presented earlier our aim with this project is to propose an anomaly detection Model for fog empowered networks. With the help of fog nodes present in the network we can detect anomalies locally. It also eases the task of detecting the malicious node for future research works. We will discuss more about this in further sections. We want to use statistics based algorithms for the purpose of anomaly detection(DDoS Detection). These algorithms are usually used for time series based datasets where packets tend to possess a trend in their incoming pattern. In a statistics based algorithm, to detect the DOS and DDOS attacks, every new traffic network measure is compared to the reference attack free traffic distribution. In simple words trends are analysed and on the basis of the trends detected anomalies are spotted.

In most of the recent works, usually one feature for example : number of data packets per unit time, is used for the detection of DDoS attacks for time series based datasets. Our aim is to use more than one parameter by analysing the dataset which will help in detection of anomalies. The IoT traffic is heterogeneous in nature so using one feature may not be a correct criteria always for judging the anomaly. Therefore, in further sections we will discuss the valuable parameters that can be extracted from the dataset and how we can detect anomalies by implementing statistics based algorithms. Our algorithm uses dimensionality reduction which will be used for extracting important information from features generated and the algorithm is based on Continuous ranked probability score(CRPS) score. CRPS is used in probabilistic forecasting. We will be using python for implementing the algorithm on a dataset and present the output.

The primary contributions of this paper are:

1. Proposed a model for anomaly detection of protocol based attacks.
2. Highlighting the importance of 3 layer architecture in IoT based networks.
3. Mentioned the importance of multiple features while detecting anomalies in IoT traffic and how it can be incorporated in the CRPS metric using dimensionality reduction techniques.

The rest of the paper is organized as follows. In Section 2, the detailed analysis of the concepts used is written. Section 3 explains the component wise working of the proposed anomaly detection model. In Section 4, the evaluation results are discussed with the help of obtained graphs and findings. Section 5 concludes the work by mentioning the key achievements and scope for future works.

2. Detailed analysis

Recently, the source code of a famous mirai malware was released, According to the experts the malware scans the web for IoT devices that have not been secured properly and infect them, as it is easier to hack them. After the devices have been compromised they are used as a part of a botnet that is directed to launch DDoS attacks with malicious intention. All the IoT devices at a global stores its information in the cloud, compromising the IoT devices means hackers can cause discrepancy in the information stored in the cloud which might cause a loss of large information. The problem is that the experts believe that the cyber security attacks on IoT are nowhere stopping here. With lack of security the attacks will increase in a vast number.

Being one of the hottest technologies in recent times, IoT should come with a great responsibility of security. Although it has numerous benefits, these devices are easily hacked which gives it a negative image in the market. IoT should be using new hardware and software which are being generated daily, eradicating the old technologies.

The main challenge with this technology is to detect anomalies in real time and check if the system is affected or not as early as possible since the system is monitoring multiple devices at all times [8,9]. Further examples of Security can be home security, autonomous vehicles security, data management using autonomous systems etc.

2.1. Advantages of fog computing

As discussed in Section 1, Fog Computing includes a 3-layered network architecture where the topmost layer is cloud layer which consists of centralized huge data centres and servers, middle layer is called fog layer where small local servers are present really close to the bottom most layer of sensory nodes to support low latency real time processing. Nodes in the fog layer are called fog nodes. These fog nodes receive data from a small environment and process that data to send desired results with very low propagation time delay. After sending the required response back, the data can then be sent to the cloud servers for further processing and analysis. Usually a group of smart homes or a small network of roads or a production line in a factory can have its own fog server node to collect and process data coming from a bunch of sensory nodes such as cameras, thermostats, pressure gauges etc. [10].

Apart from providing low latency, real-time data processing and distributed network of processing units, addition of fog layer also provides security benefits. Because the data traffic coming from the sensory nodes passes through the fog nodes first, many security checks

can be performed to prevent anomalous data or traffic from reaching cloud servers. This mechanism helps in avoiding cyber attacks like DoS or DDoS attacks to a great extent by alarming the network whenever a node tries to flood the server by sending a large amount of packets (SYN packets or Echo reply packets) in a very short amount of time [11]. On getting the alarm, traffic coming from a particular local network under the connection the fog node which raised the alarm can be dumped or blocked to keep the servers safe. The sensors also known as edge nodes are not very secure because of their simple design and application which makes them vulnerable to a lot of cyber attacks. Cyber criminals use these kinds of compromised nodes as bots to launch various kinds of distributed attacks usually on single points of failures in the networks. But the non-centralized nature of the fog empowered networks prevents these edge nodes from getting exposed in the public network and therefore improves security of the network [12].

2.2. Anomaly detection

Anomaly detection can be defined as the process of identifying abnormal or unexpected patterns within a dataset. Abnormal or unexpected patterns are basically just data points which differ from the normal trend of the dataset as a whole. For e.g. in water transportation pipelines, a particular range of pressure is important to be maintained and anything above or below that range can cause critical situations. In this case if the pressure shoots very high suddenly or maybe falls unexpectedly then it can be considered as an abnormal behaviour or an anomaly as compared to a normal scenario. These kinds of anomalies are introduced in the system either by accident or by human mistake. But there are also some cases where this is done by someone intentionally. For e.g. in cyber attacks, an attacker tries to compromise a node or some part of a network which creates anomalous behaviour in network traffic or communication.

As it can be inferred from the above mentioned examples that an anomaly is generally not desired. It is a rare situation but it can still lead to critical scenarios. So there is a need for anomaly detection mechanisms which can detect these abnormal behaviours in the system and warn the concerned authorities to take necessary actions. [13–15] Researchers have proposed many effective anomaly detection techniques based on machine learning, neural networks, statistical analysis etc. These anomaly detection algorithms usually train on unsupervised datasets and try to find the pattern or we can say distribution of the normal trend then any new data point is compared to this pattern that the algorithm has extracted out. If the new data fits in the trend and follows the similar distribution then the algorithm classifies it as normal data otherwise it is classified as an anomaly. The comparison of the data point to the distribution can be based on various metrics [16]. One such metric is called continuous ranked probability score (CRPS) which is used for probabilistic forecasting. We will discuss this metric in detail in next section.

3. Proposed approach

In this work, we are proposing a 3-step anomaly detection framework to detect DDoS attacks in fog networks. Fig. 1 shows the steps involved in the process starting from gathering data from the network traffic to finally detecting the attack. In our approach, we are using a statistical algorithm which works on features extracted from the incoming traffic. These features are first passed in a PCA component before putting the values in the algorithm. This is done because the algorithm used is single variate.

First step of the process is called the data preprocessing step where packets are bundled into time bound windows and then features like number of packets, average time between the packets, number of source IPs etc. can be extracted depending on the network state and requirements. Second step is a Principal Component Analysis(PCA) step where these features are transformed into principal components and the component with the maximum variance along its axis is selected as the main variable to be passed onto the next step.

The final step is the detection of anomalies present in the dataset. This step uses a statistical metric called CRPS which is used to calculate the difference in the distribution of testing data and training data. This difference can indicate the measure of anomaly in the network traffic,

3.1. Data preprocessing

The proposed mechanism deals with time series data which means real-time sequential flow of packets moving from a few source ip addresses to some other destination ip addresses in a network. When these packets pass through the monitoring fog nodes, traffic information is gathered usually for network analysis, security check-ups, traffic behaviour monitoring etc. In our proposed architecture, packet information such as time-stamp, source IP, destination IP, type, packet length etc. is collected. These packets i.e the collected data entries, are then bundled in the groups or we can say windows of 60 s each based on the time-stamps. This is called the Data-Aggregation step. The window size can be changed depending on the network traffic.

This time window is nothing but a hyper-parameter which can be tuned as per the need of the setup in which the algorithm is being tested. But, this tuning needs to be done keeping a few things in mind like the frequency of packets and average amount of packets getting transmitted in the attack free environment before allowing real world data that might allow an attack because if the average data load is high or we can call the throughput of the network is usually high and time window is too small, then it might become too sensitive to the normal increase in the number of packets. Similarly if the throughput is low and the time window is too large then it might miss the sudden impulses or increase in packet frequency which might be a part of a distributed attack.

In developing anomaly detection mechanisms for issues like DDoS attacks, extracting valuable features from the network or the traffic that can be exploited to detect suspicious behaviour of edge nodes or unexpected incoming network traffic is a challenging task. In most of the recently proposed time-series based DDoS detection mechanisms, the number of packets passing through the monitoring point within a fixed time limit or a time window, is used as a feature to detect attacks like TCP-SYN attacks or smurf attacks etc. This can give a rough idea about the network traffic and its behaviour but this feature is not sufficient enough to detect all the hidden patterns in the data. There can be many cases where the packet flow per time window is not exceptionally high but there are a few nodes which are consuming abnormal amount of energy from the network or if in case of DDoS attacks particularly, there can be few instances where some amount of packets were transmitted from a bunch of source node to a target destination node in a very short amount of time or even almost at the same time. Now this amount of packet flow can be detected as slightly anomalous if the total packet count within the time window is significantly large otherwise it will be considered as normal traffic. Whereas these impulses of packets from a group of different networks can accumulate and flood the target node.

Therefore, anomaly detection mechanisms based on multiple variables are more reliable and accurate. In this work, Along with the number of packets passing in the network within a time window, which is a common feature used by many researchers, the average time between the subsequent packers within the time window is used as an another parameter to tackle the above discussed issue of many nodes flooding the target by sending very high frequency of packets in a very short period of time. This another feature is important because this provides a new perspective to the situation. Suppose there is a window of say 60 s, in one case there are 100 packets passing through the network within these 60 s and the flow of packets is uniformly distributed throughout this time window, and in second case, there are no packets within first 30 s and then suddenly there are 100 packets

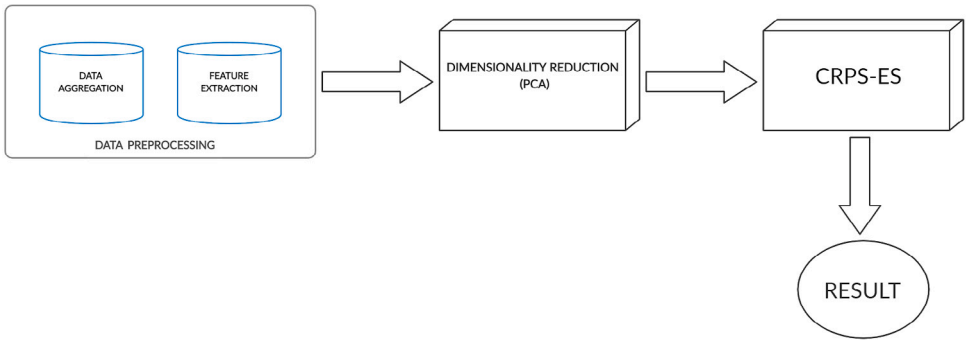


Fig. 1. The Proposed Architecture.

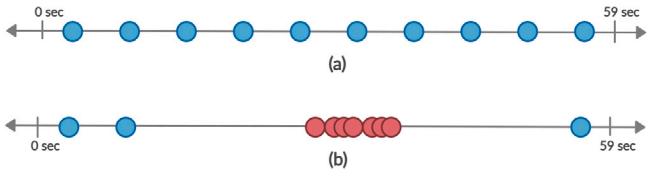


Fig. 2. Difference in the average time between the packets given the number of packets is same.

within 1 s, from 30th to the 31st second and then again no packets till 60th second. Both these cases are showing a very different behaviour though the number of packets is the same. The only difference is the average time between the packets that can project that the flow in the second case is anomalous (see Fig. 2).

The calculation of both the above discussed features for every time window is performed in the Feature Extraction step which also concludes the Data Preprocessing. After that a trend based on the combined effect of the extracted features is calculated in the dimensionality reduction component which uses principal component analysis or PCA. We have discussed more about this in the next section.

3.2. Principal component analysis

Dimensionality reduction includes transformation of data from higher dimension to a lower dimension where the lower dimension features extract useful information from the original data. In our case we required dimensionality reduction because there was a need to transform multi-dimensional feature data to a single dimensional data as CRPS accepts data with single feature. This way we combine the effect of multiple features and the most relevant information out of all the features is passed to the statistical algorithm. PCA was a perfect fit for our algorithm because it tries to extract information on the basis of maximum variance as the basis [17]. Selecting information on the basis of maximum variance will help us to extract varying information while ignoring the information not required.

As discussed , we are using two variables for our algorithm. We have used PCA to convert the two variables into a single variable to meet our requirements. Time series based anomaly detection algorithms usually consider a single variable but using PCA we can produce a single variable which is a combination of variations along different metrics which can obviously be considered a better choice for anomaly detection considering the fact that the IoT traffic is heterogeneous in nature.

There are certain conditions that have to be met to apply PCA. So it is very important to show the correlation between different variables that we have considered in the previous section. We will be using graphical results to explain the correlation in the Evaluation section.

3.3. Continuous ranked probability score

CRPS or continuous rank probability score is a statistical metric which is commonly used in probabilistic forecasting models. In our work, we have selected this metric because CRPS can be used to quantify the dissimilarity between a new observation and the attack free traffic distribution [18].

This makes CRPS a better suited algorithm for real time detection as compared to other similar metrics like KL divergence or even X2(chi square) which requires the whole data beforehand to compute distributions of training and testing data [13]. At the time of testing, the difference between the CRPS values of the incoming data points and the attack free normal data can identify the abnormal behaviour of the network traffic [7].

Let X be the random variable. Let F be the cumulative distribution function (CDF) of X, such as $F(y) = P[X \leq y]$. Let x be the observation, and F the CDF associated with an empirical probabilistic forecast. The CRPS between x and F is defined as Eq. Eq. (1) in Algorithm 1, where $1(x)$ is the Heaviside step function and denotes a step function along the real line that attains, first, the value of 1 if the real argument is positive or zero, and second, the value of 0 otherwise.

The CRPS is expressed in the same unit as the observed variable. Eq. Eq. (1) represents the measure of the difference between the underlying distribution of the given attack free data and the incoming observation data. For our case we know that the attack free data is normally distributed. Now, when we substitute the function F as a normal distribution we get Eq. Eq. (2). Therefore, Eq. Eq. (2) is just an extension of Eq. Eq. (1).

To enhance the detection efficiency of CRPS we apply Exponential smoothing to it (CRPS-ES). This is done for the inclusion of previous and current information in the decision process which helps it in uncovering even small anomalies and makes the statistic less sensitive towards noise.

Usually the distribution of data is assumed to be gaussian but here we try to figure out a more realistic distribution for the data using the Kernel Density Estimation (KDE) [19]. CRPS-ES considers a non-parametric decision threshold which is computed by analysing the flow of the underlying distribution. The major plus point of CRPS-ES metric is that it considers the past data in the detection which makes it sensitive to apparent attacks. [20] It is more suitable for real time detection as the new traffic that might contain anomalies is compared to the attack free traffic (see Fig. 3).

4. Evaluation

We have used a very popular dataset used for Wireless Sensor Networks to test our algorithm which is DARPA99 dataset. It is an anomaly detection evaluation dataset generated in 1999 by Lincoln Laboratory at MIT.

This dataset consists of data packets passing through a network over a period of 5 weeks. The first 3 weeks are a part of the training

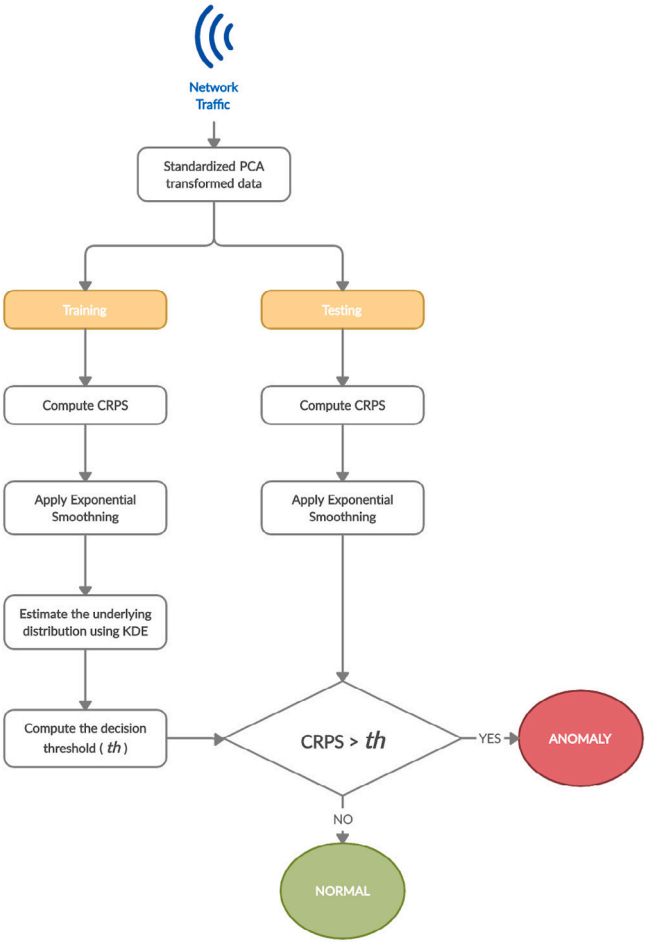


Fig. 3. Flowchart for Anomaly Detection.

dataset with various labelled information and many different types of evaluations and observations most of which are out of the scope of this work. The last 2 weeks are for testing where the many cyber attacks like TCP-SYN, ICMP etc. were launched on the network under observation.

Wireshark is a tool used for analysing data packets sent from one IP address to another. We have used Wireshark to extract out important information from the DARPA 99 dataset and use it for our experiment purpose.

As depicted in above snapshot packets in DARPA 99 contain a timestamp, Source IP, destination IP, Protocol used, Length of the message and some information about. For a particular type of attack the packets with the same type of protocol used are filtered out. For e.g. For, TCP SYN attack packets with TCP protocol are filtered out. Wireshark provides an inline command facility to filter out packets.

Here for our experiment we have used Week 1 Day 1 as an attack free traffic as there is no attack on Day 1 Week 1. For detection purposes(testing data) we have used Week 5 Day 1 as it contains a TCP SYN attack initiated once for a duration of 6 min 51 s and for ICMP Week 4 Day 1 is used it has 2 attacks for 1s each (see Figs. 4–16).

4.1. Observations for training data for TCP-SYN attack

See Figs. 4–7.

4.2. Results for testing data for TCP-SYN attack

See Figs. 8–10.

Algorithm 1: Training Algorithm

1: In this phase we will have a preprocessed data with PCA applied to it. We will use the CRPS metric on this data. CRPS is helpful in quantifying the deviations from attack free data which help us in detecting anomalies. For a observation x, the CRPS value is calculated as

$$CRPS(F,x) = \int_{-\infty}^{\infty} (F(y) - 1(y \geq x))^2 dy \quad (1)$$

where 1(y ≥ x) :

$$1(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases} \quad (2)$$

It should be noted that when the traffic is attack free its distribution is gaussian. So, the CRPS for function F as gaussian will be :

$$CRPS(N(\mu,\sigma^2),x) = \sigma[\frac{x-\sigma}{\sigma}(2\phi(\frac{x-\sigma}{\sigma})-1)+2\phi(\frac{x-\sigma}{\sigma})-1/\sqrt{\pi}] \quad (3)$$

2: Now the exponential smoothing is applied. This is done for the inclusion of previous and current information in the decision process which helps it in uncovering even small anomalies and makes the statistic less sensitive towards noise. The mathematical formula is calculated as

$$z_t^{CRPS} = \nu d_t + (1-\nu) z_{t-1}^{CRPS} \quad (4)$$

Where, z_t is the calculated value of the current data point, z_{t-1} is the calculated value of previous data , d_t is the observation of the current data point and ν is the forgetting parameter.

3: At this step, we have the calculated CRPS-ES values for the training data. Now, a non-parametric threshold (th) value needs to be calculated to detect the anomalies in the testing phase. This threshold should be able to accurately justify the results i.e. if a value is crossing the threshold, it must differ from the underlying distribution of the training CRPS-ES statistic. This distribution is calculated using Kernel Density Estimation. The threshold is calculated as the $(1-\alpha)$th quantile of the calculated distribution.

Algorithm 2: Testing Algorithm

1: In this phase we will have a preprocessed data with PCA applied to it similar to the training phase. CRPS is applied similar to the training phase.
2: Then Exponential smoothing is applied and CRPS_ES value is obtained.
3: The output of step 2 is compared to the threshold obtained from the training phase.
4: **if** CRPS_ES > th **then**
5: Anomaly Present
6: **else**
7: Normal Traffic
8: **end if**

4.3. Observations for training data for ICMP attack

See Figs. 11–13.

4.4. Results for testing data for ICMP attack

See Figs. 14–16.

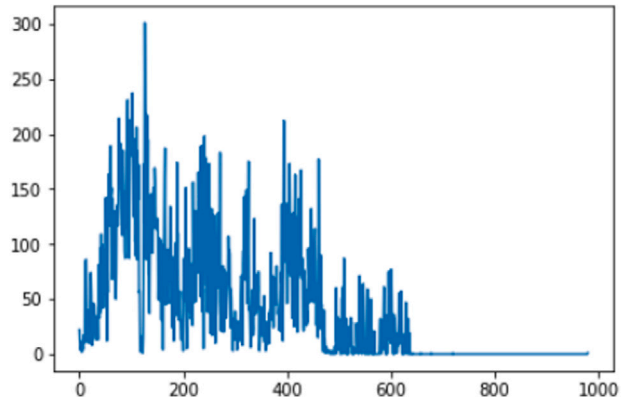


Fig. 4. First Feature: the number of packets passing through the network per 75 s window. Here, x-axis: the index of windows and y-axis: the first feature values.

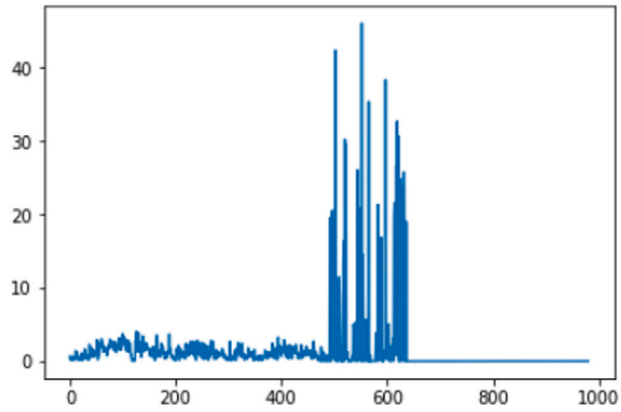


Fig. 5. Second Feature: an inverse function of average time between the packets within the 75 s window. Here, x-axis: the index of windows and y-axis: the second feature values.

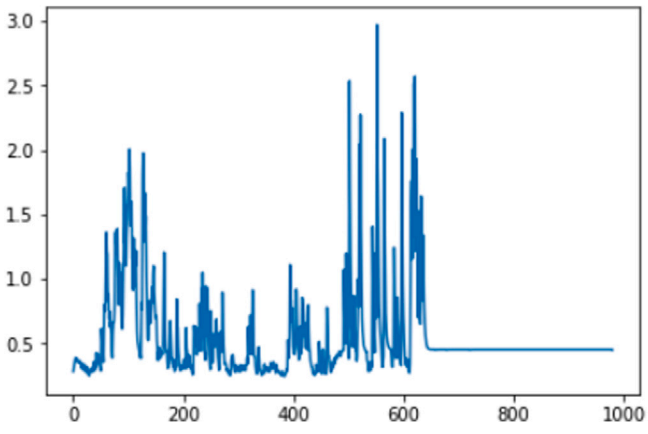


Fig. 6. CRPS-ES values for training data after applying PCA. Here, x-axis: the index of windows and y-axis: the CRPS-ES values.

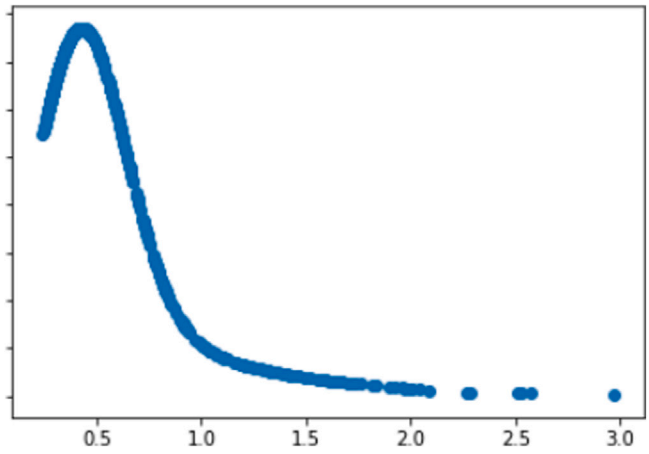


Fig. 7. Distribution of the CRPS-ES calculated using KDE.

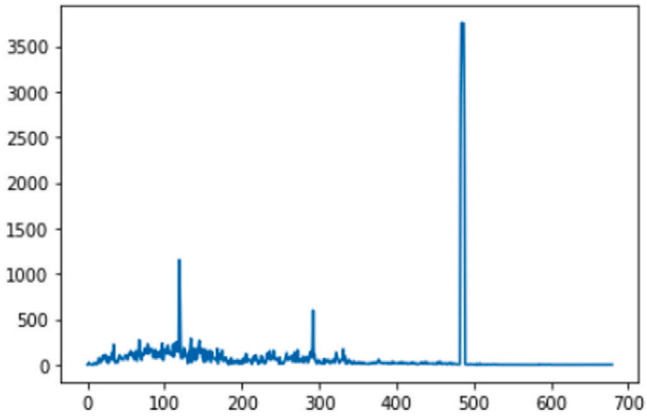


Fig. 8. First Feature: the number of packets passing through the network per 75 s window. Here, x-axis: the index of windows and y-axis: the first feature values.

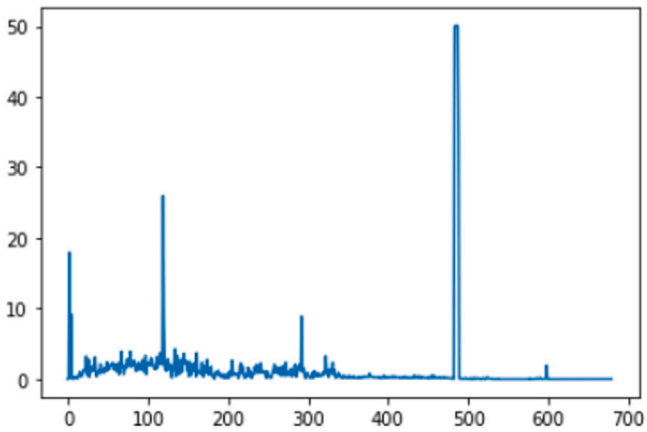


Fig. 9. Second Feature: an inverse function of average time between the packets within the 75 s window. Here, x-axis: the index of windows and y-axis: the second feature values.

5. Conclusion

We were able to detect all the TCP-SYN and ICMP attacks present in the dataset used using fast statistical algorithms as compared to deep learning mechanisms which are computationally very expensive and are not feasible for edge and fog nodes. Other protocol based attacks include UDP attacks. This algorithm can work on most of the protocol based attacks. The common idea behind all protocol based attacks remains the same, that is analysing the incoming traffic based on some window size, reducing its dimensions and applying the CRPS-ES metric. We were also able to show the importance of fog computing in the anomaly detection frameworks as it provides local monitoring of the network traffic which can help us in detection of various attacks at the lowest level. Our aim here was to highlight the importance of multiple features. in time series based algorithms. We have demonstrated why

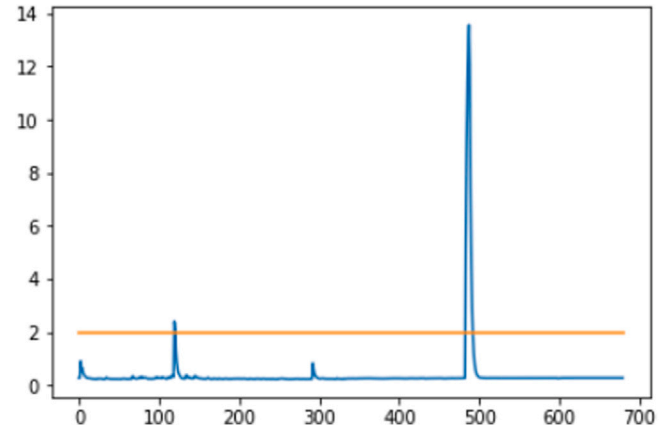


Fig. 10. CRPS-ES values for training data after applying PCA. Here, x-axis: the index of windows and y-axis: the CRPS-ES values. The orange line represents the threshold CRPS-ES value (th = 1.97, $(1-\alpha) = 0.99$). The peaks in the figure are the attack points with CRPS-ES values > th. It can be observed in the figure that all the attacks are successfully detected by the proposed algorithm.

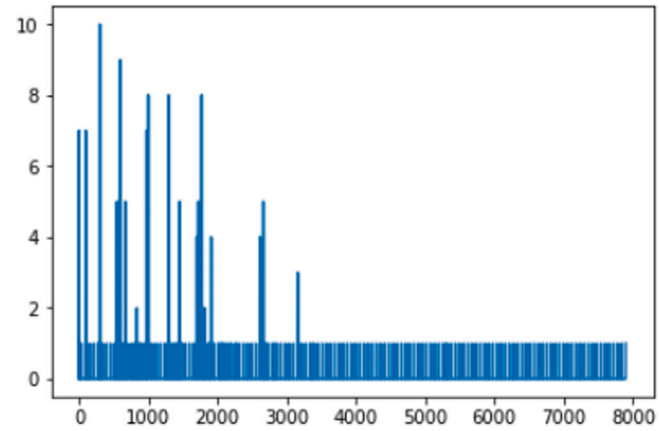


Fig. 11. First Feature: the number of packets passing through the network per 75 s window. Here, x-axis: the index of windows and y-axis: the first feature values.

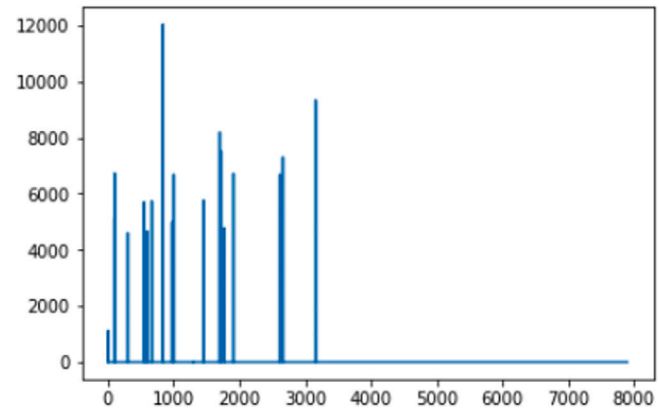


Fig. 12. Second Feature: an inverse function of average time between the packets within the 75 s window. Here, x-axis: the index of windows and y-axis: the second feature values.

we should use multiple features and how we can extract multiple features from the network. Use of multiple features plays an important role in detecting hidden attacks and exposing various unknown trends present in the data.

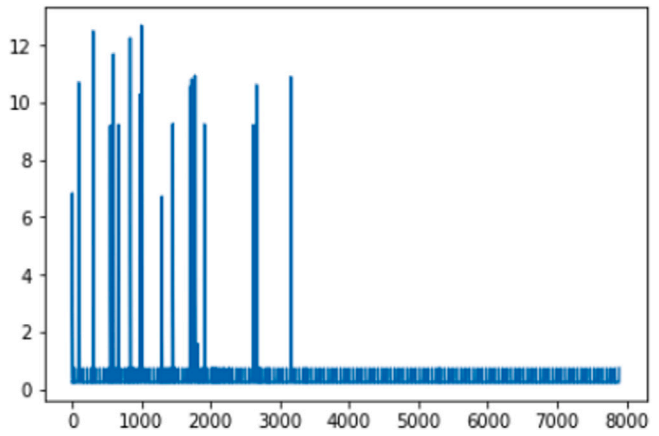


Fig. 13. CRPS-ES values for training data after applying PCA. Here, x-axis: the index of windows and y-axis: the CRPS-ES values.

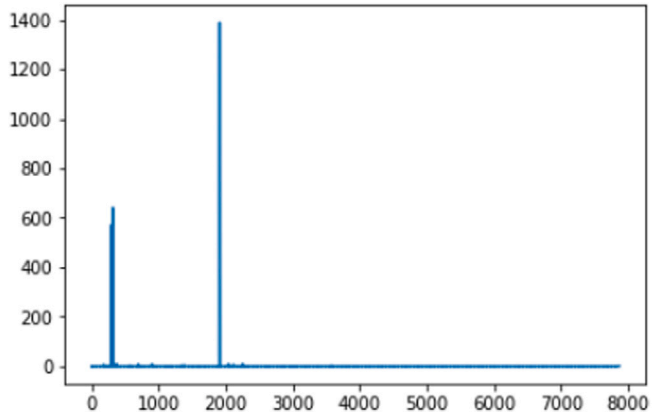


Fig. 14. First Feature: the number of packets passing through the network per 75 s window. Here, x-axis: the index of windows and y-axis: the first feature values.

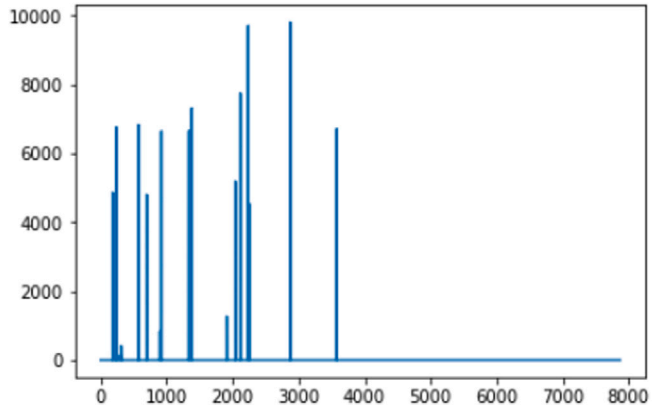


Fig. 15. Second Feature: an inverse function of average time between the packets within the 75 s window. Here, x-axis: the index of windows and y-axis: the second feature values.

But, CRPS is a single variable algorithm which is a negative point of the algorithm, though there exists a multi variable CRPS statistical method but it is not very easy to implement. Therefore , an approach was devised to incorporate multiple variables in the CRPS algorithm by using dimensionality reduction technique. But this leads to some loss of information as well which is a shortcoming of the model as there is a limit on the number of features that can be reduced to a single variable.

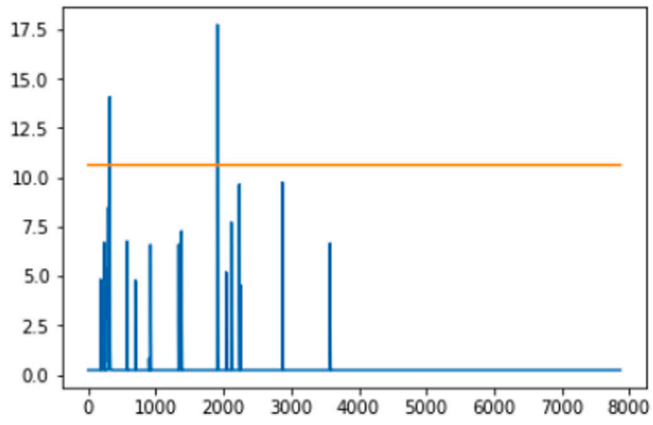


Fig. 16. Here, x-axis: the index of windows and y-axis: the CRPS-ES values. The orange line represents the threshold CRPS-ES value (th = 10.62, $(1-\alpha) = 0.99$). The peaks in the figure are the attack points with CRPS-ES values > th. It can be observed in the figure that all the attacks are successfully detected by the proposed algorithm.

In terms of objective comparative studies, previous models were also suitable as researchers were able to detect obvious attacks present in the data but if we consider niche situations and scenarios then it is clearly visible through the observations and results that our model is capturing more information from the network traffic and is more robust.

Our approach is modular as well because hyper-parameters can be tweaked and various set of features can be used depending on the use case and state of the network.

In future works, this architecture can be made more accurate and useful by,

1. Including more network parameters such as energy consumption, number of source IPs etc., as features in the anomaly detection model.
2. Using multivariate CRPS components because this can increase the capability of extracting network features and using them in analysing the data trends by manifolds.
3. Identification of the malicious nodes present in the network. In this work we have detected the anomaly present in the data but finding the source of the anomaly data is still a very challenging task that we would like to tackle in our future works.
4. Exploring the capabilities of Fog computing and figuring out how we can leverage it to enhance the efficiency of our model to the maximum.

Declaration of competing interest

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

References

[1] Bashar Ahmad Khalaf, Salama A. Mostafa, Aida Mustapha, Mazin Abed Mohammed, Moamin A. Mahmoud, Bander Ali Saleh Al-Rimy, Shukor Abd Razak, Mohamed Elhoseny, Adam Marks, An adaptive protection of flooding attacks model for complex network environments, Secur. Commun. Netw. 2021 (2021) 17, http://dx.doi.org/10.1155/2021/5542919, Article ID 5542919.

[2] Adnan Helmi A. zizan, Salama A. Mostafa, Aida Mustapha, Cik Feresa Mohd Foozy, Mohd Helmy Abd Wahab, Mazin Abed Mohammed, Bashar Ahmad Khalaf, A machine learning approach for improving the performance of network intrusion detection systems, Ann. Emerg. Technol. Comput. (AETiC) 5 (5) (2021) 201–208, http://dx.doi.org/10.33166/AETiC.2021.05.025, Print ISSN: 2516-0281, Online ISSN: 2516-029X, Published by International Association of Educators and Researchers (IAER), Available: http://aetic.theiaer.org/archive/v5/v5n5/p25.html.

[3] Shafiq Muhammad, Zhihong Tian, Ali Bashir, Alireza Jolfaei, Data mining and machine learning methods for sustainable smart cities traffic classification: A survey, Sustainable Cities Soc. 60 (2020) http://dx.doi.org/10.1016/j.scs.2020.102177.

[4] R.M.A. Ujjan, Z. Pervez, K. Dahal, et al., Towards sFlow and adaptive polling sampling for deep learning based DDoS detection in SDN, Future Gener. Comput. Syst. (2019) http://dx.doi.org/10.1016/j.future.2019.10.015.

[5] Yonghao Gu, Yongfei Wang, Zhen Yang, Fei Xiong, Yimu Gao, Multiple-features-based semi supervised clustering DDoS detection method, Math. Probl. Eng. 2017 (2017) 10, http://dx.doi.org/10.1155/2017/5202836, Article ID 5202836.

[6] S. Sumathi, N. Karthikeyan, Detection of distributed denial of service using deep learning neural networks, J. Ambient Intell. Hum. Comput. (2020) http://dx.doi.org/10.1007/s12652-020-02144-2.

[7] F. Harrou, B. Bouhaddou, Y. Sun, B. Kadri, Detecting cyberattacks using a CRPS-based monitoring approach, in: 2018 IEEE Symposium Series on Computational Intelligence, SSCI, 2018, Available: http://dx.doi.org/10.1109/SSCI.2018.8628797.

[8] S. Madakam, R. Ramaswamy, S. Tripathi, Internet of things (IoT): A literature review, J. Comput. Commun. 3 (2015) 164–173, http://dx.doi.org/10.4236/jcc.2015.35021.

[9] D. Myridakis, G. Spathoulas, A. Kakarountas, D. Schoinianakis, J. Lueken, Anomaly detection in IoT devices via monitoring of supply current, in: 2018 IEEE 8th International Conference on Consumer Electronics, Berlin (ICCE-Berlin), 2018, pp. 1–4, http://dx.doi.org/10.1109/ICCE-Berlin.2018.8576178.

[10] A.A. Mutlag, M. Khanapi Abd Ghani, M.A. Mohammed, M.S. Maashi, O. Mohd, S.A. Mostafa, K.H. Abdulkareem, G. Marques, I. de la Torre Díez, MAFC: Multi-agent fog computing model for healthcare critical tasks management, Sensors 20 (2020) 1853, http://dx.doi.org/10.3390/s20071853.

[11] V. Hassija, V. Chamola, V. Saxena, D. Jain, P. Goyal, B. Sikdar, A survey on IoT security: Application areas, security threats, and solution architectures, IEEE Access 7 (2019) 82721–82743, http://dx.doi.org/10.1109/ACCESS.2019.2924045.

[12] Shanhe Yi, Cheng Li, Qun Li, A survey of fog computing: Concepts, applications and issues, in: Proceedings of the 2015 Workshop on Mobile Big Data, Mobidata '15, Association for Computing Machinery, New York, NY, USA, 2015, pp. 37–42, http://dx.doi.org/10.1145/2757384.2757397.

[13] I. Alrashdi, A. Alqazzaz, E. Aloufi, R. Alharthi, M. Zohdy, H. Ming, AD-IoT: Anomaly detection of IoT cyberattacks in smart city using machine learning, in: 2019 IEEE 9th Annual Computing and Communication Workshop and Conference, CCWC, 2019, pp. 0305–0310, http://dx.doi.org/10.1109/CCWC.2019.8666450.

[14] Zhongguo Yang, Irshad Ahmed Abbasi, Elfatih Elmubarak Mustafa, Sikandar Ali, Mingzhu Zhang, An anomaly detection algorithm selection service for IoT stream data based on tsfresh tool and genetic algorithm, Secur. Commun. Netw. 2021 (2021) 10, http://dx.doi.org/10.1155/2021/6677027, Article ID 6677027.

[15] A.A. Cook, G. Mısırlı, Z. Fan, Anomaly detection for IoT time-series data: A survey, IEEE Internet Things J. 7 (7) (2020) 6481–6494, http://dx.doi.org/10.1109/JIOT.2019.2958185.

[16] G. Spanos, K.M. Giannoutakis, K. Votis, D. Tzovaras, Combining statistical and machine learning techniques in IoT anomaly detection for smart homes, in: 2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks, CAMAD, 2019, pp. 1–6, http://dx.doi.org/10.1109/CAMAD.2019.8858490.

[17] T. Jolliffe Ian, Cadima Jorge, Principal component analysis: a review and recent developments, Phil. Trans. R. Soc. A 374 (2016) http://dx.doi.org/10.1098/rsta.2015.0202, 2015020220150202.

[18] B.A. Khalaf, S.A. Mostafa, A. Mustapha, M.A. Mohammed, W.M. Abduallah, Comprehensive review of artificial intelligence and statistical approaches in distributed denial of service attack and defense methods, IEEE Access 7 (2019) 51691–51713, http://dx.doi.org/10.1109/ACCESS.2019.2908998.

[19] Weglarczyk Stanislaw, Kernel density estimation and its application, in: ITM Web of Conferences, Vol. 23, 2018, p. 00037, http://dx.doi.org/10.1051/itmconf/20182300037.

[20] Tilmann Gneiting, Adrian E. Raftery, Strictly proper scoring rules, prediction, and estimation, J. Amer. Statist. Assoc. 102 (477) (2007) 359–378, http://dx.doi.org/10.1198/016214506000001437.

Deepak Kumar Sharma is working as an Assistant Professor in the Department of Information Technology, Netaji Subhas University of Technology (Formerly NSIT), Dwarka, New Delhi, India. He obtained his Ph.D. in Computer Engineering from University of Delhi, India in 2016. His research interests include opportunistic networks, wireless ad hoc and sensor networks, Software Defined Networks and IoT Networks. He has over 15 years of experience in Academics. He has published various research papers in reputed international journals like ETT Wiley, IEEE Systems Journal, Computer Communication Elsevier, IJCS Wiley etc. and conferences of repute like IEEE AINA, GLOBECOM etc. He has also authored various book chapters in edited books of IET, Wiley, Springer, Elsevier etc. He is also a reviewer of various reputed journals like ETT Wiley, AIHC Springer, IJCS Wiley etc.

Tarun Dhankhar is currently pursuing is undergraduate Bachelor's degree in Information Technology from Netaji Subhas Institute of Technology, New Delhi, India. His research interests are in the areas related to Fog Computing, the Internet of Things and the application of Statistical techniques in these areas. His specific interests are the application of Information Security in these research areas. Previously He has worked in the fields of Machine Learning and Deep Learning as well.

Gaurav Agrawal is currently pursuing his undergraduate Bachelor's degree in Information Technology from Netaji Subhas Institute of Technology, New Delhi, India. His research interests are in the areas related to the Internet of Things, Fog Computing and the application of Statistical techniques in these areas. His specific interests include topics like Machine Learning, Deep Learning.

Satish Kumar Singh is working as an Assistant Professor in the Department of Information Technology, Netaji Subhas University of Technology (Formerly NSIT), Dwarka, New Delhi, India. He is currently pursuing his Ph.D. in Information Technology from Netaji Subhas University of Technology, Delhi, India. He has obtained M.Tech Integrated degree in Information Technology from University School of Information Technology, GGSIPU, New Delhi. His research interests include Cloud Computing, Fog Computing, Wireless ad hoc and sensor networks, Software Defined Networks and IoT. He has over 10 years of experience in Academics. He has published various research papers in conferences of repute like IEEE, etc.

Deepak Gupta received a B.Tech. degree in 2006 from the Guru Gobind Singh Indraprastha University, India. He received M.E . degree in 2010 from Delhi Technological University, India and Ph. D. degree in 2017 from Dr. APJ Abdul Kalam Technical University, India. He has completed his Post-Doc from Inatel, Brazil. With 13 years of rich expertise in teaching and two years in the industry; he focuses on rational and practical learning. He has contributed massive literature in the fields of Intelligent Data Analysis, BioMedical Engineering, Artificial Intelligence, and Soft Computing. He has served as Editor-in-Chief, Guest Editor, Associate Editor in SCI and various other reputed journals (IEEE, Elsevier, Springer, & Wiley). He has actively been an organizing end of various reputed International conferences. He has authored/edited 50 books with National/International level publishers (IEEE, Elsevier, Springer, Wiley, Katson). He has published 186 scientific research publications in reputed International Journals and Conferences including 99 SCI Indexed Journals of IEEE, Elsevier, Springer, Wiley and many more. He has also filled/published 4 patents.

Jamel Nebhen received the M.Sc. in Microelectronics from the National Engineering School of Sfax, Tunisia in 2007, and the Ph.D. degrees from the Aix-Marseille University, France, in 2012, all in Microelectronics. From 2012 to 2018, he worked as a Postdoctoral Researcher in France in LIRMM-Lab Montpellier, IM2NP-Lab Marseille, ISEP Paris, LE2I-Lab Dijon, Lab-Sticc Telecom Bretagne Brest, and IEMN-Lab Lille. Since 2019, he joined the Prince Sattam bin Abdulaziz University in Alkharj, Saudi Arabia, as an Assistant Professor. His research interests are mainly in the design of analog and RF integrated circuits, IoT, biomedical circuit, and sensors instrumentation.

Imran Razzak is currently Sr. Lecturer at School of Information Technology at Deakin University, Australia. Imran's research interest is machine learning and data analytics in general, particularly in healthcare industry. He is a passionate health informatician who wants to make the healthcare industry a better place through informatics. He has published more than 70 refereed research publications in international journals and conferences. He is an editorial board member of many reputable international journals as well as session co-chair, session chair and TPC member of dozens of conferences. He is also working as a consultant in various projects involving deep learning models in Big Data, IoT, medical Imaging and BCI applications.

Application of Swarm-Based Feature Selection and Extreme Learning Machines in Lung Cancer Risk Prediction

Priya Garg*, Deepti Aggarwal

Department of Computer Science, Delhi Technological University, Bawana Road, Delhi, India

*Corresponding author
doi: https://doi.org/10.21467/proceedings.115.1

ABSTRACT

Lung cancer risk prediction models help in identifying high-risk individuals for early CT screening tests. These predictive models can play a pivotal role in healthcare by decreasing lung cancer's mortality rate and saving many lives. Although many predictive models have been developed that use various features, no specific guidelines have been provided regarding the crucial features in lung cancer risk prediction. This study proposes novel risk prediction models using bio-inspired swarm-based techniques for feature selection and extreme learning machines for classification. The proposed models are applied on a public dataset consisting of 1000 patient records and 23 variables, including sociodemographic factors, smoking status, and lung cancer clinical symptoms. The models, validated using 10-fold cross-validation, achieve an AUC score in the range of 0.985 to 0.989, accuracy in the range of 0.986 to 0.99 and F-Measure in range of 0.98 to 0.985. The study also identifies smoking habits, exposure to air pollution, occupational hazards and some clinical symptoms as the most commonly selected lung cancer risk prediction features. The study concludes that the developed lung cancer risk prediction models can be successfully applied for early screening, diagnosis and treatment of high-risk individuals.

Keywords: Lung Cancer Risk Prediction, Machine Learning, Extreme Learning Machine, Swarm-based Techniques, Feature Selection

1 Introduction

According to WHO [1], lung cancer is the leading cause of cancer deaths worldwide, with around 2.09 million cases and 1.76 million deaths in 2018. In 2011, The National Lung Screening Trial (NLST) found that low-dose helical computed tomography (CT) can reduce the risk of lung cancer deaths by 15% to 20 % [2]. Although these screening tests can lower lung cancer mortality, they are only beneficial if given to the right target individuals who have a high risk of developing lung cancer. The reason is that the screening process involves exposure to radiations, invasive diagnostic procedures and false positives [3].

In recent years, lung cancer risk prediction models have been developed for identifying the individuals who are at a high risk of developing lung cancer in the future using various variables like socio-demographic factors, socio-economic factors, smoking history, family history of lung cancer, biomarkers and lung cancer clinical symptoms [4]. Such models help early diagnosis and treatment as they identify the target population for low-dose computed tomography (LDCT) screening. These models can increase the CT process's overall cost-effectiveness while minimizing the associated effects of harmful CT-radiations [3, 5]. Many lung cancer risk prediction models have been developed by past studies using different classification techniques, independent variables and validation procedures [3-6]. Although these prediction models achieve good performance, there

© 2021 Copyright held by the author(s). Published by AIJR Publisher in the "Proceedings of Intelligent Computing and Technologies Conference" (ICTCon2021) March 15th–16th, 2021. Jointly organized by Assam Science and Technology University (ASTU), and Central Institute of Technology Kokrajhar (CITK).

Proceedings DOI: 10.21467/proceedings.115; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-947843-5-7

are certain sources of bias due to improper variable selection, lack of model validation and poor research design [3]. Additionally, different sets of independent variables have been used for the model development, but there is little analysis regarding which factors play an essential role in determining lung cancer risk.

Specific feature selection techniques have been applied to study and identify the significant variables in lung cancer prediction. Xie et al. [7] applied a combination of ReliefF and Fisher for feature selection, followed by BP Neural Networks, to classify lung cancer data and control health data. They determined 15 top risk factors from the 41 factors for lung cancer risk prediction and achieved an accuracy of 83.816%. Dallri [8] proposed a model using Genetic Algorithms (GA) for feature selection and fuzzy Extreme Learning Machine (ELM) for classification for lung cancer prediction and achieved an accuracy of 97.5%. Chen et al.[9] used Deep Neural Networks to identify the risk factors for lung cancer incidence in older people. The study found that cancer history, smoking-related factors and physical exercise level were essential factors and achieved an accuracy ranging from 0.927 to 0.962. Markaki et al. [10] proposed a lung cancer screening model using backward feature selection and Cox regression. The model identified seven risk factors for lung cancer risk prediction and achieved an externally validated AUC score of 0.87 within six years.

This study proposes novel lung cancer prediction models that use bio-inspired swarm-based techniques for feature selection, followed by Extreme Learning Machines for classification. Bio-inspired techniques are often used for optimal feature subset selection, but such techniques have not been employed in this research field. These techniques mimic real-life phenomena and are known for their global search capability and low computational time compared to exhaustive search [11]. The extensive use of these techniques in other medical fields [12, 13] motivated us to use these in lung cancer risk prediction. ELMs are randomization based neural network techniques that show the predictive power of deep neural networks while minimizing the learning time and large dataset requirements [14]. ELMs have been used as the classification technique in lung cancer prediction in previous studies and have shown high accuracy and generalization capability, which motivated us to use ELM as the classification technique in this study.

The major contributions of the study in the field of lung cancer risk prediction are:

- The study investigates the importance of feature selection in the improvement of risk prediction. This study also presents a comparison between various bio-inspired feature selection techniques.
- The results of the study also identify the most important features that are effective and contribute the most towards prediction.

The rest of the paper has the following sections. Section 2 describes the three bio-inspired swarm-based techniques (Particle Swarm Optimization, Artificial Bee Colonization and Firefly algorithm) used for feature reduction, Section 3 explains ELM, which is the classification technique, Section 4 presents the experimental setup, Section 5 shows the experimental results and observations, Section 6 concludes the study and possible future extensions.

2 Bio-Inspired Swarm Swarm-Based Techniques for Feature Selection

In this study, bio-inspired swarm-based techniques have been used for effective feature selection. These techniques use the properties and behaviour of biological forms like optimal food source location, luminescence attraction etc., to search for the optimized solution in the problem space. In this study, three such techniques have been used - Particle Swarm Optimization, Firefly algorithm, and Artificial Bee Colonization. These techniques have been further explained in the following subsections.

Proceedings DOI: 10.21467/proceedings.115
ISBN: 978-81-947843-5-7

Series: AIJR Proceedings
ISSN: 2582-3922

2.1 Particle Swarm Optimization

PSO [15] is a swarm-based optimization technique based on bird flocks' behaviour, which tends to optimize their position according to the food source available. For feature reduction problems, every particle in the swarm denotes a vector of size m, where m is the number of features or independent variables in the dataset. The algorithm optimizes each particle's position to get the highest fitness value which in this study is the AUC-ROC value achieved by selecting the features that are represented in the position vector of each particle. The pseudo-code of PSO is explained in Figure 1.

Algorithm 1: Pseudo Code of PSO

Initialization:

1. *Number of particles:* P
2. *Threshold:* θ
3. *Initial position of each particle:*
   Pos = {pos_i}, i= 1,2,3...P and pos_i is an m-dimensional vector.
4. *Initial Velocity of each particle:*
   V = {v_i} i= 1,2,3...P and v_i is an m-dimensional vector
5. pbest = [$pbest_i$], i= 1,2,3...P and $pbest_i$ is a m-dimensional vector
6. gbest = {0} m-dimensional vector
7. maxiteration = it_{max}

Functions:

nof(pos): Calculate number of features according to θ and pos.
fit(pos): Calculate the fitness function according to θ and pos.

Algorithm:

```
for each particle i in (0, 1..., P):
   pbest_i = pos_i
   if(fit(pbest_i)>fit(gbest))
      gbest = pbest_i
Repeat for it_max iterations:
   for each particle i in (0, 1..., P):
Calculate fit(pbest_i), fit(pos_i), fit(gbest), nof(pbest_i), nof(pos_i), nof(gbest)
      if(fit(pos_i)>fit(pbest_i))
         pbest_i = pos_i
      if(fit(pbest_i)>fit(gbest))
         gbest = pbest_i
   for each particle i in (0,1,.., P):
Update v_i to v_i' and pos_i to pos_i' according to equation (1) & (2) respectively
      v_j^{t+1} = v_j^t*w + c_1*r_1j*(pbest_j - x_j^t) + c_1*r_2j*(gbest - x_j^t) (1)
      x_j^{t+1} = x_j^t + v_j^{t+1}          (2)
return gbest
```

Figure 1: Pseudo Code for PSO

2.2 Firefly Optimization

The firefly (FF) algorithm [16] is based on the attraction of fireflies towards each other according to their luminescence intensity. In feature selection problems, the intensity of fireflies represents the achieved AUC-ROC value using the position vector. The position vector represents a vector of m features, and each value ranges from 0 to 1 and features greater than a threshold is selected for the final problem and less are not. The pseudo-code of the FF optimization technique is given in Figure 2.

Algorithm 2: Pseudo Code of FF

Initialization:

1. *Number of fireflies:* F
2. *Threshold:* θ
3. *Alpha a*
4. *Beta β(Random Parameter)*
5. *gamma γ (Light intensity Coefficient)*
6. *Intensity vector I ={Ii} where i = 1...F*
7. *Pos = {Pos$_i$} where i = 1… F, Pos$_i$= {P$_{i1}$...P$_{im}$} where m is number of features*
8. *maxiteration = it$_{max}$*

Functions:

obj(): Objective function to calculate Fitness value AUC-ROC

Algorithm:

Determine Intensity (I) of each firefly using objective function obj().
Repeat for it$_{max}$iterations:
  For each firefly i = 1,2, … N:
    For each firefly j = 1,2, ...N:
      if(Ij > Ii)
        Move firefly i towards j according to equation (3)
        r = Distance between i and j firefly
        Dominance = (βeγr2 + *a(ξ-0.5) + 1)/2*
        Pos(i) = (1-Dominance*)Pos(i)* + Dominance**Pos(j) (3)*
    Evaluate new solution and update Intensity (I)
  Rank the fireflies according to intensity and find gbest
return gbest

Figure 2 : Pseudo Code for FF

2.3 Artificial Bee Colonization

ABC [17] is a swarm-based optimization technique based on the food searching behaviour of bees. In this technique, different types of bees perform various iterations to search for the optimal position where nectar is present. The optimal position is defined by a fitness function that calculates the optimal selected feature that gives the best AUC-ROC value. The pseudo-code of the technique is given in Figure 3.

Algorithm 3: Pseudo Code of ABC

Initialization:

1. *Number of employed bees = Number of onlooker bees:* B
2. *Threshold:* θ
3. *Abandonment Limit = L*
4. *Posb = {Posb$_i$} where i = 1… B, Posb$_i$ = {P$_{i1}$...P$_{im}$} where m is number of features*
5. *maxiteration = it$_{max}$*

Functions:

Fitness(Posb$_i$): Objective function to calculate Fitness value AUC-ROC using position of a bee

Algorithm:

Repeat for *it$_{max}$* iterations:
  // Employed Bees
  For each employed bee 1…B:
    New food source vi initialization
      Calculate Fitness(vi)

Update $Posb_i$ to vi if Fitness(vi) > Fitness($Posb_i$)
Fit_i = 1/Fitness($Posb_i$)
Calculate probability $pi = Fit/\sum Fit_i$

//Onlooker Bees
For each onlooker bee 1…B:
Select a food source with maximum probability (p)
Produce a new food source at vi
Update $Posb_i$ to vi if Fitness(vi) > Fitness($Posb_i$)

//Scout Bees
If any employed bee reach its abandonment limit Then
Replace it with a new random position
Store the best solution so far in gbest
return gbest

Figure 3: Pseudo Code for ABC

3 Extreme Learning Machine

Extreme Learning Machine (ELM) [14] is a class of randomization-based neural networks that use randomized weights and biases for the evaluation and analytically calculate the weights between the hidden and the output layers. This technique reduces the learning time to a great extent and maintains the predictive capability of neural networks. In this study, a single hidden layer ELM has been used with one input layer and one output layer as shown in Figure 4.

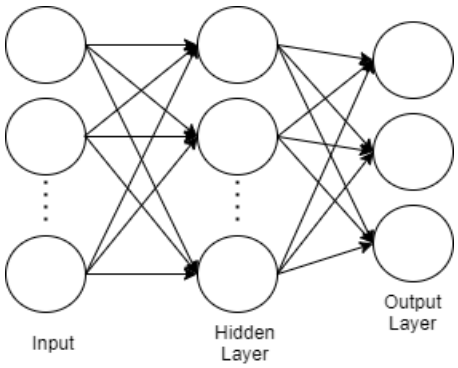


Figure 4 : Structure of Extreme Learning Machine

Consider dataset D = [X, Y] where $X = \{x_1, x_2, ... x_N\}^T$ and $x_i = \{x_{i1}, x_{i2} \dots x_{if}\}$, $Y = \{y_1, y_2, \dots y_N\}^T$ and $y_i = \{y_{i1}, y_{i2} \dots y_{im}\}$ is the output vector, f be the number of independent variables in the dataset, N be the number of samples in the dataset, and m be the output classes in the dataset.

The ELM output equation is defined as:

$$\sum_{t=1}^{h} \beta_t * \text{act}\,(w_t . x_i + \text{bias}_t) = o_i \qquad (4)$$

where h is the number of hidden layer nodes, β is the required weights between the hidden and the output layer, w_t are the randomly initialized weights between input and hidden layer nodes, x_i is the input feature vector, $bias_t$

is the bias of hidden layer, act() is the activation function, and o_i is the output vector of size m for sample i = 1, 2, 3, …, N.

Equation (4) can be re-written as:

$$\beta * H = O \tag{5}$$

$$\text{where, } H = \begin{bmatrix} act(w_1 x_1 + bias_1) & \cdots & act(w_k x_1 + bias_h) \\ \vdots & \vdots & \vdots \\ act(w_1 x_N + bias_1) & \cdots & act(w_k x_N + bias_h) \end{bmatrix}_{NXh} \tag{6}$$

The goal of the problem is to minimize the output error i.e. $||Y-O|| = 0$ where Y is the target output vector and O is the expected output vector:

$$H * \beta = Y \tag{7}$$

The optimal value of β is given as:

$$\beta\hat{} = H^{+}Y \tag{8}$$

In this study, we use Moore-Penrose generalized inverse H^{+} of the hidden layer output matrix. The optimal value of $\beta\hat{}$ gives the minimum output error and are the final weights between the hidden and output layer. The pseudo code of ELM is shown in Figure 5.

Algorithm 4: Pseudo Code of ELM

Initialization:

1. *Randomization Range for weights and bias :* (-1,1)
2. *Number of hidden layer nodes h*
3. *Weights W:* $\{w_t\}$ *where t = 1…h*
4. *Biases bias:*$\{bias_t\}$ *where t =1…h*
5. *Hidden layers: 1*

Algorithm:

1. Create H matrix i.e. output from hidden layer matrix using equation (6).
2. Calculate β matrix using H and Y according to equation (8).
3. Using β matrix and new H matrix created using the testing dataset, calculate the predicted output O using equation (5).

return O

Figure 5 : Pseudo Code for ELM

4 Experimental Setup

4.1 Datasets and Variables

The lung cancer risk prediction dataset used in the study is taken from the open-source data repository [18]. The dataset contains 1000 patient records and has 23 independent variables and one dependent variable. The dependent variable denotes the risk of the patient having lung cancer and has three values - "Low", "Medium", and "High", making this a multi-class classification dataset. The 23 independent variables are shown in Table 1.

Proceedings DOI: 10.21467/proceedings.115
ISBN: 978-81-947843-5-7

Series: AIJR Proceedings
ISSN: 2582-3922

Table 1: *Variables in the dataset*

Variable	Range	Variable	Range
Age	between 14-73 years	Passive Smoker	between 1-8
Gender	binary male and female	Chest Pain	between 1-9
Air Pollution	between 1-8	Coughing of Blood	between 1-9
Alcohol use	between 1-8	Fatigue	between 1-9
Dust Allergy	between 1-8	Weight Loss	between 1-8
Occupational Hazards	between 1-8	Shortness of Breath	between 1-9
Genetic Risk	between 1-7	Wheezing	between 1-8
Chronic Lung Disease	between 1-7	Swallowing Difficulty	between 1-8
Balanced Diet	between 1-7	Clubbing of Finger Nails	between 1-9
Obesity	between 1-7	Frequent Cold	between 1-7
Smoking	between 1-8	Dry Cough	between 1-7
Snoring	between 1-7		

4.2 Performance Measures

In this study, five performance measures are used to evaluate risk prediction models. These measures are - Precision, Recall, F-Measure, AUC-ROC and Accuracy. To define these measures, we first define the four terms that form a confusion matrix.

- True Positive (TP) - Number of positive samples that are predicted positive
- True Negative (TN) - Number of negative samples that are predicted negative
- False Positive (FP) - Number of negative samples that are predicted positive
- False Negative (FN) - Number of positive samples that are predicted negative.

The performance measures used are defined as:

- Recall is defined as the number of correctly predicted positive samples (TP) to the total positive samples predicted as positive (TP + FP).
- Precision is the ratio of the number of correctly predicted positive samples in the dataset (TP) to the total number of positive samples in the dataset (TP+FN).
- F-Measure is the harmonic mean of Precision and Recall.
- AUC-ROC is the area under the plot of True Positive Rate (TPR), and the FPR(False Positive Rate) where TPR is the same as the Recall and FPR is the ratio of the number of negative samples that are predicted positive to the total negative class samples in the dataset. It measures the predictive power of the classification technique.
- Accuracy is the sum of the correctly predicted positive and negative samples to the dataset's total samples.

The mathematical representation of these measures is given in Table 2. Since this is a multi-class classification problem, these metrics are evaluated using One-Vs-Rest method and micro-averaging of the values is performed on the individual class values to obtain the final results.

4.3 Validation Technique

This study uses 10-fold cross validation technique for training and evaluation of the classification techniques. In 10-fold cross validation technique, the dataset is randomly divided in 10 equal sized subsets. In one iteration, training is done on 9 subsets and testing is done on the remaining subset. This process is performed for 10 iterations allowing every subset to be a testing set once and the results of each iteration are macro-averaged to give the final performance measures.

4.4 Parameter Settings

The parameter settings of all the techniques have been carefully chosen using a grid search technique on various combinations. The final parameter settings of the techniques are shown in Table 2.

Table 2: *Parameter Settings used in the Study*

ML Technique	**Parameter Settings**
PSO	c1 = c2 = 1.496, w = 0.7298, number of particles = 30, number of iterations = 50, vmax =2, threshold for feature selection = 0.5.
ABC	number of employed bees = number of onlooker bees = 20, number of iterations = 50, limit for abandonment = 10, threshold for feature selection = 0.6
FF	population size = 100, alpha = 1, beta = 0.5, gamma = 1, number of function iterations = 10, threshold for feature selection = 0.7
ELM	initial randomization range for hidden weights and biases = (-1,1), activation function = sigmoid, number of hidden layer nodes = 80

5 Experimental Results and Observations

This section presents the results and observations of the experiments conducted in the study. All the experiments are conducted according to the framework described in Section 4. Figure 6 shows the AUC-ROC, F-Measure, Precision, Recall and Accuracy values for the three proposed models, i.e. PSO-ELM, ABC-ELM, and FF-ELM. The table also displays the performance measure values for a case where no feature selection technique is applied. PSO-ELM has achieved the highest value for all the performance measures, followed by ABC-ELM and then FF-ELM. It is observed that the case where no feature selection technique is applied performs the worst in terms of all performance measures. The application of swarm-based feature selection techniques has led to an increase in AUC-ROC in the range of 0.22% (for FF-ELM) - 0.68% (for PSO-ELM) the case where no feature selection technique is applied. Similarly, the accuracy is increased between 0.20% (for FF-ELM) - 0.61% (for PSO-ELM). A similar increase of more than 0.30% is observed for other performance indicators like Precision, Recall and F-Measure.

Table 3 shows the features selected by PSO, ABC and FF for training ELM classifiers. Factors like smoking habits, exposure to air pollution, occupational hazards, genetic risks are commonly selected. FF and ABC also include some clinical symptoms like snoring, weight loss, fatigue, shortness of breath and cold as strong predictors for lung cancer risk prediction.

Proceedings DOI: 10.21467/proceedings.115
ISBN: 978-81-947843-5-7

Series: AIJR Proceedings
ISSN: 2582-3922

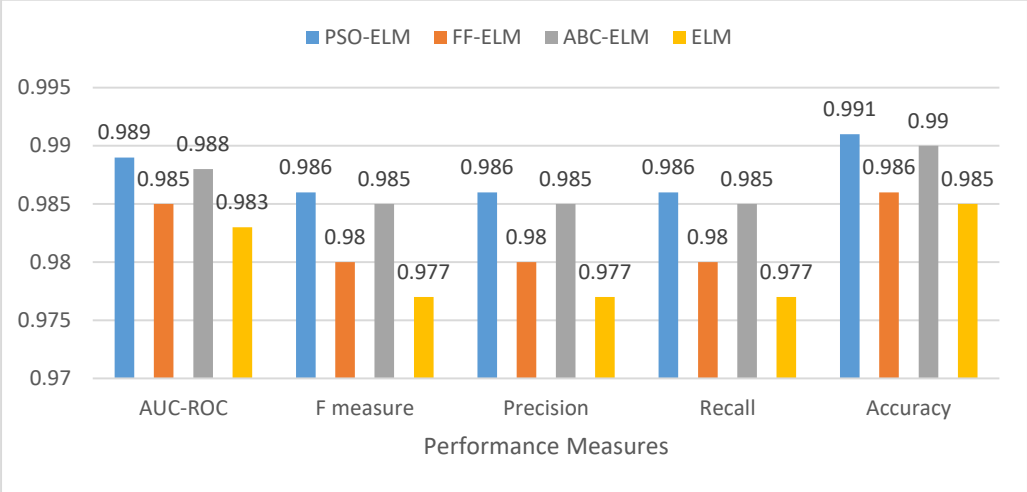


Figure 6: Results of ML Techniques

Table 3: *Features Selected by Different Techniques*

ML Technique	Selected Features
PSO-ELM	Age, Air Pollution, Genetic Risk, Chronic Lung Disease, Balanced Diet, Passive Smoker
FF-ELM	Gender, Air Pollution, Occupational Hazards, Chronic Lung Disease, Balanced Diet, Coughing of Blood, Fatigue, Shortness of Breath, Snoring
ABC-ELM	Alcohol use, Dust Allergy, Occupational Hazards, Genetic Risk, Obesity, Passive Smoker, Weight Loss, Swallowing Difficulty, Frequent Cold, Snoring

6 Conclusion and Future Work

The study proposed novel lung cancer risk prediction models using swarm-based techniques for feature selection and extreme learning machines for classification. A dataset of 1000 records and 23 variables was used where the risk was categorized into three classes according to severity. Five performance measures, i.e., AUC-ROC, Accuracy, Precision, Recall and F-Measure, were used to evaluate the model. The models achieved an AUC score in the range of 0.985 to 0.989, Accuracy in the range of 0.986 to 0.99 and F-Measure in the range of 0.98 to 0.985. PSO-ELM performed the best, followed by ABC-ELM in the second position and FF-ELM in the third position in terms of performance. All three models showed significant improvement over the baseline ELM with no feature selection in all performance measures. This shows the importance of identifying the effective subset of independent variables for lung cancer risk prediction through feature selection techniques. The improvement in performance of measures also show that the effective feature subset is much smaller than the actual number of features initially calculated. The most commonly selected features include exposure to air pollution, smoking, occupational hazards and clinical symptoms like snoring, shortness of breath and weight loss.

In the future, this study can be extended by including more categories of feature selection techniques, and the proposed models can be validated on more if publicly available datasets.

References

[1] World Health Organization. Cancer; 2018. Available from: https://www.who.int/news-room/fact-sheets/detail/cancer

[2] Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. New England Journal of Medicine, 365(5), 395–409 (2011). doi:10.1056/nejmoa1102873

[3] Tang, W., Peng, Q., Lyu, Y., Feng, X., Li, X., Wei, L., Li, N., Chen, H., Chen, W., Dai, M., Wu, N., Li, J., Huang, Y. Risk prediction models for lung cancer: Perspectives and dissemination. Chin J Cancer Res 2019. 31(2) 316-328 (2019). doi: 10.21147/j.issn.1000-9604.2019.02.06

[4] Toumazis, I., Bastani, M., Han, S. S., & Plevritis, S. K. (2020). Risk-Based Lung Cancer Screening: A Systematic Review. Lung Cancer. doi:10.1016/j.lungcan.2020.07.007

[5] Field, J. K., Chen, Y., Marcus, M. W., Mcronald, F. E., Raji, O. Y., & Duffy, S. W. The contribution of risk prediction models to early detection of lung cancer. Journal of Surgical Oncology. 108(5), 304–311 (2013). doi:10.1002/jso.23384

[6] Spitz, M. R., Hong, W. K., Amos, C. I., Wu, X., Schabath, M. B., Dong, Q., … Etzel, C. J. A Risk Model for Prediction of Lung Cancer. JNdCI Journal of the National Cancer Institute, 99(9), 715–726 (2007). doi:10.1093/jnci/djk153

[7] Xie, N.-N., Hu, L., Li, T.-H. .Lung Cancer Risk Prediction Method Based on Feature Selection and Artificial Neural Network. Asian Pacific Journal of Cancer Prevention. 15(23), 10539–10542 (2015).

[8] Daliri, M. R.. A Hybrid Automatic System for the Diagnosis of Lung Cancer Based on Genetic Algorithm and Fuzzy Extreme Learning Machines. Journal of Medical Systems. 36(2), 1001–1005 (2011). doi:10.1007/s10916-011-9806-y

[9] Chen S, Wu S, Identifying Lung Cancer Risk Factors in the Elderly Using Deep Neural Networks: Quantitative Analysis of Web-Based Survey Data, J Med Internet Res 2020;22(3):e17695 (2020), doi: 10.2196/17695

[10] Markaki, M., Tsamardinos, I., Langhammer, A., Lagani, V., Hveem, K., & Røe, O. D. A Validated Clinical Risk Prediction Model for Lung Cancer in Smokers of All Ages and Exposure Types: A HUNT Study. EBioMedicine, 31, 36–46 (2018). doi:10.1016/j.ebiom.2018.03.027

[11] Selvaraj, C., A survey on application of bio-inspired algorithms. Int. J. Comput. Sci. Inf. Technol. **67**, 366–370 (2014)

[12] Sakri, S. B., Abdul Rashid, N. B., Muhammad Zain, Z., Particle Swarm Optimization Feature Selection for Breast Cancer Recurrence Prediction. IEEE Access, 6, 29637–29647 (2018). doi:10.1109/access.2018.2843443

[13] Sawhney, R., Mathur, P., Shankar, R., A Firefly Algorithm Based Wrapper-Penalty Feature Selection Method for Cancer Diagnosis. In: Gervasi O. et al. (eds) Computational Science and Its Applications – ICCSA 2018. ICCSA 2018. Lecture Notes in Computer Science, vol 10960. Springer, Cham. (2018) doi: 10.1007/978-3-319-95162-1_30

[14] Huang, G. B., Zhu, Q.Y., Siew, C.K., Extreme learning machine: Theory and applications, Neurocomputing 70 (1-3), 489–501 (2006) doi: 10.1016/j.neucom.2005.12.126

[15] Kennedy, J., Eberhart, R. (n.d.). Particle swarm optimization. Proceedings of ICNN'95 - International Conference on Neural Networks (1995). doi:10.1109/icnn.1995.488968

[16] Emary, E., Zawbaa, H. M., Ghany, K. K. A., Hassanien, A. E., & Parv, B. Firefly Optimization Algorithm for Feature Selection. Proceedings of the 7th Balkan Conference on Informatics Conference - BCI '15 (2015). doi:10.1145/2801081.2801091

[17] Karaboga, D., & Akay, B. A comparative study of Artificial Bee Colony algorithm. Applied Mathematics and Computation, 214(1), 108–132 (2009). doi:10.1016/j.amc.2009.03.090

[18] [dataset] https://data.world/cancerdatahp/lung-cancer-data

2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP - 2021)

2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP) | 978-1-6654-3277-1/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCCSP52374.2021.9465497

ArtiMarker: A Novel Artificially Inflated Video Marking And Characterization Method on YouTube

1st Deepika Varshney
Department of Information Technology
Delhi Technological University
Delhi, India
deepikavarshney06@gmail.com

2nd Dinesh Kumar Vishwakarma
Department of Information Technology
Delhi Technological University
Delhi, India
dinesh@dtu.ac.in

***Abstract*—YouTube is on demand platform, crowd sourced by design and its huge popularity encourages users to adopt fraudulent methods to inflate viewcounts of their videos, thereby increasing the revenue of the video uploader. It is very important to mark artificially inflated videos as it is one of the stepping stone to encourage false information over web and to gain monetary benefits.To address this issue, in this paper we present "ArtiMarker" a novel artificially inflated video marking and characterization (less,medium,high) method on YouTube based on user level feature.The proposed technique described in the paper will help researchers to understand a new efficient method to mark and characterize artificially inflated videos. The proposed parameters in this paper can also be used in identification of spam videos, fake profiles on YouTube. The performance of proposed work is analyzed and demonstrated using various classifier such as Decision Tree, Function-Based and Bayes.The comparison is based on the testing strategy utilized during the experiments;percentage split,cross validation. It has been observed that Bayes network,J48 and Random Forest are best suited for percentage-split while Random Forest is more suited for cross validation schemes.**

***Index Terms*—Fraudulent views, Suspicious, Fraud ,YouTube, Artificial Inflation.**

I. Introduction

Online video platforms such as YouTube, Dailymotion etc. provide a way to the users to publish their videos that can be viewed by others on the web. These platforms provide on-demand video services which allow users to share, upload and watch videos depending upon their interest. From the YouTube statistics [1] , [2] it can be seen that YouTube has over a billion users almost one-third of all people on the Internet and each day those users watch 5 billion videos, thereby generating large number of views[for ex. 10,113 YouTube videos generated over 1 billion views].[3]. Due to such a scale of popularity, lot of malicious and scam activities are targeted over YouTube. Malicious entities spread malicious content that compromise system reputation and degrades user experiences[1]. Due to heavy use of social media network a certain group of internet users called cyber criminal make use of this opportunity to create spam fraud and other attacks on social media platform.

A. Artificial Traffic Spam for Monetary Benefit

YouTube has a monetization system to reward video uploaders. To get the monetary benefits users have to link their YouTube account with adsense account. Successfully linking of the two accounts means you will be eligible to receive payments[4]. YouTube's algorithm works very hard in rewarding quality content instead of content that games the system. YouTube reward videos with respect to their watched time. Higher watch time videos are ranked more higher on the algorithm[5].The one of the important things to understand is that YouTube reward videos based on the number of minutes a video is watched. So, the ultimate goal is to increase the watch time [6]. After the deployment of this policy, malicious users apply illegal activities to flood the platform by undesired content for monetary benefits. There are various kinds of undesired content causing spam on YouTube like via text comments,fake likes ,views..etc. YouTube faces problem to manage large scale user text comments[23]. User posted undesired text comments to self promote their videos or to disseminate malicious link to steal private data. Lot of research has been done to automatically detect/filter undesired text comments posted on YouTube[2,3]. Along with the comment spam the

[1]https://www.youtube.com/yt/about/press/
[2]https://merchdope.com/youtube-statistics/
[3]http://videonitch.com/2017/12/13/36-mind-blowing-youtube-facts-figures-statistics-2017-re-post/
[4]https://support.google.com/adsense/answer/2530958?hl=en
[5]https://creatoracademy.youtube.com/page/lesson/impact-metrics
[6]https://mag.octoly.com/how-to-understand-the-youtube-algorithm-in-2018-c435136abb97

978-1-6654-3277-1/21/$31.00 ©2021 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:36:00 UTC from IEEE Xplore. Restrictions apply.

other way that malicious users adopted to self promote their videos to gain popularity and rewards is via artificial views. One of the spam fraud that we have included in this paper is known as ”Artificial Traffic Spam”. By Artificial we mean that anything that artificially increases the number of views, likes, comments, or other metric either through the use of automatic systems or by serving up videos to unsuspecting viewers[7]. Artificial traffic spam greatly affects online advertising. From the Interactive Advertisement Bureau(IAB) report, it was found that online advertising generated revenue of $49B in 2014 in the US alone. With such a revenue it is not surprising that online advertising attract frauds[4]. From the study it was found that 15-30% of add-impression are fraudulent leading losses in order of billions of dollars to advertisers[5]. Online video platform have a monetization policy, which means one can put an uploaded video for monetization. Depending upon the factors related to video popularity and context, video platforms put advertisements over these uploaded videos. Advertiser typically pay some money to video service provider depending upon number of clicks received on their videos, referred to as Cost Per Click(CPC) and number of views, referred as Cost Per View(CPV). Video publisher earn revenue from the video platform based on advertisement which gets rendered on his videos by the video platform. Hence, there is an incentive for dishonest publisher to inflate the number of impressions on their video [6,7,8,9,10,11,12,13]. This will cause a huge loss to the advertisers, who pay to get their ads displayed[4].

There are many companies (like YTMONSTER.net, Boostwolf.com.etc) who are building their businesses based on selling YouTube related services, like for $3,1000 views[8]. From the previous study it is observed that YouTube has effective means to identify fake views, study reveals that whenever the permissive threshold for the detection of fake monetized views is crossed, YouTube severely penalizes the video uploader of the video by suspending their Adsense account[4].The Fig 1 clearly depicts the example of such instances. Though YouTube is tracking and suspending suspicious video accounts, still these fraudulent activities are on place. The other major impact of artificial traffic spam is in promoting false/fake news. Social media enables the wide propagation of ”fake news” means the news with intentionally false information. Low cost of creating social media profiles makes it easy to create malicious accounts, such as social bots, trolls etc. all of which can become a powerful source for proliferation of fake news[14]. Existing solutions worked on network level features[4][15] to detect artificially inflated views. However the main limitations in the existing solution is the lack of user perspective. In addition to it no practical system(Ytmonster.net, Boostwolf.com,etc) has been explored yet, that involved in selling fraud services. The existing solution is doing the same as what YouTube is doing, tracking of views and penalize video publisher by cutting down number

[7]https://support.google.com/youtube/answer/2801973?hl=en

[8]http://boostwolf.com/youtube.php

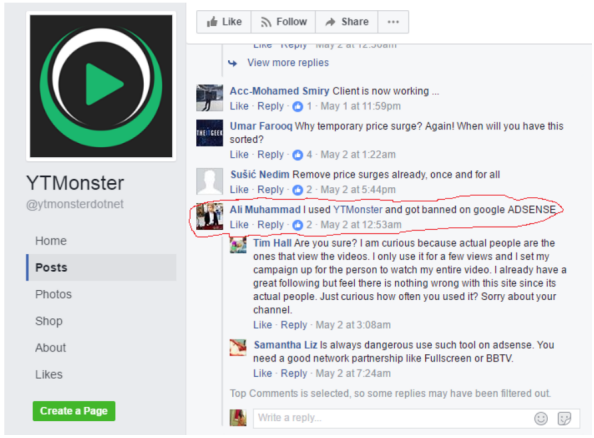


Fig. 1: The comment posted by an YTMonster user that his account got banned by YouTube

of views and sometimes by removing video too from the platform whenever any fake is encountered, [9], due to large scale of videos, many of these videos are still running over the platform. The internal information is not revealed to normal users. By only seeing number of views on a video, one cannot identify the credibility of video(views are real or fake). To overcome the above mentioned problems, in this paper we propose a new mechanism that will help video service provider to accurately mark as well as characterize fraudulently inflated videos.The novelty of the proposed method lies in exploring a real time practical system selling artificial views and encouraging results are also reported. We strongly believe that the proposed scheme of this kind have been reported for the first time in this paper.

In summary, our paper makes the following contribution:

1) We analyze, explore and automate the third party platform , selling fraud services to inflate video popularity. This analysis further help in building ground truth data set to mark and characterize artificially inflated videos.
2) We provide concrete suggestion to video service providers as how to provide more transparent system and automatically reporting artificial traffic spam to the normal users.

The rest of this paper is organized as follows: Related work is elaborated in Section II. Our proposed approach ”Artimarker” is discussed in section III. Section IV describe experimental analysis and results. Conclusions and Future work are described in section V.

II. Related Work

Network anomaly detection attracted a huge interest from commercial concern as well as academic research. Many researchers have worked on analysis of flow level and traffic level anomalies for network monitoring and network security[16].Whereas concerning to this some of the studies are also keen towards analyzing malicious videos on social

[9]https://support.google.com/youtube/answer/2801973?hl=en

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:36:00 UTC from IEEE Xplore. Restrictions apply.

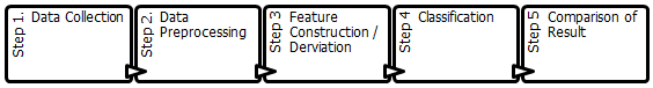


Fig. 2: Block Semantic Diagram of Proposed Approach

media.There appears to be similar issues when one wants to lure user eyeballs on the internet. Cao et al.[17] detected fake accounts on the online social network by ranking users according to their perceived likelihood of being fake based on social graph properties. Whereas the author's of [18] employed three sets of latent features based on user-profiling, video-content and human consensus in order to retrieve cognitive evidences for the detection of clickbait videos. The author's of [19] employed K- means approach has been proposed to classify unlabeled data into different categories for anomaly based intrusion detection. To detect network anomaly researcher proposed statistical and signal analysis approaches. Whereas, Bolton and Hand[20] presented a review of statistical fraud detection. Most of the anomaly detection problems has been solved using an Entropy function. Gu et al.[21] used two phase entropy measures to detect network anomalies by comparing the current network traffic against a baseline distribution. There is a concept of maximum entropy principle that is applied to estimate the distribution for normal network operation using pre-labeled training data. The efficient streaming algorithms has been proposed to implement the entropy measurement on high-speed links with low requirements on CPU and memory [22]. Chen et al.[23] proposed an efficient entropy-based approach to detect fake views. A preliminary report of this work appeared in nossdav 2014. Metwally et al. [10] presented an early study in which they used the IP address as the parameter to detect coalition of fraudulent users or fraudsters.

From the above mentioned work it has been observed that various researchers has used network level features to counter anomaly/frauds traffic.Our proposed work is novel and quite different from the above mentioned work. The past work done by the researchers to detect fraudulent views via network level features. Though they are detecting and removing fake views, but still frauds are ongoing. Our proposed technique ArtiMarker used to mark artificially inflated videos using user level features. Our approach analyze the behaviour of impression on a video and then according to our proposed features(NC(Num count of rate of decrease),MD(Maximum decrease count value during analysis),(CBD) (city block distance) characterize video into Less(L), Medium(M) and High(H) suspicious categories. This categorization will help to identify those video that are frequently involved in purchasing fake views. From the analysis it has been observed that Daily analysis and public display of video performance measure will greatly reduce such fraudulent activities.

III. Proposed Methodology

In this section, we have discussed "Artimarker" scheme which is proposed to mark and characterize fraudulently inflated videos in an online video platform. The main aim is to characterize the videos into three categories(LMH). Various machine learning classifier such as LMT, J48, Random Forest, Simple Logistic, SMO, Logistic, Naive Bayes and Bayes Network are utilized to classify the input videos [10] and multiple thresholds are analyzed based on CBD parameter for the classification of input videos into LMH Category. The block semantic diagram of our proposed approach is shown in Fig 2.

A. Abbreviations Used

In this section we elaborate the detail description of the functions and parameters used in our work to categorize the video into less, medium and high suspicious category.

1) NC = Number of times video found suspicious by YouTube[Penalize by YouTube] algorithm during our analysis period.
2) MD = Identifying the maximum value of views cut down via YouTube algorithm during our analysis period.
3) Weight_Score(WS)= Weight score parameter for k number of videos can be calculated as: Weight Score(WS)= X_ia W_ia + X_ib W_ib,where, X_ia = Value of NC for îth video, W_ia= Weight Percentage assigned to NC, X_ib= Value of MD for îth video, W_ib= Weight Percentage assigned to MD,
   where *i= 1 to k*
4) Extract_WS_parameter() = It is a function which takes four parameters (NC,MD,W_ia,W_ib) as an input for each video and return weight score value as a result for each video.
5) Threshold = It is a parameter which takes weight score parameter as an input for v number of videos shown below.

$$Threshold(T) = (WS)/v \tag{1}$$

6) CBD : The CBD between two points, a and b, with k dimensions can be calculated as shown by eq.2:

$$CBD = \sum_{j=1}^{k} |a_j - b_j| \tag{2}$$

B. Data Collection

In order to construct features time series data sequence for all suspicious videos that has been collected. YouTube provides a platform for a huge repository of video data and engagements (likes, views , dis-likes, comments, etc) around it.There are various third party sites that provides YouTube related services[views, subscribers..etc] to increase

[10]https://gerardnico.com/data_mining/algorithm

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:36:00 UTC from IEEE Xplore. Restrictions apply.

TABLE I: Characterization Accuracy[NC-60:MD-40] : Percentage Split

Classifier	60:40	70:30	80:20	90:10
LMT	99.78	100	99.57	100.
J48	99.57	100	100	100
Random Forest	99.57	100	100	100
Simple Logistic	99.78	100	99.57	100
SMO	95.12	96.32	98.30	83.89
Logistic	99.78	99.71	99.57	100
Bayes Network	100	100	100	100
Naive Bayesian	99.15	99.71	99.15	93.22

TABLE II: Characterization Accuracy[NC-40:MD-60] : Percentage Split

Classifier	60:40	70:30	80:20	90:10
LMT	100	99.43	99.57	100.
J48	100	100	100	100
Random Forest	100	100	100	100
Simple Logistic	100	99.43	99.57	100
SMO	95.12	96.32	95.76	84.74
Logistic	99.78	99.43	99.57	100
Bayes Network	100	100	100	100
Naive Bayesian	98.30	96.61	98.30	98.30

TABLE III: Characterization Accuracy[NC-70:MD-30] : Percentage Split

Classifier	60:40	70:30	80:20	90:10
LMT	99.15	99.71	99.57	100.
J48	99.78	99.71	99.57	100
Random Forest	100	86.15	100	100
Simple Logistic	99.15	99.71	99.57	100
SMO	95.55	96.32	95.76	84.74
Logistic	99.15	99.71	100	100
Bayes Network	99.78	99.71	99.57	100
Naive Bayesian	98.09	96.32	98.72	99.15

video/channel popularity of the Registered user. One of the third party site that we have considered in this paper is YTmonster [11]. The suspicious video URLs are collected from ytmonster.net programmatically using Selenium Webdriver. The features were then collected for these URLs from YouTube using YouTube Data API.2 [12]. The Data API of YouTube incorporate functions normally executed on the YouTube website into your own website or application. To build the ground truth dataset, 1180 suspicious videos were collected form YT-MONSTER. We have analyzed video based features(view_count,comment_count,like_count...etc) with an interval of 10 minutes for a duration of 9 days (26th-July-2017- 03-Aug- 2017). After every 10 minutes,the given set of features has been collected for every individual video and from the analysis we got a certain discrimination in views concerning to fake from real ones.

C. Technique to Mark Suspicious Video

In order to characterize the suspicious video into LMH category, the threshold values based on CBD concept is evaluated to distinguish the boundaries of all three. The name city block distance is explained if you consider two points in the xy-plane. The shortest distance between the two points is along the hypotenuse, which is the Euclidean distance. The city block distance is instead calculated as the distance in x plus the distance in y, which is similar to the way you move in a city where you have to move around the buildings instead of going straight through. Along with CBD here we consider video views as a primary feature used to derive other features. In order to calculate the weight score, we have given weights to two parameters NC and MD as 40% and 60% respectively. The comparative accuracy results of the suggested scheme has been also analyzed on various percentage ratios of NC and MD i.e(60:40),(40:60) and (70:30) to evaluate the Weight Score as shown in Table I, II and III respectively. One can observe here that considering NC= 40 and MD= 60 most of the classifier are performing better for different data proportions.

The flowchart of the proposed scheme is shown in Fig 3. In the first step suspicious video_ids has been collected from YTMONSTER.NET. Behavioral analysis has been performed on the collected video feature data set given in step 2. After analysis based on derived features and threshold values video have been categorized into LMH suspicious category.

[11]https://www.ytmonster.net/

[12]https://developers.google.com/youtube/v3/docs/

IV. Experimental Analysis and Results

This section depicts the experimental analysis and results of the proposed technique. In this section we have described the experimental steps taken for the classification of videos. To classify the videos into LMH category, firstly 1180 suspicious videos are analyzed from the collected dataset, then their CBD values are compared with first threshold (T1).The values for Threshold and CBD is calculated using Eq. 1 and Eq.2 respectively. The videos having CBD value more than T1 will belongs to High suspicious category. The same procedure has been applied on the remaining videos. This will again separate videos into two parts with respect to threshold (T2). The videos having CBD value more than the T2 will come under medium suspicious and the remaining videos will belong to less suspicious category as shown in Fig 3. After performing the training on the collected features, various features such as NC, MD and CBD of collected data set are evaluated and are used to train the machine based on these feature set. As we have already observed in section C that considering NC=40 and MD=60, most of the classifier are performing best for different data proportion. So, for accuracy analysis we have considered the weight percentage ratio of NC and MD as 40,60 respectively. In our experimentation we have considered four data proportions such as 60-40, 70-30, 80-20, and 90-10 for the percentage split testing strategy and it reveals that the Bayes Network, J48 and Random Forest are best in utilizing the proposed features for classification as shown in Table III. The other experiments were performed by varying folds required

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:36:00 UTC from IEEE Xplore. Restrictions apply.

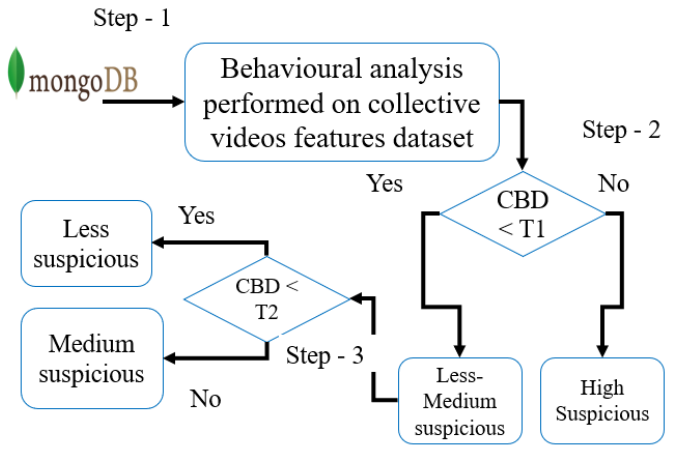


Fig. 3: Flowchart of Proposed Scheme

TABLE IV: Characterization Accuracy[NC-40:MD-60] : Cross Validation

Classifier	K=10	K=15	K=20
LMT	100	99.71	99.91
J48	99.91	99.71	99.91
Random Forest	100	100	100
Simple Logistic	100	100	99.91
SMO	86.34	85.92	85.83
Logistic	99.57	99.66	99.74
Bayes Network	99.91	99.91	99.91
Naive Bayesian	95.58	95.33	95.33

in cross-validation testing. Details of the results are shown in Table IV. It has been observed that the highest accuracy was obtained using the Random Forest classifier while worst in the case of SMO.

V. Conclusions and Future Work

This paper presents an ArtiMarker scheme based on user level features where we have taken NC, MD and CBD as a major parameters to mark and characterize fraudulently inflated videos into LMH category. The performance of the proposed work is analyzed and compared using various classifier such as Decision Tree, Function-Based and Bayes etc. In the evaluation phase, detection accuracy obtained by all given classifiers are compared. The comparison is based on the testing strategy utilized during the experiments;i.e, percentage split, cross validation. It has been observed that Bayes network, J48 and Random Forest are best suited for percentage-split while Random Forest is more suited for cross validation. Therefore we conclude that our proposed approach efficiently characterize videos into Less, Medium and High using Bayes Network, J48 and Random Forest in percentage-split, while with Random Forest in Cross -validation testing strategy. This analysis is new as it has not been done in earlier studies. In future, work should be done in order to build a tool/extension for automatic detection of fraudulent views in real time that will help normal user/advertisers.

References

[1] Miss. Sayali S. Karmode.et al. "DETECTION AND BLOCKING SOCIAL MEDIA MALICIOUS POSTS ." International journal of modern trends in engineering and research, vol-4 issue- 5 May 2017.

[2] JEzpeleta, Enaitz, et al. "A Mood Analysis on Youtube Comments and a Method for Improved Social Spam Detection." International Conference on Hybrid Artificial Intelligence Systems. Springer, Cham, 2018.

[3] Aiyar, Shreyas, and Nisha P. Shetty. "N-Gram Assisted Youtube Spam Comment Detection." Procedia Computer Science 132 (2018): 174-182.

[4] Miriam Marciel , Rubén Cuevas , Albert Banchs , Roberto González , Stefano Traverso , Mohamed Ahmed , Arturo Azcorra,"Understanding the Detection of View Fraud in Video Content Portals", Proceedings of the 25th International Conference on World Wide Web, April 11-15, 2016, Montréal, Québec, Canada Pg- 357-366.

[5] Abdullah, Abdullah O., et al. "A comparative analysis of common YouTube comment spam filtering techniques." Digital Forensic and Security (ISDFS), 2018 6th International Symposium on. IEEE, 2018.

[6] V. Anupam, A. Mayer, K. Nissim, B. Pinkas, and M. Reiter. On the Security of Pay-Per-Click and Other Web Advertising Schemes. In Proceedings of the 8th WWW International Conference on World Wide Web, pages 1091–1100, 1999.

[7] C. Blundo and S. Cimato. SAWM: A Tool for Secure and Authenticated Web Metering. In Proceedings of the 14th ACM SEKE International Conference on Software Engineering and Knowledge Engineering, pages 641–648, 2002.

[8] M. Jakobsson, P. MacKenzie, and J. Stern. Secure and Lightweight Advertising on the Web. In Proceedings of the 8th WWW International Conference on World Wide Web, pages 1101–1109, 1999.

[9] D. Klein. Defending Against the Wily Surfer-Web-based Attacks and Defenses. In Proceedings of the 1st USENIX ID Workshop on Intrusion Detection and Network Monitoring, pages 81–92, 1999.

[10] A. Metwally, D. Agrawal, and A. El Abbadi. Duplicate Detection in Click Streams. In Proceedings of the 14th WWW International World Wide Web Conference, pages 12–21, 2005

[11] A. Metwally, D. Agrawal, and A. El Abbadi. Using Association Rules for Fraud Detection in Web Advertising Networks. In Proceedings of the 31st VLDB International Conference on Very Large Data Bases, pages 169–180, 2005.

[12] M. Naor and B. Pinkas. Secure and Efficient Metering. In Proceedings EUROCRYPT International Conference on the Theory and Application of Cryptographic Techniques, pages 576–590, 1998.

[13] D. Vise. Clicking To Steal. Washington Post Magazine, page F01, April 17 2005.

[14] Shu, Kai, H. Russell Bernard, and Huan Liu. "Studying Fake News via Network Analysis: Detection and Mitigation." arXiv preprint arXiv:1804.10233 (2018).

[15] Liang Chen, Yipeng Zhou, and Dah Ming Chiu. 2014. Fake view analytics in online video services. In Proceedings of Network and Operating System Support on Digital Audio and Video Workshop (NOSSDAV'14). ACM, New York, Article 1. DOI:http://dx.doi.org/10.1145/2578260.2578263.

[16] Lakhina, Anukool, Mark Crovella, and Christiphe Diot. "Characterization of network-wide anomalies in traffic flows." Proceedings of the 4th ACM SIGCOMM conference on Internet measurement. ACM, 2004.

[17] Cao, Qiang, et al. "Aiding the detection of fake accounts in large scale social online services." Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation. USENIX Association, 2012.

[18] Varshney, Deepika, and Dinesh Kumar Vishwakarma. "A unified approach for detection of Clickbait videos on YouTube using cognitive evidences." Applied Intelligence (2021): 1-22.

[19] Ensafi, Roya, Soheila Dehghanzadeh, and Mohammad-R. Akbarzadeh-T. "Optimizing fuzzy k-means for network anomaly detection using pso." Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on. IEEE, 2008.

[20] Bolton, Richard J., and David J. Hand. "Statistical fraud detection: A review." Statistical science (2002): 235-249.

[21] Gu, Yu, Andrew McCallum, and Don Towsley. "Detecting anomalies in network traffic using maximum entropy estimation." Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement. USENIX Association, 2005.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:36:00 UTC from IEEE Xplore. Restrictions apply.

[22] Lall, Ashwin, et al. "Data streaming algorithms for estimating entropy of network traffic." ACM SIGMETRICS Performance Evaluation Review. Vol. 34. No. 1. ACM, 2006.
[23] Chen, Liang, Yipeng Zhou, and Dah Ming Chiu. "Analysis and detection of fake views in online video services." ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 11.2s (2015): 44.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:36:00 UTC from IEEE Xplore. Restrictions apply.

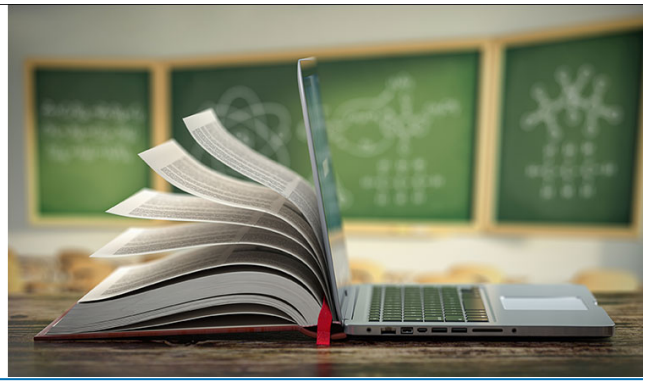
PAPER • **OPEN ACCESS**

CFD Modelling and Simulation of an Indirect Forced Convection Solar Dryer

To cite this article: Rupanshu Singh *et al* 2021 *IOP Conf. Ser.: Earth Environ. Sci.* **795** 012008

View the article online for updates and enhancements.

The Electrochemical Society
Advancing solid state & electrochemical science & technology
2021 Virtual Education
Fundamentals of Electrochemistry:
Basic Theory and Kinetic Methods
Instructed by: **Dr. James Noël**
Sun, Sept 19 & Mon, Sept 20 at 12h–15h ET
Register early and save!



This content was downloaded from IP address 14.139.251.100 on 28/07/2021 at 10:37

CFD Modelling and Simulation of an Indirect Forced Convection Solar Dryer

Rupanshu Singh[1], Pranav Salhan[1] and Anil Kumar[1,2]

[1]Department of Mechanical Engineering, Delhi Technological University, Delhi, India
[2]Centre for Energy and Environment, Delhi Technological University, Delhi, India
Email: anilkumar76@dtu.ac.in

Abstract. Solar dryer is a conventional device used for solar drying products. There are several types of solar dryers being employed for drying fruits and vegetables. In this research, the indirect type solar dryer is studied and computational fluid dynamics is employed to simulate the process. Ansys fluent CFD software is used to simulate and obtain dynamic and thermal performance of the dryer at different operating conditions (mass flow rate). The predicted results are validated with the help of experimental results. The average velocity at the exit of the collector was 2.89 m/s for 0.0872 kg/s mass flow rate, while for a mass flow rate of 0.0447 kg/s was 1.36 m/s. The maximum temperature at the collector outlet was 44.6°C and the maximum absorber plate temperature was 69.57°C for a magnitude of 0.0872kg/s mass flow rate. The temperature profiles of the trays are also analysed and the tray closest to the collector is found to have the maximum average temperature. The results also revealed that CFD simulations are a reliable method to predict and design solar dryers.

1. Introduction

The depletion of natural resources of Earth, have made it necessary to shift towards the renewable resources alternative for the energy extraction purposes. Solar energy is one of the free, abundant renewable resources. It has several productive applications like concentrated solar power plants for generating high amounts of energy, while low energy generation devices like solar furnaces, solar dryers etc., are used. A solar dryer is a conventional device used to dry food products and vegetables to preserve and maintain the integrity for a longer period [1]. The fruits and vegetables are dried using direct or indirect solar irradiance. A flow of dry air is allowed to pass through a closed chamber containing the products, the moisture is extracted from the food product and a dry output is obtained. For increasing the efficiency and making the dryer operational at night and in all weather conditions, an integration is added with a secondary heater like an electric heater, biomass stove, etc. [2].

Solar dryer is preferred over the sun drying process. The food product is placed in a closed chamber situated at a certain height, away from biological reactions, insects and dust [3]. In a solar dryer, depreciation in relative humidity content was observed while the drying temperature was increased, resulting in less moisture content [4]. The solar dryer application includes the drying of agricultural products and textile materials, dehydrating vegetables, preservation of dairy products, etc. [5]. Lingayat et al. (2020) [6] reviewed the different models of ITSD used for agricultural purposes. The dry air flowing through the solar dryer can be due to nature's convection; in case of insufficient wind flow, a blower fan can be used to produce forced convection. A glass surface is present on the front side of the solar dryer to allow the solar radiation to penetrate the product and heat the inner part, and dry the outer cover; this amplifies the heat transfer process. The factors affecting the solar drying

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

operation include solar radiation absorptance, thermal conductivity of crops, and excessive temperature [1].

Solar dryers are broadly categorized into passive solar drying systems (natural air circulation) and active solar drying systems (forced air convection), which are further divided into direct, indirect, mixed, and hybrid solar dryers. Prakash et al. (2016) [5] have reviewed and discussed the numerous types of solar dryers and the development and condition of solar dryers in India. This includes the usage of solar panels for operating the fan to induce forced air flow, which operates entirely on solar resources. The pharmaceutical and nutritional value of any crop can be preserved by drying if a wide range of controllable parameters of drying are present [7].

Many studies have shown the experimental work on drying agricultural products like vegetables and fruits. Vijayan et al. (2020) [8] analyzed the environmental and exergy impact by constructing an active type indirect solar dryer for drying slices of food products like bitter gourd. The mass flow rate's effect on solar dryer's pickup and exergetic efficiency in drying slices of bitter gourd were calculated.

Lingayat et al. (2020) [9] developed an indirect type solar dryer and analyzed the drying kinetics of watermelon and apple. The results showed that overall thermal efficiency of the dryer came out to be 28.76% for watermelon and 25.39% for apple slices. The depreciation in the moisture content of apple slices was 5.361 kg/kg of db and 10.264 kg/kg of db (dry basis) for watermelon slices. Jiskani et al. (2020) [4] conducted the experiment on chilies and grapes using passive-direct type solar dryer and observed a reduction in the moisture content of around 75% in grapes & 85% in green chilies, while examining for 15 hours. Hamdi et al. (2018) [10] analyzed energy and exergy involved in drying tomatoes using solar greenhouse dryers. A reduction in moisture of around 19.91g water/g dry matter resulted in drying upto 48h and overall dryer efficiency 42%. Apart from the agricultural applications, solar dryers can dry different materials based on their absorptivity property. Luo et al. (2010) [11] did an experimental study on solar drying the ignite (brown coal) material and observed the characteristic property of the material. Direct type solar dryer was used. The results concluded that the optimal condition for the solar drying process requires a thin layer having small particles of material with high strength solar radiation.

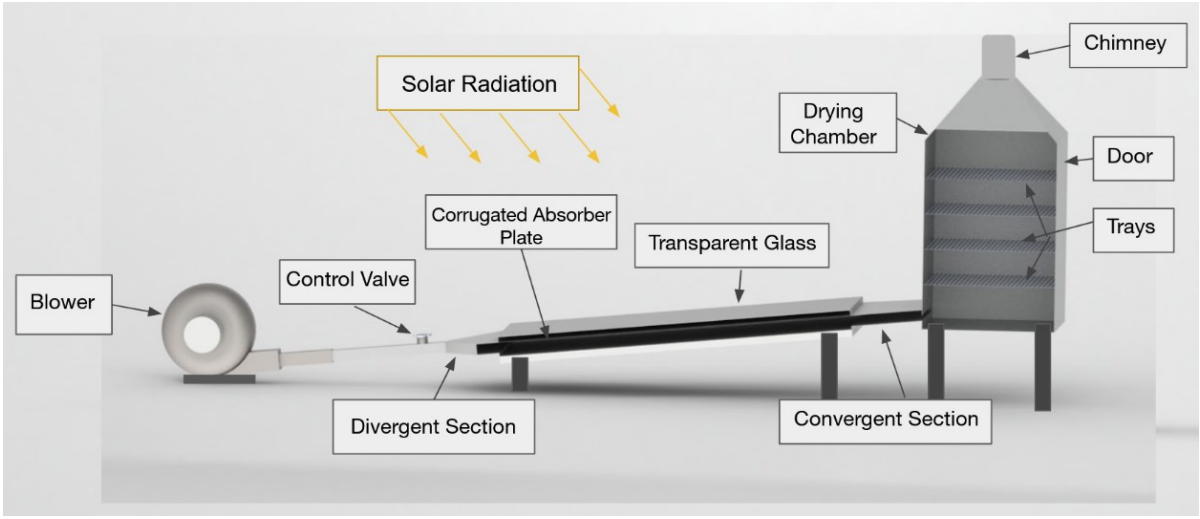


Figure 1. Detailed diagram of solar dryer setup [8]

Computational fluid dynamics (CFD) develops a mathematical model based on the governing equations & boundary conditions and provides the numerical solution [12]. The 3D air flow field and temperature variation over the drying setup can be predicted by using the CFD analysis [13]. CFD analysis software develops the result contours and graphs by applying the boundary conditions on a

virtual model. It can help in optimizing the design and help researchers to analyze different characteristics of CFD simulation.

Jain et al. (2018) [14] developed a CAD model of -shelf direct solar dryer and simulated the drying process using ANSYS fluent. The energy variation and heat flux contours were obtained and compared with the experimental work. M. et al. (2019) [15] analyzed the optimal shape of the roof for a greenhouse solar dryer. Different roof shapes (hemispherical, triangle, trapezoidal) were virtually designed and temperature variation was observed using the analysis software. Results showed that the trapezoidal shaped roof achieved the maximum temperature.

Meenakshi Reddy et al. (2018) [16] developed a solar crop dryer having a recovery mechanism of exhaust's waste heat. The results were compared with the normal solar dryer and the flow field & temperature distribution within the solar dryer was demonstrated using CFD analysis. Moghimi et al. (2020) [17] proposed an optimal solar dryer design for household purposes. The experiment was conducted on a physically designed device, and the air flow and temperature contours were obtained by modeling and analyzing the design on the CFD software.

This current research work demonstrates the CFD simulation of active type indirect solar dryer operated at different mass flow rates. The tentative temperature distribution and flow field is simulated, which can help the researchers visualize different features of the indirect solar dryer. A brief review of solar dryer and the methodology involved in simulation is presented in this paper and the simulation results are compared with the experimental data provided by [8].

2. Design and material of solar dryer

The solar dryer's design and construction require considering certain variables like solar irradiation, humidity and ambient temperature [18]. An active type indirect solar dryer was constructed by [8] at Coimbatore, India. The solar dryer consisted of a drying chamber having four trays at a specified distance on which the product was placed. (Figure 1.) shows the experimental setup of an active type indirect solar dryer with different parts annotated separately. The air flow is transferred by the blower fan, which gets heated up by solar radiations while traveling through the collector and is released through the chimney, drying the products in the drying chamber.

Table 1. Solar dryer's component dimensions & materials

S. No.	Component	Material	Dimensions (mm) (lxbxh)
1	Drying Chamber	Galvanized iron	750×750×1050
2	Flat plate collector	-	2000×1000×150
3	Tray ×4	Aluminium	750×750×1.2
4	Corrugated absorber plate	Galvanized iron sheet	2000×1000×6
5	Window glass cover	Glass	2000×1000×4
6	Glass wool insulation	Glass wool	25 mm thick
7	Chimney	Galvanized iron	150 mm diameter
8	Convergent section	Galvanized iron	1000×400×60

The tray's area was 0.5625 m^2, placed at a distance of 0.2m from each other. For ensuring maximum collection of solar radiations, the collector was aligned to 11° to the horizontal, in the latitude's direction of the corresponding area. For reducing the heat losses due to convection and conduction, insulation of glass wool is provided in the whole setup.

The setup is shown in Figure 2 (a) was developed as a CAD model in Solidworks 2020. All the solar dryer parts were developed and assembled separately according to the dimensions given in (Table 1). The collector and drying chamber consisting of four trays were designed and connected with a convergent part. The drying chamber consisted of a vent at the top, which acts as the chimney outlet and an inlet vent was provided at the opening of the collector part.

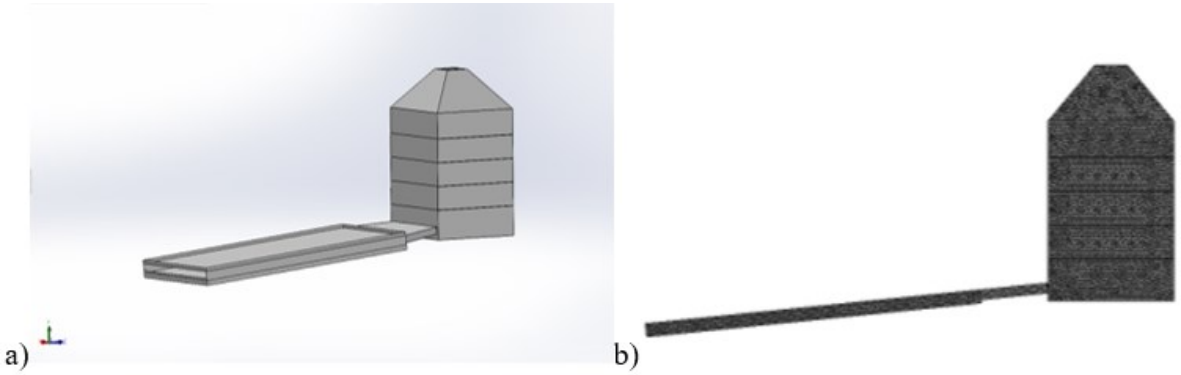


Figure 2. (a) solar dryer's 3D CAD model developed in Solidworks 2020; (b) refined mesh generated on the symmetric solid model.

3. Simulation

3.1. Methodology

Simulating a model on ANSYS, requires the implementation of correct governing equations and applying appropriate boundary conditions. In CFD, a finite volume method is used to resolve the governing equations of flowing fluid. This method involves dividing the flow domain into several control volumes and applying the fundamental laws of conservation on them [19].

For determining the characteristic properties of the fluid domain, several conservation laws are considered. The laws involved are stated below [13]:

Continuity equation: $\frac{\partial \rho}{\partial t} + \nabla.(\rho v) = 0$ (1)

Momentum conservation equation: $\frac{\partial}{\partial t}(\rho v) + \nabla.(\rho v v) = -\nabla p + \rho g + F$ (2)

Energy conservation equation: $\frac{\partial}{\partial t}(\rho v) + \nabla.[v(\rho E + p)] = 0$ (3)

where 'ρ' is the fluid's density, 'v' is the velocity with which the fluid flows, 'ρg' is the gravitational force, 'F' being the external force applied, static pressure is represented by 'p' and 'E' is the fluid's total energy.

CFD is applied to determine the airflow and temperature variation within the solar dryer. A mathematical model is generated by providing the necessary boundary conditions, and result data and contours are obtained. The flow of the CFD simulation process is depicted (Figure 3).

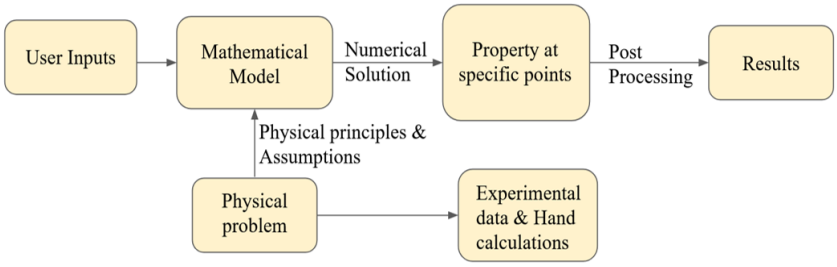


Figure 3. Flowchart representing CFD simulation process [14]

3.2. Meshing

Meshing is done for calculating the governing equations at the required points and defining the domain. To simulate the result on the analysis software, a refined tetrahedral shaped mesh with 120649 nodes and 621106 elements were generated over the completely assembled solid model [Figure 2(b)].

3.3. Simulation setup

For conducting simulation on ANSYS all the prerequisite information of the boundary conditions and variables should be known. As this study is demonstrating the simulation of experimental work presented by [8], the conditions are analyzed before setting up the model.

The solar dryer setup was constructed in Coimbatore Institute of Engineering and Technology, India, with latitude and longitude as 11.028831373336152 and 77.0272806. A realizable k-Ɛ model was selected to carry out turbulent computation as it produces accurate and satisfactory results. Similarly, energy equations were also turned on to account for thermal calculations. The conditions were applied considering 1300 hours of the day at the experiment location, which corresponded to maximum solar insolation of 996 W/m^2 and the ambient temperature of 32°C. The simulations were carried out for the mass flow rate of 0.0872, 0.0636, and 0.0447 kg/s.

Table 2. Boundary conditions

Region	Boundary condition	Value
Inlet	Mass flow- inlet	0.0872 kg/s
Walls	No-slip, convection	5 W/m^2
Absorber plate	Heat flux	996 W/m^2
Outlet	Pressure- outlet	0 Pa
Ambient temperature	-	32°C

4. Results and Discussions

The airflow velocity contours for the maximum mass flow rate are depicted in Figure 4(a) on the symmetry plane. It shows the novel design of ITSD to supply drying airflow in the dryer chamber. Figure 5 shows the velocity streamlines of the air flowing within the model. It is evident that the airflow travels smoothly through the collector before passing through the convergent passage and speeding up the flow. The porous aluminium trays with the agricultural products act as a barrier for the airflow and drastically slows it down. It is evident from Figure 4(a) that after passing through tray 1 the air velocity is reduced drastically. The airflow then speeds up again after passing through tray 4 to exit through the chimney. The average velocity at the collector's exit for 0.0872 kg/s mass flow rate is calculated at 2.89 m/s, while for a mass flow rate of 0.0447 (Figure 4(b)) is 1.36 m/s. From the experimental work carried out by [8], the average velocity at the exit of the collector was recorded as 3.09 m/s and 1.58 m/s for the mass flow rate of 0.0872 kg/s and 0.0447 kg/s, respectively. This helps to validate the results obtained from simulations with experimental results.

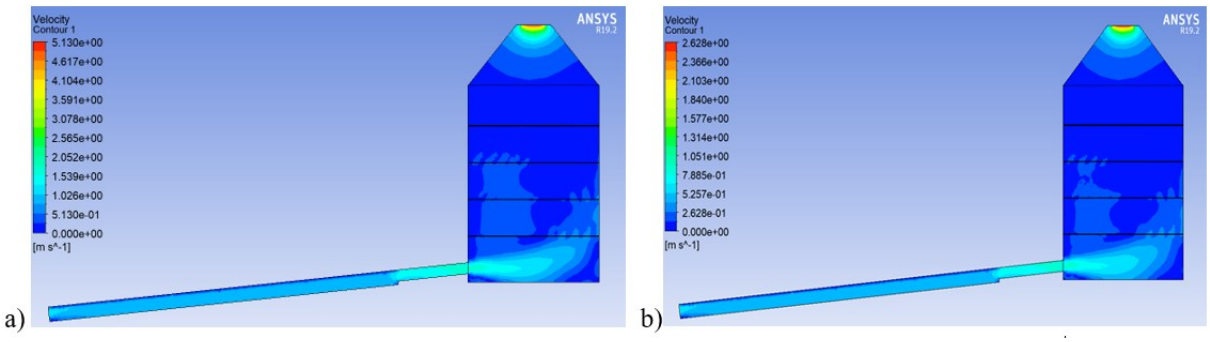


Figure 4. (a) Velocity contour corresponding to 0.0872kg/s rate (b) Velocity contour for mass flow rate of 0.0447 kg/s.

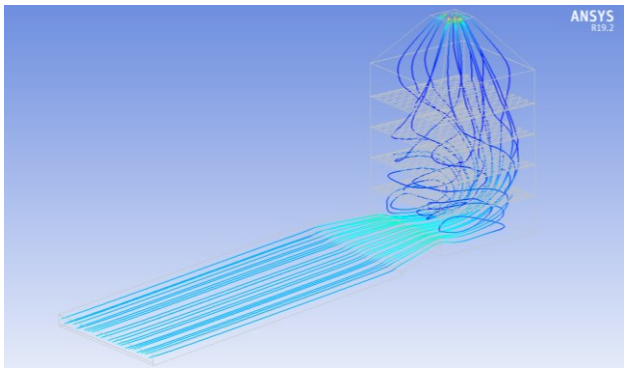


Figure 5. Velocity streamlines.

Utilizing and providing the right amount of heat to the dryer chamber is very important for an efficient ITSD. Figure 6 shows the temperature contours on the symmetry plane for different values of mass flow rates. It is evident that as the wind velocity is increased, a depreciation in the temperature in the collector and dryer chamber is observed (Figure 7). This trend is observed because as the velocity increases, the airflow has less time to absorb heat from the absorber plate and increase its temperature. The maximum collector outlet temperature and the maximum absorber plate temperature for the 0.0872kg/s mass flow rate was observed as 44.6°C and 69.57°C, respectively (Figure 6(a)), which also helps to verify the simulation results (Figure 8). Table 3 represents the variation in velocity & temperature at collector outlet and absorber plate with different mass flow rates input.

Table 3. CFD simulation results corresponding to different mass flow rate

Mass flow rate of air (kg/s)	Average velocity at the outlet of collector (m/s)	Maximum absorber plate temperature (°C)	maximum collector outlet temperature (°C)
0.0872	2.89	69.57	44.6
0.0636	2.12	73.72	46.2
0.0447	1.36	76.02	48.5

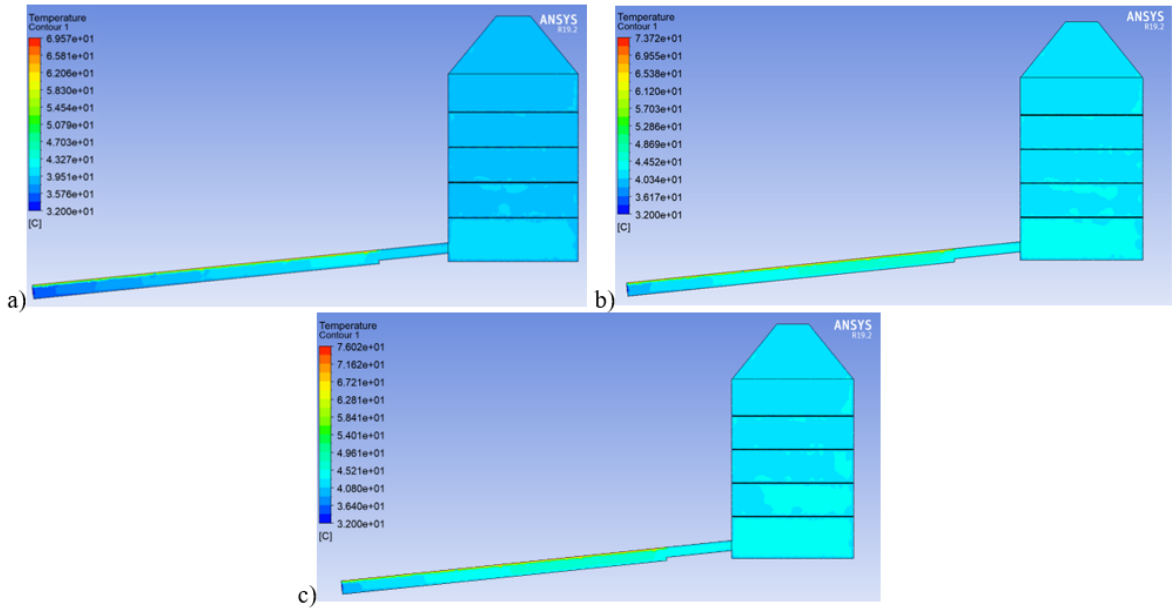


Figure 6. Temperature contours for various inlet mass flow rates (a) Flow rate = 0.0872 kg/s; (b) Flow rate = 0.0636 kg/s; (c)Flow rate = 0.0447 kg/s.

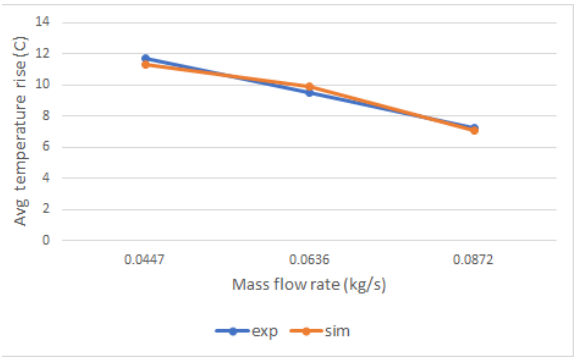


Figure 7. Average temperature rise for different air mass flow rates.

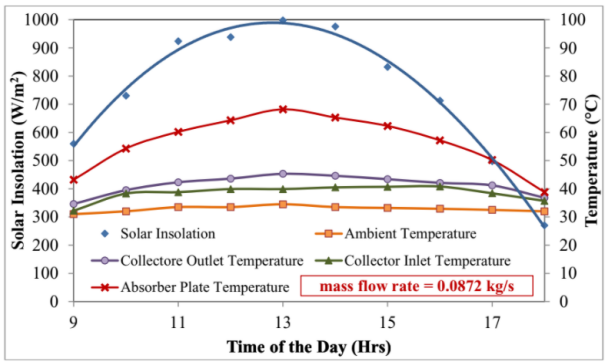


Figure 8. Solar insolation and temperature of several elements [8].

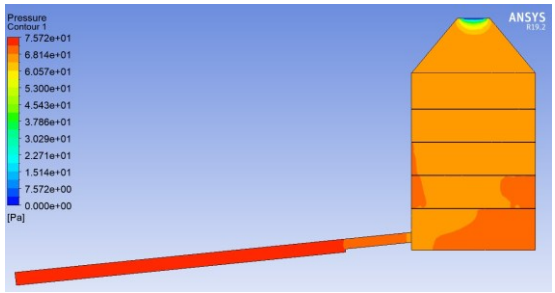


Figure 9. Pressure contour for inlet mass flow rate of 0.0872 kg/s

Figure 7 shows the comparative graph plot between the CFD simulation and the experimental results conducted by [8]. The results gained from the simulation resemble the experimental results. From gauge pressure contours as evident from Figure 9, it can be concurred that the pressure remains more or less the same inside the dryer chamber. Figure 10 displays the temperature contours of the 4 trays in

the dryer chamber. As evident from the figure, tray 1 (bottom most) temperature is maximum since it is the first tray to interact with the high temperature airflow from the collector. After passing the first tray, a considerable amount of heat is absorbed by the agricultural products, and the tray itself obstructs some flow. Therefore, the trend of lower temperature for higher placed trays is observed. The average temperature of tray 2 is considerably lower than tray 1 and trays 3 and 4 are at lower values than tray 2. Figure 11 shows the temperature distribution in four trays, corresponding to different mass flow rates.

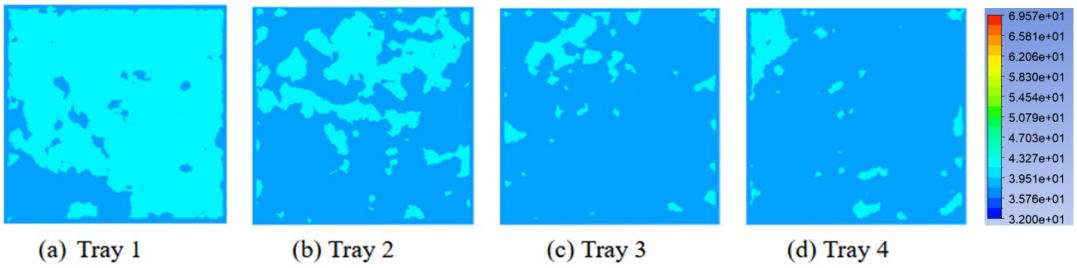


Figure 10. Temperature contours of trays developed at 0.0872 kg/s mass flow rate.

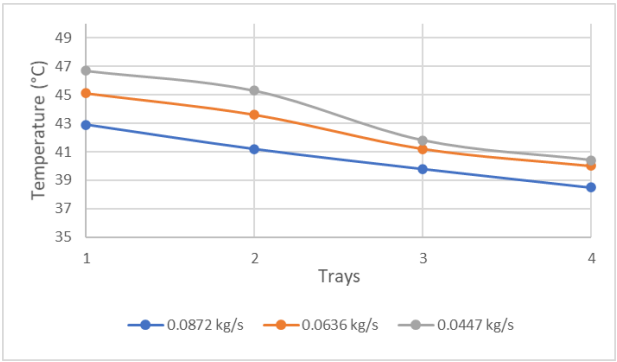


Figure 11. Average temperature of trays developed at different mass flow rates.

5. Conclusion

A CFD modelling and simulation on an active type indirect solar dryer was attempted. The temperature variation and flow velocity at different mass flow rates (0.0872, 0.0636 & 0.0447 kg/s) were predicted and compared with the experimental data. The velocity contours and streamlines were obtained and the results showed that the average velocity at the exit of the collector was 2.89 m/s for 0.0872 kg/s mass flow rate, while for a mass flow rate of 0.0447 kg/s was 1.36 m/s. The maximum collector outlet temperature was 44.6°C and the maximum absorber plate temperature 69.57°C for 0.0872kg/s mass flow rate. The average temperature rise difference was around 4°C for the two flow rates (0.0872 & 0.0447 kg/s). The temperature contours in the trays were analysed and it was observed that the bottom most tray attained the maximum temperature, while the uppermost tray had the lowest temperature. The CFD simulation provided an insightful output of the temperature distribution and velocity profile within the solar dryer.

Acknowledgments

The authors would like to thank Centre for Energy and Environment, Delhi Technological University, to provide the workspace and necessary equipment to carry out the required research.

References

[1] Solar Drying Technology. (2017). *Green Energy And Technology*

[2] D. K. Rabha and P. Muthukumar, "Feasibility Study of the Application of a Latent Heat Storage in a Solar Dryer for Drying Green Chili," 2018 2nd International Conference on Power, Energy and Environment: Towards Smart Technology (ICEPE), Shillong, India, 2018, pp. 1-6

[3] Prakash, O., & Kumar, A. (2013). Historical Review and Recent Trends in Solar Drying Systems. *International Journal Of Green Energy*, *10*(7), 690-738

[4] Jiskani, S., Ahmed Chandio, I., Mehdi, G., Memon, A., Raqeeb Bhutto, A., & Sandilo, U. (2020). Fabrication & performance analysis of direct type passive solar dryer for chilies and grapes drying. *2020 3Rd International Conference On Computing, Mathematics And Engineering Technologies (Icomet)*

[5] Prakash, O., Kumar, A., & Sharaf-Eldeen, Y. (2016). Review on Indian Solar Drying Status. *Current Sustainable/Renewable Energy Reports*, *3*(3-4), 113-120

[6] Lingayat, A., Chandramohan, V., Raju, V., & Meda, V. (2020). A review on indirect type solar dryers for agricultural crops – Dryer setup, its performance, energy storage and important highlights. *Applied Energy*, *258*, 114005

[7] Lan, N. (2017). Improvement of conventional solar drying system. *2017 International Conference On System Science And Engineering (ICSSE)*

[8] Vijayan, S., Arjunan, T., & Kumar, A. (2020). Exergo-environmental analysis of an indirect forced convection solar dryer for drying bitter gourd slices. *Renewable Energy*, *146*, 2210-2223

[9] Lingayat, A., Chandramohan, V., Raju, V., & Kumar, A. (2020). Development of indirect type solar dryer and experiments for estimation of drying parameters of apple and watermelon. *Thermal Science And Engineering Progress*, *16*, 100477

[10] Hamdi, I., & Kooli, S. (2018). Exergy and energy analysis of the solar drying processes of tomatoes in Tunisia. *2018 9Th International Renewable Energy Congress (IREC)*

[11] Luo, L., Liu, J., Cheng, J., Xi, J., Zhou, J., & Cen, K. (2010). Experimental Study on Solar Drying of Lignite. *2010 International Conference On Digital Manufacturing & Automation*

[12] Sanghi, A., Ambrose, R., & Maier, D. (2017). CFD simulation of corn drying in a natural convection solar dryer. *Drying Technology*, *36*(7), 859-870

[13] Demissie, P., Hayelom, M., Kassaye, A., Hailesilassie, A., Gebrehiwot, M., & Vanierschot, M. (2019). Design, development and CFD modeling of indirect solar food dryer. *Energy Procedia*, *158*, 1128-1134

[14] Jain, A., Sharma, M., Kumar, A., Sharma, A., & Palamanit, A. (2018). Computational fluid dynamics simulation and energy analysis of domestic direct-type multi-shelf solar dryer. *Journal of Thermal Analysis And Calorimetry*, *136*(1), 173-184

[15] M., P., T.N., V., & Santhosh, P. (2019). CFD Analysis of Greenhouse Solar Dryer with Different Roof Shapes. *2019 Fifth International Conference On Science Technology Engineering And Mathematics (ICONSTEM)*

[16] Meenakshi Reddy, R., Siva Reddy, E., Uma Maheswari, C., & Krishna Reddy, K. (2018). CFD and experimental analysis of solar crop dryer with waste heat recovery system of exhaust gas from diesel engine. *IOP Conference Series: Earth And Environmental Science*, *164*, 012010

[17] Moghimi, P., Rahimzadeh, H., & Ahmadpour, A. (2020). Experimental and numerical optimal design of a household solar fruit and vegetable dryer. *Solar Energy*, *214*, 575-587.

[18] Parikh, Darshit & Agrawal, G. (2011). Solar Drying In Hot and Dry Climate of Jaipur, India. International Journal of Renewable Energy Research

[19] Iranmanesh, M., Samimi Akhijahani, H., & Barghi Jahromi, M. (2020). CFD modeling and evaluation the performance of a solar cabinet dryer equipped with evacuated tube solar collector and thermal storage system. *Renewable Energy*, *145*, 1192-1213

2021 2nd International Conference for Emerging Technology (INCET)
Belgaum, India. May 21-23, 2021

CFOA Based Negative Floating Fractional Order Capacitance Multiplier

Rahul Gupta
Department of Electronics and Communication
Delhi Technological University
Delhi, India
rahul.gupta9246@yahoo.in

Charvi Agrawal
Department of Electronics and Communication
Delhi Technological University
Delhi, India
charviagrawal4@gmail.com

Ankit Gupta
Department of Electronics and Communication
Delhi Technological University
Delhi, India
ayangupta137@gmail.com

Rajeshwari Pandey
Department of Electronics and Communication
Delhi Technological University
Delhi, India
rpandey@dce.ac.in

2021 2nd International Conference for Emerging Technology (INCET) | 978-1-7281-7029-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/INCET51464.2021.9456355

***Abstract*- In this paper, a Negative Floating Fractional Order Capacitance Multiplier has been presented using a fractional-order capacitor and Current Feedback Operational Amplifier (CFOA). The paper depicts the working of the proposed design through PSPICE simulations. Simulations of three instances of Negative Capacitor are presented.**

The proposed designs have been verified using a commercially available Current Feedback Op-amp macro model (IC AD-844A) and a fractional capacitance (FC) using RC Domino Ladder Network. The simulation results closely verify our theoretical propositions.

Keywords- CFOA, Capacitance, Fractance, Multiplier, Fractional-order, RC-circuit, Impedance, Phase

I. INTRODUCTION

A negative capacitance is synthetic impedance which can be created using an active block and few passive circuit components. The negative capacitances find wide applications in electrical/electronic signal processing such as to cancel the parasitic capacitance of inductors, transformers and heat-sinks [1]. The successful applications of these cancellations in the circuits can greatly benefit EMI noise suppression. The concept of negative capacitance can be used for bandwidth enhancement of a circuit by cancelling the speed bottleneck capacitance. The negative capacitance shunts can be used to reduce structural vibrations of flexural systems [2].

It is well known that realizing a physically large capacitor in integrated circuits (ICs) is a very costly affair and involves complex fabrication process steps. In many applications the physical size of the capacitor needed presents a major challenge such as CMOS image sensor which requires large noise attenuating capacitor and monolithic phase locked loop filter. A capacitance multiplier circuit is an alternate solution to realize large capacitances and is based on the principle of tunable impedance converter. The capacitance multiplier circuits are built around active blocks and small physical capacitance. A variety of capacitance multiplier circuits are available in literature [3-6] and references cited their in. The concept of capacitance multiplication can be extended to negative capacitance circuits to obtain the desired value of negative capacitance.

It has been established through research works proposed in past two decades that fractional calculus [7] acts as a more efficient mathematical tool for modeling real-world objects and processes than its integral counter-part. The fractional- order (FO) circuits and systems are widely being used in the field of signal processing, control systems, biomedical instrumentation, and many more.

The aim of this paper is to generalize the narrow integer-order negative capacitance multiplier circuits to more general fractional-order counterparts having flexibility of varying its fractional order. Thus, this paper presents a fractional order negative capacitance multiplier using two CFOA, a fractional capacitor [8] and two resistances.

The CFOA [9] is an active element that offers relatively wider range of operation frequency and higher value of slew-rate as compared to conventional op-amps owing to the use of current as the feedback signal.

Rest of the paper is structured as below: Section II highlights the design methodology of proposed negative fractional order floating capacitance multiplier. The simulation results are compiled in Section III to draw constructive inferences. The application of the proposed circuit is shown in Section IV and the conclusions have been presented in section V.

II. PROPOSED CONFIGURATION

A fractional order Laplacian operator is used to represent FO circuits. Using Laplace transform the fractional order operator can be simplified into integer approximation and thus it becomes physically realizable with fractional order elements (FOE).

The impedance function of a simple fractional-order element is given as,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:27:12 UTC from IEEE Xplore. Restrictions apply.

$$Z(s) = (s^{\alpha}F) \quad (1)$$

where F is known as the fractance and α (a fractional number) is known as the exponent of the FOE. A positive α corresponds to a fractional inductor (FI) while negative fraction corresponds to fractional capacitor (FC). The impedance and phase angle of an FC are expressed respectively as [8]:

$$|Z(j\omega)| = 1/(\omega^{\alpha}F) \quad (2.a)$$

$$\theta = -\alpha\pi/2 \quad (2.b)$$

As of now the FCs are not commercially available however, can be emulated based on various structures [10] such as passive resistor-capacitor (RC) elements arranged in the form of RC tree, cross RC ladder, domino ladder etc. obtained using various rational approximations. In this work a truncated RC (resistor–capacitor) domino ladder-type network based on continued fraction Expansion (CFE) is used. The RC ladder network and the symbol for FC are shown Fig.1.

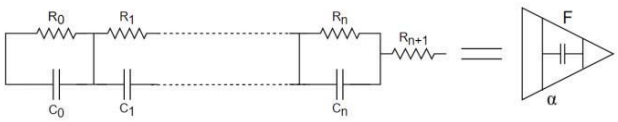


Fig. 1. RC Domino Ladder for Fractance and its corresponding symbol [8]

The symbolic diagram of CFOA is shown in Fig. 2. It consists of four ports. Considering the direction of port currents towards the CFOA block the port relationship, in matrix form [11] is given as:

$$\begin{bmatrix} I_y \\ V_x \\ I_z \\ V_w \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \pm 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} V_y \\ I_x \\ V_z \end{bmatrix}$$

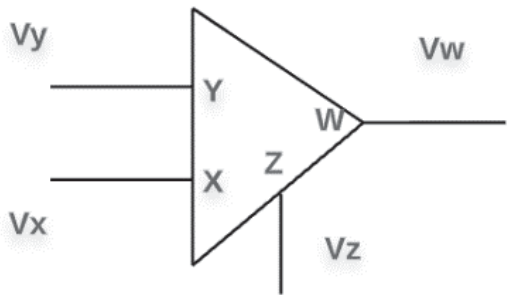


Fig. 2. Block Diagram of CFOA

The proposed Negative Floating Fractional Capacitance Multiplier is shown in fig. 3.

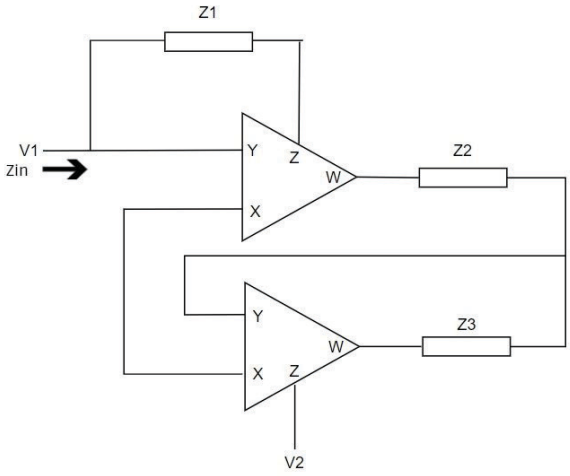


Fig. 3. Negative Floating Fractional Capacitance Multiplier

Expression for input impedance Z_{in}, is expressed as (3) [5]:

$$\frac{V_1 - V_2}{I_{in}} = -\frac{Z_1 Z_3}{Z_2} \quad (3)$$

From Equation (3), it can be observed that two configurations of the circuit are possible which are shown below:

Configuration 1: Selecting Z_1 as FC and Z_2 and Z_3 are resistors as shown in Fig.3 the impedance of the structure can be expressed as:

$$Z_{in} = -\frac{R_3}{s^{\alpha}FR_2} \quad (4)$$

Thus, Z_{in} is representing a Negative Fractional-order Capacitance of value,

$$C_{eq} = -K_{m1}s^{\alpha}F \quad (5)$$

Here K_{m1} denotes the multiplication factor whose mathematical representation is given in (5)

$$K_{m1} = \frac{R_2}{R_3} \quad (6)$$

The magnitude and phase angle of the proposed circuit are expressed as (6):

$$|C_{eq}(j\omega)| = K_{m1}(\omega)^{\alpha}F \quad (7.a)$$

$$\theta = \pi - \alpha\pi/2 \quad (7.b)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:27:12 UTC from IEEE Xplore. Restrictions apply.

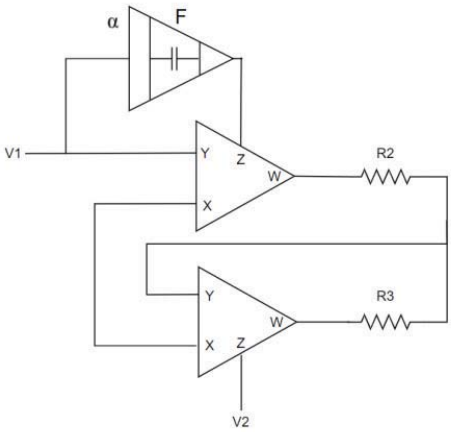


Fig. 4. Proposed configuration 1 of Negative Floating Fractional Capacitance

Configuration 2: If Z_3 is taken as FC and Z_1 and Z_2 are resistors as shown in fig. 4 then the impedance can be calculated as:

$$Z_{in} = -\frac{R_1}{s^{\alpha} F R_2} \tag{8}$$

thus, Z_{in} is representing a Negative Fractional-order Capacitance of value,

$$C_{eq} = -K_{m2} s^{\alpha} F \tag{9}$$

Here K_{m2} denotes the multiplication factor whose mathematical representation is given in (10)

$$K_{m2} = \frac{R_2}{R_1} \tag{10}$$

The magnitude and phase angle of the proposed circuit are expressed as (11):

$$\left|C_{eq}(j\omega)\right| = K_{m2}(\omega)^{\alpha} F \tag{11.a}$$

$$\theta = \pi - \alpha\pi/2 \tag{11.b}$$

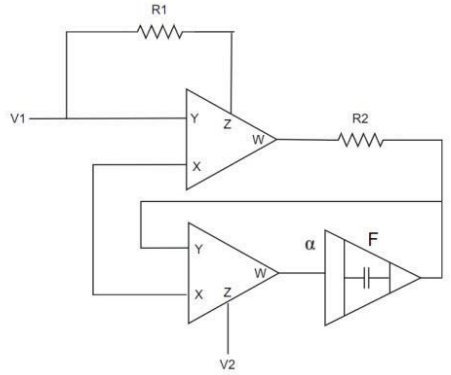


Fig. 5. Proposed configuration 2 of Negative Floating Fractional Capacitance

Thus, the magnitude and the order of negative FC can be varied by varying the multiplication factor K_{m1} or K_{m2} (as the case may be) the value of α. Thus, two degrees of freedom are possible in this configuration.

III. Simulation Results

The functionality of the proposed circuit is verified through SPICE simulations using AD844A macro model of CFOA with ±5 V supply voltages. Simulations are performed for two FC instances (i) F=12.5μF, α=0.3; (ii) F=3.75 μF, α=0.5 for which the components are listed in Table 1.

For F=12.5μF and α=0.3, the impedance magnitude is plotted for different values of K_m in Fig. 6 and it can be observed that the impedance decreases with the increase in frequency which confirms the expression for capacitance relation. The phase is plotted for the same in fig. 7 and yields a nearly constant phase which closely resemble the theoretical value of 153° for values of K_m 1-10 and starts deviating for larger K_m. Lastly for the same F and α, the transient response for phase relation between voltage and current is shown in fig. 8.

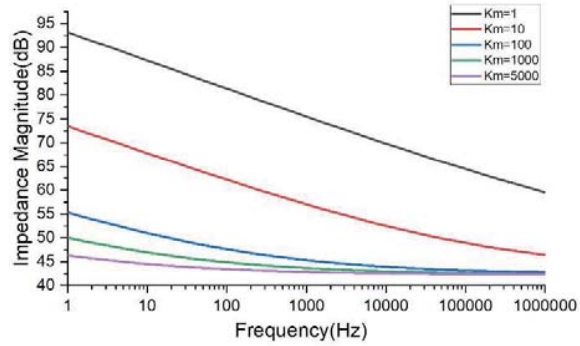


Fig. 6. Impedance Magnitude Response for **α**=0.3 and F=12.5u

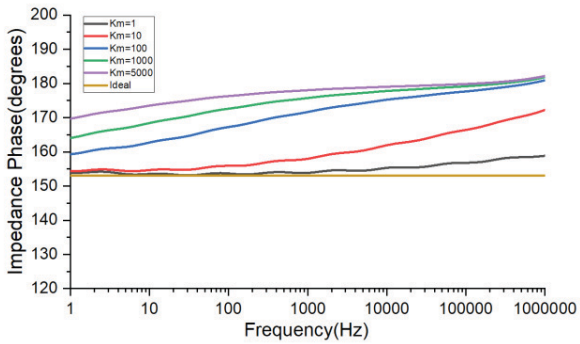


Fig. 7. Phase Magnitude Response for **α**=0.3 and F=12.5u

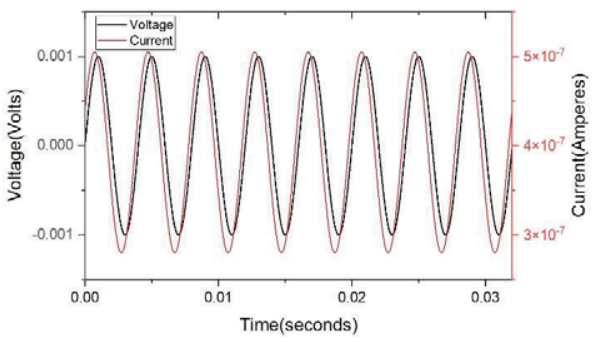


Fig. 8. Transient Response of Negative Capacitance with α=0.3. F=12.5u and K_m =1

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:27:12 UTC from IEEE Xplore. Restrictions apply.

For F=3.75u and α=0.5, the impedance magnitude is plotted for different values of K_m in fig. 9 and it can be observed that the impedance decreases with the increase in frequency which confirms the expression for capacitance relation. The phase is plotted for the same in fig. 10 and yields a nearly constant phase which closely resemble the theoretical value of 135° for values of K_m 1-10 and starts deviating for larger K_m. Lastly for the same F and α, transient response for phase relation between voltage and current is shown in fig. 11.

By examining the magnitude and phase plot of the input impedance of all the different instances, the variability of the multiplication factor can be defined up to 1000. Some deviations from ideal plot are observed, which are majorly due to internal parasitic elements of CFOA, thus limiting the operational bandwidth. Therefore, the multiplier circuit has a practical frequency range from 1 Hz to 10 KHz.

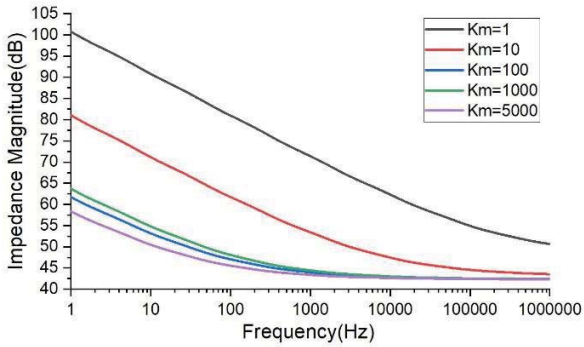


Fig. 9. Impedance Magnitude Response for α=0.5 and F=3.75u

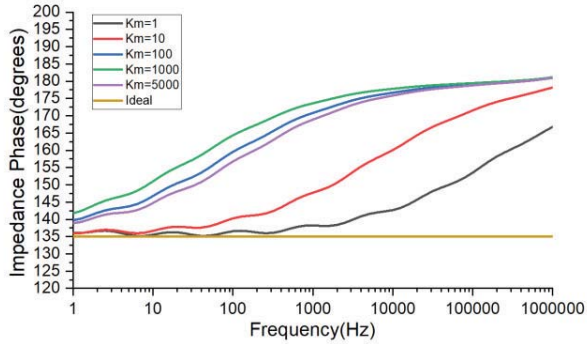


Fig. 10. Phase Magnitude Response for α=0.5. and F=3.75u

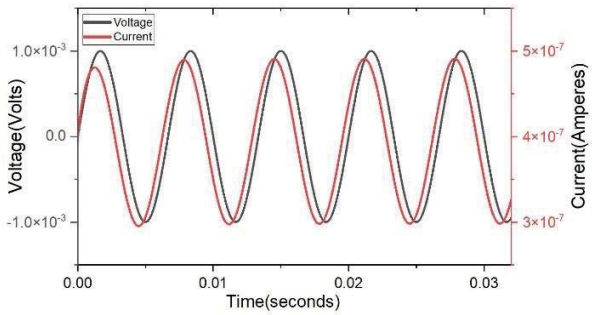


Fig. 11. Transient Response of Negative Capacitance with α=0.5. F=3.75u and K_m =1

IV. APPLICATION

During circuit designing often stray capacitances are generated which can hamper the performance. These capacitances need not be of integral order, therefore in such case a negative fractional capacitance multiplier can be used to design cancellation circuits to get rid of such problems. Fig. 12 below shows an accurate description of this application. The circuit consists of a resistor in series with a parallel combination of a parasitic positive fractional capacitor C_1 and proposed negative fractional capacitor C_2, which can be easily tuned to the exact value in order to cancel the effect of C_1 using the proposed design.

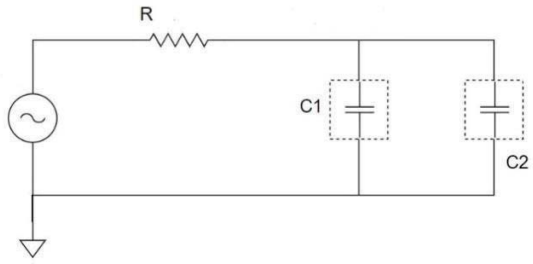


Fig. 12. Circuit for Capacitor Cancellation

The input impedance of the circuit in fig. 12 is expressed as (12)

$$Z_{in} = R + \frac{1}{s(C_1 - |C_2|)} \tag{12}$$

By properly tuning the domino capacitance using the proposed fractional multiplier circuit we can make the circuit purely resistive, $C_1 = |C_2|$, reducing expression (12) to (13).

$$Z_{in} = R \tag{13}$$

TABLE I. VALUES OF ELEMENTS OF RC LADDER [2]

	Ladder elements	Block Numbers												
		0	1	2	3	4	5	6	7	8	9	10	11	12
F=12.5 μF α=0.3	R (Ω)	46.9k	21k	12.6k	7.61k	4.58k	2.76k	1.66k	999	601	362	218	65.6	259
	C (F)	22.2u	9.12u	2.79u	854n	261n	80n	24.5n	7.49n	2.29n	701p	215p	131p	
F=3.75 μF α=0.5	R(Ω)	336k	83.7k	32.5k	12.6k	4.89k	1.90k	736	286	111	43	8.34	17.6	
	C (F)	4.98u	3u	1.16u	452n	176n	68.1n	26.4n	10.3n	3.98n	1.54n	1.2n		
	C (F)	149.3n	151.2n	77.1n	39.3n	20.1n	10.2n	5.2n	2.66n	1.36n	1.38n			

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:27:12 UTC from IEEE Xplore. Restrictions apply.

The simulation of the cancellation of capacitors as proposed in fig. 12 is shown in fig. 13 wherein the current and voltage are in phase as the effect of the fractional capacitor is canceled by the negative fractional capacitor tuned to the appropriate value thereby making the circuit purely resistive.

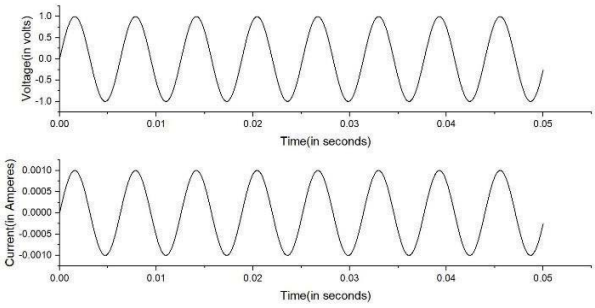


Fig. 13. Transient Response showing cancellation

V. CONCLUSION

In this paper, a negative floating fractional capacitance multiplier designed using CFOA has been presented. The proposed design is verified with the use of SPICE simulations for fractional-order of 0.3 and 0.5. The behavior of the proposed design was tested by comparing the theoretical impedance magnitude and phase responses for different values of the multiplication factor with the ideal ones.

The proposed capacitance multiplier shows promising results in generating high value fractional capacitances with minimum hardware requirement and also showed successful outcomes in cancellation circuit applications.

REFERENCES

[1] S. Wang and F. C. Lee, "Negative Capacitance and its Applications on Parasitic Cancellation for EMI Noise Suppression," *2007 IEEE Power Electronics Specialists Conference*, Orlando, FL, 2007, pp. 2887-2891

[2] Benjamin S Beck, Kenneth A Cunefare and Manuel Collet, "The power output and efficiency of a negative capacitance shu nt for vibration control of a flexu ral system", Published 30 April 2013, 2013 IOP Publishing Ltd

[3] S. Unhavanich, O. Onajn, and W. Tangsrirat “Tunable Capacitance Multiplier with a Single Voltage Differencing Buffered Amplifier,” In Proc. of the International MultiConference of Engineers and Computer Scientists (IMECS), Hong Kong, China, 2016, pp. 1–4.

[4] W. Tangsrirat, “Resistorless tunable capacitance multiplier using single voltage differencing inverting buffered amplifier,” Rev. Roum. Sci. Techn.– Électrotechn. et Énerg., vol. 62, no. 1, pp. 72–75, 2017.

[5] W. Tangsrirat, and O. Channumsin, “Tunable Floating Capacitance Multiplier Using Single Fully Balanced Voltage Differencing Buffered Amplifier,” J. Commun. Technol. Electron, vol. 64, pp. 797–803, 2019.

[6] M. A. Al-Absi and A. A. Al-Khulaifi, “A New Floating and Tunable Capacitance Multiplier with Large Multiplication Factor,” IEEE Access, vol. 7, pp. 120076–120081, 2019.

[7] A. G. Radwan and K. N. Salama, "Passive and Active Elements Using Fractional $L_\alpha C_\beta$ Circuit," IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 58, no. 10, pp. 2388-2397, Oct. 2011.

[8] A. Adhikary, P. Sen, S. Sen and K. Biswas, “Design and Performance Study of Dynamic Fractors in Any of the Four Quadrants,” in Circuits, Systems, and Signal Processing , vol. 35, no. 6, pp. 1909–1932, Jun. 2016.

[9] F.J. Lidgey, K. Hayatleh, 'Current-feedback operational amplifiers and applications', Electronics & Communication Engineering Journal, 1997, 9, (4), p. 176-182.

[10] R. Verma, N. Pandey, & R. Pandey, "Realization of a higher fractional order element based on novel OTA based IIMC and its application in filter" in Analog Integr Circ Sig Process 97, 177–191 (2018).

[11] S. Singh, N. Pandey, R. Pandey. (2020) CFOA Based Negative Floating Capacitance Multiplier. In: Zakaria Z., Ahmad R. (eds) Advances in Electronics Engineering. Lecture Notes in Electrical Engineering, vol 619. Springer, Singapore

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:27:12 UTC from IEEE Xplore. Restrictions apply.

2021 2nd International Conference for Emerging Technology (INCET)
Belgaum, India. May 21-23, 2021

CNN-based Person Recognition System for Masked Faces in a post-pandemic world

Parangat Mittal
Dept. of Electronics & Communication Engineering
Delhi Technological University
Delhi, India
parangatmittal1@gmail.com

Kartikeya Pandey
Dept. of Electronics & Communication Engineering
Delhi Technological University
Delhi, India
kartiikey1711@gmail.com

Peeyush Tawani
Dept. of Electronics & Communication Engineering
Delhi Technological University
Delhi, India
tawanipeeyush@gmail.com

Prof. Rajesh Rohilla
Dept. of Electronics & Communication Engineering
Delhi Technological University
Delhi, India
rajesh@dce.ac.in

2021 2nd International Conference for Emerging Technology (INCET) | 978-1-7281-7029-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/INCET51464.2021.9456416

***Abstract*—With everyone now covering their faces with a face mask to avoid contagion of the COVID-19 virus, it becomes an increasingly difficult challenge for the facial recognition systems to identify people wearing masks. Algorithms devised before the pandemic for facial recognition often fail in this context, and consequently a need to understand the working of facial recognition algorithms when presented with occluded faces arises. This project aims to develop a new lightweight Convolutional Neural Network-based algorithm to resolve this issue. The proposed model gives a comparable accuracy with similar models developed in the past. Further, the proposed algorithm is used to create a robust system to ensure adherence of COVID-19 protocols in a real-world environment.**

Keywords— Artificial Intelligence, Convolutional Neural Networks, Face Mask Detection, Masked Face Recognition

I. INTRODUCTION

COVID-19 is an unprecedented crisis leading to a huge number of casualties and security concerns. Several studies on this new virus have proven that the act of wearing a face mask by a healthy as well as an infected person considerably reduces the transmission of this virus. Many guidelines have been issued regarding the mandatory use of masks in public places as well as maintaining a certain distance between people but the main cause of concern is how to enforce these guidelines in a real-world setup where the authorities are responsible for the general health and safety. Another cause of concern is that since the virus is highly communicable, the standard tactile security protocols in any organization such as biometric systems based on fingerprints and passcodes are not safe.

Face recognition is one of the best methods to deal with the above-mentioned problems as it does not require any physical touch and once the person has been identified, he could easily be notified if the mandatory guidelines are not being followed. The problem which arises here is that the face mask makes face recognition a very troublesome task since parts of the face are concealed.

An important difference that needs to be noted at this point is that between Face Mask Detection and Masked Face Recognition. The former indicates whether or not someone is wearing a face mask, while the latter is indicative of who is wearing the face mask. The two must not be confused and must be dealt with separately.

Wearing a face mask causes the following problems:

- Criminals and Fraudsters commit various crimes without being recognised, seeking the perfect cover of a face mask.
- Facial authentication becomes a very tough task when a major portion of one's face is concealed by a mask.
- Present-day face recognition algorithms are not efficient when treated on someone wearing a mask which fails to provide the whole facial image for description.
- Exposing merely the periocular region poses a tough challenge to the task of face recognition.

Due to these problems, face masks have adversely affected the performance of current face recognition methods [1]. While occlusion-resistant Convolutional Neural Networks are suited best for this task, concealed parts of the face make it somewhat difficult for distinct recognition [2]. To increase the reliability of the system, image samples of an individual with and without a face mask can been incorporated in the facial recognition systems to uniquely recognize a person without their knowledge or physical interaction.

In this paper, we are proposing a complete methodology to tackle all the above-mentioned issues. We are proposing a person identification system for detecting and recognising masked faces which would replace the existing identification systems in a post-pandemic world. The proposed system is designed to tackle the following tasks:

1. Detect a person in a frame (Human Detection)
2. Detect if a person is wearing a face mask or not. (Face Mask Detection)
3. Recognise a person who is wearing a mask. (Masked Face Recognition)

The rest of this paper is organized as follows: Section 2 presents the previous/related works. The proposed

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:25:26 UTC from IEEE Xplore. Restrictions apply.

methodology is detailed in Section 3. Experimental results are presented and discussed in the subsequent section (Section 4). Conclusion ends the paper.

II. Previous/Related Works

The main contribution of this work is proposal of new models for Face Mask Detection and Masked Face Recognition, and a complete person identification system incorporating both. We will first review state-of-the-art face detection models, and then move to the datasets available for this new problem.

A. Face Recognition

Face recognition is a computer vision task on which several studies have been performed till now [3]. When Face recognition is compared with other physical recognition systems like biometrics, fingerprint and iris, the former has a much better chance to recognize the identity of a person in a discrete manner. Hence, widely used and has many applications such as security monitoring, forensic analysis, and border control. With the growth of deep neural network-based techniques [4]-[6], and the publicly as well as freely available large-scale face datasets [7],[8], face recognition performance has improved by leaps and bounds [9],[10]. Performance of all the deep learning models as well as State-of-the-Art techniques is barely satisfactory when facing challenges in the form of varying facial expressions, poor resolution, diverse illumination, large-pose differences and occlusion. A summary [11] of various state-of-the-art networks has been presented in Table 1.

TABLE I. State-of-the-art Face Recognition Models

Method	Architecture	Accuracy
FaceNet [12]	GoogleNet-24	99.63 %
DeepFace [13]	AlexNet	97.35 %
DeepID2 [14]	AlexNet	99.15 %
DeepID3 [15]	VGGNet-10	99.53 %
VGGFace [16]	VGGNet-16	98.95 %
SphereFace [17]	ResNet-64	99.42 %

The ongoing pandemic has also affected these face recognition models adversely. All people are forced to wear a Face Mask to curb the spread of the virus. Extensive research is being conducted to incorporate the presence of masks. Performance of all the state-of-the-art face recognition models dealing with people without the mask give great results [18] but fail when people are wearing masks. One such study [19] discusses this problem and proposes a method for masked face recognition with an accuracy of 91.3%.

B. Datasets for Face Recognition

Several datasets have been developed over the years to train and evaluate facial recognition models. Earlier datasets for the same problem were very limited and controlled since the image samples were collected manually. Recent datasets like MegaFace [7], CASIA WebFace [20], IMDB-Faces [21], etc. collect the image samples from the internet as well. These datasets, hence, capture the real-world faces and scenarios with a greater precision and detail. The newer datasets also provide a lot of information about the faces. VGGFace2 [22] provides annotations in the form of depth, cross pose, age, sex and ethnicity. Similarly, WIDER FACE [23] provides various attributes like positions, gender, landmarks, and poses.

For the purpose of this paper, facial datasets containing images of individuals wearing masks as well as without masks are required. Not many such datasets were developed prior to the COVID-19 pandemic since this problem arose as a result of the same. The publicly-available datasets for this problem are MAsked FAces (MAFA) [2] dataset, MaskedFace-Net [24] and Real-world Masked Face Dataset (RMFD) [25] consisting of Masked Face Detection Dataset (MFDD), Real-world Masked Face Recognition Dataset (RMFRD) and Simulated Masked Face Recognition Dataset (SMFRD). All these datasets have been compared in Table 2.

TABLE II. Datasets for Masked Face Recognition

Dataset	Number of Images	Key Features
Masked Faces (MAFA) Dataset [2]	35,806	Images contain masks as well as other occlusions
Masked Faces Detection Dataset [25]	24,711	Contains coordinates of masked faces and if mask is not present
Real-world Masked Face Recognition Dataset (RMFRD) [25]	5,000 masked and 90,000 unmasked	Contains masked as well as unmasked faces
Simulated Masked Face Recognition Dataset (SMFRD) [25]	500,000	Masked counterparts of LFW and WebFace datasets
MaskedFace-Net [24]	137,016	Contains information about correctly and incorrectly worn face masks

III. Methodology

In this paper, we present a person recognition system which will aid in recognizing individuals with whatever part of their face is visible due to the face mask. This system comprises a three-step as shown in Fig. 1, with each stage catering to a specific standalone task.

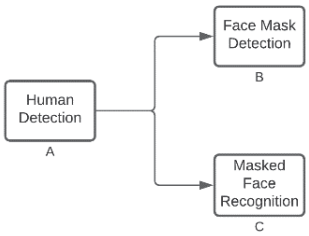


Fig. 1. Stages of the Proposed System.

The first stage, Human Detection, detects the distinct people in a video frame (generally from a CCTV). Once the people are detected, the second stage, Face Mask Detection, checks if the person is wearing a face mask. The last stage, i.e., Masked Face Recognition, confirms the identity of a person who is wearing a face mask for proper authentication.

A. Human Detection

The first function in this task is detecting the people in a video frame, for which already a lot of CNN-based object detection models have been established [26],[27]. Out of all the state-of-the-art methods, YOLO [28] outperforms everyone. This paper uses the same algorithm to achieve the first task.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:25:26 UTC from IEEE Xplore. Restrictions apply.

YOLO v3 model trained on the MS COCO dataset was used in this study, with the detection label "Person". The architecture for the model is shown in Fig. 2.

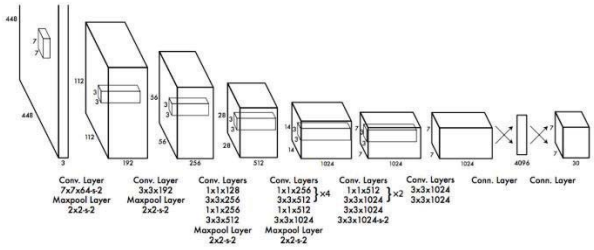


Fig. 2. YOLO v3 Architecture [28].

B. Face Mask Detection

Face mask detection performs the task of verifying whether or not the detected person is covering his face with a mask or not. To achieve this task, we propose the use of a simple Convolutional Neural Network described as follows and illustrated in Fig. 3, which has been trained on images with labels being as masked or unmasked. The output of this stage is binary, since only two classes are being considered – mask or no mask.

The model proposed for this task is as follows:

- **Convolutional Layer 1** consists of 32 kernels of size (3, 3) and a padding of (1, 1), combined with a ReLU operator, and followed by a maxpool with a kernel size of (2, 2).
- **Convolutional Layer 2** consists of 64 kernels of size (3, 3) and a padding of (1, 1), combined with a ReLU operator, and followed by a maxpool with a kernel size of (2, 2).
- **Convolutional Layer 3** consists of 128 kernels of size (3, 3), stride 3 and a padding of (1, 1), combined with a ReLU operator, and followed by a maxpool with a kernel size of (2, 2).
- **Linear Layer** consists of 1024 neurons combined with ReLU.
- **Output Layer** consists of 2 neurons for each of the output class.

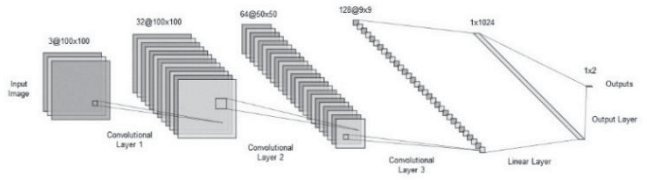


Fig. 3. Proposed Model for Face Mask Detection.

C. Masked Face Recognition

The task of masked face recognition, as explained in the preceding sections, is different from the conventional face recognition task mainly due to unavailability of the complete facial region for processing. Dlib [29] has devised a 68-point facial landmarks model which are considered for CNN-based facial recognition tasks, but in the scope of the current problem maximum 28 facial landmarks can be available in an input image, as shown in Fig. 4. This is with the underlying assumption that the face mask is being properly worn.

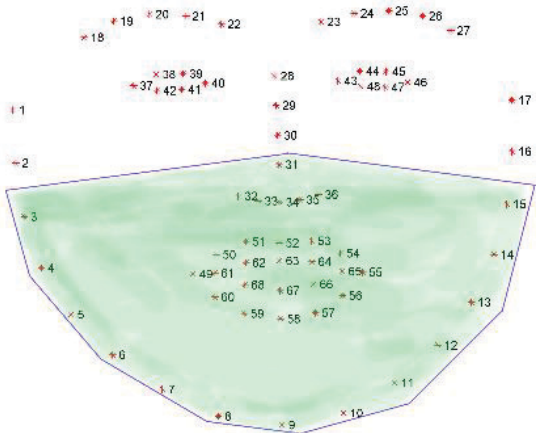


Fig. 4. Facial landmarks in the presence of a mask.

As explained already, a number of occlusion-resistant facial recognition methods have been devised which take into account every kind of obstructions possible. Since we need to account for occlusion only on the lower part of the face, we propose a Convolutional Neural Network as shown in Fig. 5. Since it tackles a specific section obstruction, it is significantly smaller in size and resource-friendly.

The details for the proposed Convolutional Neural Network for this task are as follows:

- **Convolutional Layer 1** consists of 24 kernels of size (10, 10), stride 2 and a padding of (1, 1), combined with a leaky ReLU operator, followed by a maxpool with a kernel size of (2, 2) and a batch normalisation operation.
- **Convolutional Layer 2** consists of 36 kernels of size (9, 9), stride 2 and a padding of (1, 1), combined with a leaky ReLU operator, followed by a maxpool with a kernel size of (2, 2) and a batch normalisation operation.
- **Convolutional Layer 3** consists of 48 kernels of size (8, 8), stride 2 and a padding of (1, 1), combined with a leaky ReLU operator, followed by a maxpool with a kernel size of (2, 2) and a batch normalisation operation.
- **Linear Layer 1** consists of 256 neurons combined with ReLU.
- **Linear Layer 2** consists of 64 neurons combined with ReLU.
- **Output Layer** consists of neurons combined with ReLU equal to the number of output classes.

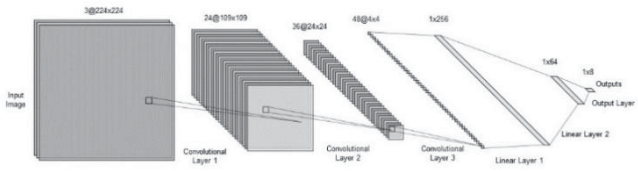


Fig. 5. Proposed Model for Masked Face Recognition

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:25:26 UTC from IEEE Xplore. Restrictions apply.

Due to the limited number of facial landmarks available, the image cannot be processed directly by the network; some level of pre-processing is required in the input image to make it ready for the network.

For a uniform feature extraction from the partially occluded images, it is therefore necessary to align all the images in a similar fashion. Since eyes are the only major facial feature left exposed to the real world while wearing a face mask such that it covers nose, mouth and chin, image alignment can be done using eyes as a reference. In this paper, we consider the line joining the two detected eyes using a Haar Cascade Classifier [30] as the horizontal axis with which the images are aligned.

The pipeline for image processing is as follows (Fig. 6):

1. Detect the face under question and crop it out from the whole image as Region of Interest (ROI)
2. Detect Eyes in the input image and align the image with the line joining both the eyes as horizontal axis
3. Resize and Center Crop the image to 224 x 224 pixels size.

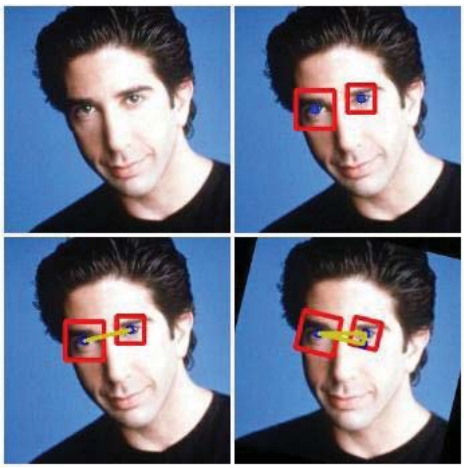


Fig. 6. Recognition Pipeline.

IV. Results And Discussion

A. Dataset

Real-world Masked Faces Dataset (RMFD) [25], a publicly available free dataset, has been used in this study for the tasks of Face Mask Detection and Masked Face Detection. This dataset has laid a virtual mask on all the images of the CASIA-WebFace Dataset [20]. Similar to the CASIA WebFace, our complete dataset also contains 494,414 images over 10,575 identities after face detection.

Images of masked faces from the RMFD [25] and those of unmasked faces from the CASIA-WebFace [20] for each of the identities were included while creation of the training dataset, since in the real-world, not everyone is seen wearing the face mask, detection of which is one of the purposes of this study. The validation or the testing set contained only the images of masked faces. Image samples from the dataset containing masked images and unmasked images of 4 distinct classes have been presented in Fig. 7.

Fig. 7. Unmasked and Masked Image samples.

B. Training and Evaluation

All the images in the dataset were resized to 224 x 224 pixels for uniformity. The grayscale images were removed from all the classes since the models were designed for 3-channel inputs. In this study, the Stochastic Gradient Descent (SGD) and Adaptive Moment Estimation (ADAM) were used as optimizers. Cross Entropy loss was minimized while training the network. The hyper-parameters such as Batch Size, Training Epochs, Learning Rate, Learning Rate Decay Step Size were suitably modified to achieve higher accuracies and lower losses.

The Human Detection Task [28] has already been tested and evaluated since it is the standard detection model. The results for the tasks of Face Mask Detection and Masked Face Recognition are presented in Table 3. Both the proposed models have been compared against a standard state-of-the-art Facenet [12] Inception Resnet v1 model in Table 4.

TABLE III. Results of the Proposed Models

Task	Proposed Model Accuracy	Proposed Model Estimated Size
Face Mask Detection	99.4%	8.23 MB
Masked Face Recognition	97.7%	4.46 MB

TABLE IV. Comparison with Standard Facenet Model

Model	Estimated Size	Floating Point Operations	No. of Parameters
FaceNet Inception Resnet v1	281.39 MB	3,170,217,856	23,486,728
Proposed Model for Masked Face Recognition	4.46 MB	128,352,088	254,476
Proposed Model for Face Mask Detection	8.23 MB	64,460,096	2,193,474

C. Analysis

a) Masked Face Recognition:

In this paper, the proposed CNN for the masked face detection achieves an accuracy of 97.7% which is far better than the earlier state-of-the-art models that were used in the pre-pandemic era. The model has a size of merely 4.46 MB.

It was noted that when the training set consisted of only the masked images of each individual, the model did not achieve a high accuracy in recognition. But when the unmasked images, i.e., images containing the full face, as well as masked images of the individuals were used in the training set, the model achieved stellar accuracy. This allows the feature extractor in the proposed model to

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:25:26 UTC from IEEE Xplore. Restrictions apply.

understand the face completely, and recognise it even when some of it is covered with a mask, as depicted in Fig. 8.

A study [1] in reference to these pre-pandemic state-of-the-art models showed their accuracy to decrease significantly under the current situation. The main problem arising with the earlier state-of-the-art models was that the part of the face occluded by the face mask was being considered by the model for feature extraction. Since no facial feature was available, the model performed poorly. The model proposed in this paper uses only the visible part of the face for extracting the information, resulting in a better performance for this task.

The other models aimed for this task tried to reconstruct the part of the face that was concealed by the mask. Since this is a very resource intensive task such models were unfit and inefficient for real time scenarios. As mentioned above, the proposed model is smaller in size and such a small size makes it suitable for real time applications.

b) Face Mask Detection:

The model proposed for the task of face mask detection in this paper achieves a very high accuracy and is extremely compact in size as well. The accuracy achieved by the face mask recognition is 99.4% with the model size being around 8MB.

The proposed model can be used efficiently in real-life situations due to its compact size. The model detects the face mask only when the mask is worn properly by an individual, i.e., when it completely covers the nose, mouth as well as the chin, as shown clearly in Fig. 9. Whenever any of these facial regions is visible, the model detects it as a "No-Mask" condition, thus making it ideal for ensuring COVID-19 protocols in the real world.

This model acts as a classifier differentiating between the "Mask" and "No-Mask" conditions. It has been trained on the masked images from the RMFRD [26] dataset as a whole nested under the "Mask" condition, while the unmasked images from the CASIA-WebFace [20] dataset as the "No-Mask" condition. Since the images of all the individuals were under a single class, the image samples significantly increased, thus leading to a high accuracy.

Fig. 8. Experimental Results for Masked Face Recognition Task.

Fig. 9. Experimental Results for Face Mask Detection Task.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:25:26 UTC from IEEE Xplore. Restrictions apply.

V. Conclusion

This paper proposes a novel person identification system and details the layout of that using Masked Face Detection and Face Mask Detection. The system can be used in any real-world setup where it is mandatory to ensure the COVID-19 related guidelines. Each component of the system has been tested and verified individually. The above proposed system can also be used at any public place to detect any notorious and suspicious person whose details are available with the authorities and generate an alert for the concerned authorities as well to catch that person. The algorithms for masked face detection and face mask detection proposed in this paper result in a better accuracy than existing algorithms.

Due to the lack of a dataset with all the required parameters, the system could not be implemented as a whole and for future works it is suggested to create a unified dataset which can be used for the above-mentioned task. A dataset of all the individuals with their masked images as well as unmasked images can be created and then the proposed models can be trained on the dataset to achieve the desired results.

References

[1] M. Ngam, P. Grother, and K. Hanaoka, "Ongoing Face Recognition Vendor Test (FRVT) Part 6A: Face Recognition Accuracy with Masks using pre-COVID-19 Algorithms," *NIST IR 8311*, National Institute of Standards and Technology, 2020.

[2] S. Ge, J. Li, Q. Ye and Z. Luo, "Detecting Masked Faces in the Wild with LLE-CNNs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 426-434, 2017.

[3] L. Best-Rowden and A. K. Jain, "Longitudinal study of automatic face recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(1), pp. 148-162, 2018.

[4] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, pp 1097-1105, 2012.

[6] E. Real, A. Aggarwal, Y. Huang, and Q. V. Le, "Aging evolution for image classifier architecture search," in *AAAI Conference on Artificial Intelligence*, 2019.

[7] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, "The megaface benchmark: 1 million faces for recognition at scale," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4873–4882, 2016.

[8] P. J. Phillips, "A cross benchmark assessment of a deep convolutional neural network for face recognition," in *Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pp. 705-710, 2017.

[9] Y. Cheng, J. Zhao, Z. Wang, Y. Xu, K. Jayashree, S. Shen, and J. Feng, "Know you at one glance: A compact vector representation for lowshot learning," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 1924–1932, 2017.

[10] L. Tran, X. Yin, and X. Liu, "Disentangled representation learning gan for pose-invariant face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1415–1424, 2017.

[11] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep Face Recognition: A Survey" in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp 471–478, 2018.

[12] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823, 2015.

[13] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1701–1708, 2014.

[14] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Advances in Neural Information Processing Systems*, pp. 1988–1996, 2014.

[15] Y. Sun, D. Liang, X. Wang, and X. Tang, "DeepID3: Face Recognition with Very Deep Neural Networks," arXiv:1502.00873, 2015.

[16] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," in *Proceedings of the British Machine Vision Conference (BMVC)*, pp. 41.1-41.12, 2015.

[17] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 212–220, 2017.

[18] M. A. Hmani and D. Petrovska-Delacrétaz, "State-of-the-art face recognition performance using publicly available software and datasets," in *Proceedings of the International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pp. 1-6, 2018.

[19] W. Hariri, "Efficient Masked Face Recognition Method during the COVID-19 Pandemic," Preprint (Version 3) available at Research Square, 2020.

[20] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning Face Representation from Scratch," arXiv:1411.7923, 2014.

[21] F. Wang, L. Chen, C. Li, S. Huang, Y. Chen, C. Qian, and C. Change Loy, "The Devil of Face Recognition is in the Noise," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

[22] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pp. 67–74, 2018.

[23] S. Yang, P. Luo, L. C. Change, and X. Tang, "Wider face: A face detection benchmark," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 5525–5533, 2016.

[24] A. Cabani, K. Hammoudi, H. Benhabiles, and M. Melkemi, "MaskedFace-Net - A Dataset of Correctly/Incorrectly Masked Face Images in the Context of COVID-19", arXiv:2008.08016, 2020.

[25] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei, H. Chen, Y. Miao, Z. Huang, and J. Liang, "Masked face recognition dataset and application," arXiv:2003.09093, 2020.

[26] Z. Zhao, P. Zheng, S. Xu, X. Wu, "Object Detection with Deep Learning: A Review," arXiv:1807.05511, 2019.

[27] L. Jiao, F. Zhang, F. Liu, S. Yang, L. Li, Z. Feng, R. Qu, "A Survey of Deep Learning-Based Object Detection," in *IEEE Access*, vol. 7, pp. 128837–128868, 2019.

[28] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," arXiv:1506.02640, 2016.

[29] Davis E. King, "Dlib-ml: A Machine Learning Toolkit," in *Journal of Machine Learning Research*, vol. 10, pp. 1755-1758, 2009.

[30] P. I. Wilson, and J. Fernandez, "Facial feature detection using Haar classifiers," in *Journal of Computing Sciences in Colleges*, vol. 21, no. 4, pp. 127–133, 2006.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:25:26 UTC from IEEE Xplore. Restrictions apply.

2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP - 2021)

Content Based Bot Detection using Bot Language Model and BERT Embeddings

Shubham Kumar*, Shivang Garg†, Yatharth Vats‡, Anil Singh Parihar§
Department of Computer Engineering
Delhi Technological University
New Delhi, India
Email: *shubhamk0027@gmail.com, †shivanggarg1998@gmail.com, ‡vatsyatharth46@gmail.com
, §parihar.anil@gmail.com

2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP) | 978-1-6654-3277-1/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCCSP52374.2021.9465506

***Abstract*—Micro-blogging sites such as Twitter, with 187 million active users daily, have become an essential source for getting news and trends, sharing information, and gathering various insights. Many user accounts on Twitter are automated with software to share content automatically without human intervention. Such accounts are called bots. However, the open nature of Twitter has resulted in people misusing it for changing the information dynamics, as seen in the US 2016 elections. This research proposes a neural network ensemble of Text CNN and LSTM model with BERT embeddings to classify tweets as bot tweets or not based on the tweets' textual content. Along with this, we handle the problem of data imbalance present in the Twitter dataset, using the Language Model Oversampling Technique to create a bot language model that will mimic bot tweets over the traditional techniques like Synthetic Minority Oversampling technique. A comparison and analysis of the tweets generated by the above two methods is made. The comparison with the various state-of-the-art model shows that the proposed model outperforms the previous models.**

***Index Terms*—Natural Language Processing, Information Systems, Attack Detection, Social Media**

I. Introduction

People use Micro-blogging sites like Twitter, Reddit, and Pinterest to share information and be up to date with the local, national and international news. Users of such platforms can be small to big businesses, students, traders, researchers, politicians, policymakers, government, actors, and even the general public. The primary use of these social media sites involves gathering business insights, product reviews, general public opinion, sentiment, and sharing information and developments.

Such social media channels are filled with automated social accounts called bots that share content without real human interaction. Considering the micro-blogging site Twitter, 15% of all the accounts are bots [1]. They play a valuable part in the ecosystem by posting real time updates, stories and answering user queries such as the Netflix Bot (@netflix_bot) that automatically tweets when a new content is added to the online streaming service and Grammar Checker (@_grammar_) that identifies grammatically incorrect tweets and offers suggestions for correct usage.

However, this evolution in social media and technology has also resulted in people misusing it. When the user accounts are controlled by the software algorithms, the accounts can then be used for spreading fake information, spam and, malicious content among the masses. With the spread of fake information from these accounts, people's opinions, views, and perceptions can be shaped in a particular direction. Such information can not only change the news trend dynamics but can also disturb the stock markets, change the general public opinion, promote or demote a product, individual, businesses or party changing the whole review and rating systems and the political discourse as in the 2016 US elections, where about one-fifth of the conversation on Twitter was produced by bots and, more than 5K bots were suspended by Twitter [2]. Another example is when an idle company raised its market value of 5B$ after a bot campaign, which was later suspended [3].

This is called "platform manipulation" mixed with the social engineering practices with the help of automated accounts. Researchers are using advancements in information sciences, machine learning, and natural language processing to detect such content over the internet. Lee et al. [4] were the first ones to notice the problem of rising bots in Twitter and MySpace.

Research shows that detecting social media bots is a challenge as their feature changes as botmasters evolve them [5]. In this research, we try to build a novel classifier for detecting bots in social media platforms like Twitter, Facebook, and Reddit based on the textual contexts such as tweets and posts posted by the user accounts without considering other parameters such as their likes, retweets, friend circles, usernames etc. while handling the data imbalance present in the S. Cresci [6] Twitter dataset to achieve a near to perfect accuracy. The contributions of this research are as follows:

- To detect social accounts as bot or not purely based on the social media account's contents. We propose that textual features can be compelling for detecting bots with great accuracy.
- To employ Language Model oversampling technique [7] in NLP to handle data imbalance in the publicly available Cresci dataset over the traditional methods like SMOTE [8]. This will help in creating a bot model that will generate the bot text for the minority classes.
- Propose a novel user content-based Twitter bot detection model using BERT embeddings with an ensemble of neural nets.
- Compare the results of the proposed model on the balanced dataset with state-of-the-art methods.

978-1-6654-3277-1/21/$31.00 ©2021 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:39:28 UTC from IEEE Xplore. Restrictions apply.

The rest of the paper is organized as follows- section II presents the literature review of the related research works. Section III describes the motivation behind the research. Section IV describes the proposed model's methodology with results and analysis in Section V. Finally, Section VI concludes the work.

II. Literature Review

The detection methods can be classified as account-level detection or textual level detection. Account-level detection methods are complex and expensive and require a lot of labeled data. Each account is characterized by all the user's textual contents, his activities such as likes, retweet counts, social circle, and account metadata. Text-level approaches are simple yet powerful for detecting bots. Over the state-of-the-art account level classifiers, textual content detectors have seen many improvements for detecting bots with high accuracy.

Lee et al. [4] were the first to notice the problem of rising bots in Twitter and the MySpace community. They called them honeypots and employed a simple supervised technique called Decorate for detecting social spammers and honeypots based on the content, account-usage and tweet account. Loyla et al. [9] used contrast-pattern-based supervised classifier for Twitter bot detection and introduced a new feature model out of Twitter content usage and tweet content sentimental analysis.

Sneha et al. [10] employed deep neural networks for detecting bots using contextual Long Short Term Memory exploiting based on the content, and the metadata of the user account handled using GloVe Embeddings. They proposed a new technique for handling the data imbalance problem in the dataset based on the Synthetic Minority Oversampling Technique by combining batches of existing ENN and TOMEK generated samples. SMOTENN samples were generated by oversampling the minority and by undersampling with ENN. And the same technique was used with TOMEK. However, these techniques only generated the content, not the metadata of the accounts.

Jorge et al. [1] proposed a one-class classification approach for detecting bots on Twitter. Their supervised method is based on the idea that the "bot and human features can be separated from the labeled dataset effectively" and were indeed able to achieve high accuracy. They proposed it as an anomaly detection problem in one class classifier and used Bagging-TPMiner and Bagging-RandomMiner for detecting the anomaly.

More recently, Bin et al. [11] handled the data imbalance problem using an Improved Conditional Generative Adversarial Network (improved CGAN). They used GANs to generate new samples for the minority class and applied training classifiers to improve the accuracy of detecting the social bots. To avoid imbalance within the social bot distributions, they proposed a modified clustering algorithm called Gaussian kernel density peak clustering algorithm (GKDPCA). They also extended the proposed work using traditional Wasserstein distance with a gradient penalty to avoid gradient disappearance.

A Shared task for detecting bots was published at CLEF 2019 as CLEF 2019 Bots Profiling Subtask (English language). Most of the methods employed there were statistically based on the non-textual parameters of the social accounts. Michael et al. [12] were the first to employ Convolution Neural Networks for such a task over the traditional methods of Random forests, Bayes Net, and Support Vector Machines. They used superficial 2 one dimensional CNN layers and sigmoid activation final layer with just text contents of the accounts. They were able to achieve high accuracy of 90.34% utilizing this technique.

Feng et al. [13] used deep neural networks based on a contextual BiLSTM for detecting bots with Glove Embeddings. They employed three stacked BiLSTM layers with around 200 recurrent units. They achieved an accuracy of 92% but didn't handle the data imbalance present in the Cresci dataset. Such neural networks have millions of features, but not all are important. Attention mechanisms are used to concentrate on a few relevant things, ignoring the others.

III. Motivation

The research community has proposed different approaches for the detection of bots. Though they are able to achieve good results with high accuracy, their performance can not be considered accurate until the highly imbalanced nature of the datasets is taken into account. The research works that handled the imbalance still lack the reasoning behind their work. Overall, the primary motivations that fueled this research are as follows:

1) One of the significant challenges in detecting bots is handling the data imbalance in the datasets. In real datasets like the Twitter Dataset [6], the number of bot users is significantly less as compared to the number of non-bot users. Most of the previous research in this domain used Synthetic Minority Oversampling Technique Chawla et al. [8] for handling the problem of data imbalance. Recently Bin et al. [11] handled this problem using GANs.
   Techniques such as SMOTE use KNN for generating new samples, which works well for less feature space. Still, for generic structured data like the text, the feature space is enormous, and so the samples generated loose their semantic structure.
   Sneha et al. [10] handled the data imbalance using SMOTE but by adding undersampling by ENN and TOMEK but could not get such improvements because of the underlying SMOTE technique.
2) Michael et al. [12] were the first ones to employ Convolution Neural Networks as their classifier, and Feng et al. [13] used deep neural networks based on a contextual bidirectional long short-term memory (Bi-LSTM) for detecting bots. Using neural networks, they were able to achieve high accuracy of 92% Feng et al. [13]. Sneha et al. [10] used contextual LSTM for the same. Such neural networks have millions of features, but not all are important. Attention mechanisms are employed to

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:39:28 UTC from IEEE Xplore. Restrictions apply.

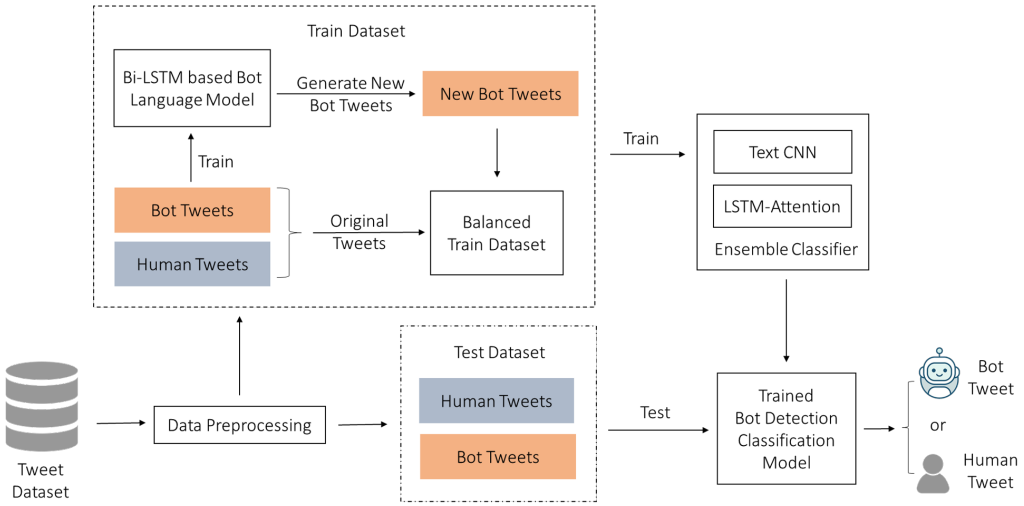


Fig. 1. Twitter Bot Detection Framework

TABLE I
DATASET DESCRIPTION

Database	Description	Accounts	Tweets	Year
Genuine Accounts	Manually Verified Accounts	3474	8.4M	2011
Social Spambots #1	Italian political candidate Retweeters	991	1.6M	2012
Social Spambots #2	Paid applications spammers	3457	0.4M	2014
Social Spambots #3	Amazon Sale Product Spammers	464	1.4M	2011
Traditional spambots #1	Malware Spammers	1000	1.4M	2009

concentrate on a few relevant things while ignoring the others.

3) Using Bi-LSTM networks Feng et al. [13] captured the input context but did not focus on the n-gram spatial correlation among the words. Michael et al. [12] succeeded in capturing the n-gram relationships using their CNN, but the accuracy achieved is still far from actual due to the imbalance present in the dataset.

IV. METHODOLOGY

For this research, we use the publicly available Twitter dataset by Cresci et al. [6]. It presents a labeled dataset after manually monitoring the interactions, shared content, friends, and profile pictures of various user accounts. Most of the tweets are manually selected for being produced by a bot or not. The dataset involves overall 8.4 million user tweets and 4.8 million bot tweets. That is, around 60% of the total tweets are human tweets and remaining as bot tweets. So, the dataset is imbalanced by 3.3 million tweets. A short overview of the dataset used for training is given in Table 1. The bot tweets are generally the spammers collected from different sources - from retweeters of Italian political candidates, application spammers, product spammers, and malware spammers.

Fig. 1 summarizes the various steps involved in the process. The problem of the data imbalance in the dataset is addressed by training a new bot model using the Language Model Oversampling Technique [7] to mimic the bot language and generate new samples. With this technique, we will generate synthetic samples using a trained model for the minority classes. It will be trained on the tweets from the text classified as bots in the dataset. A simple BiLSTM network will be used to generate new sequences from the most frequent n-grams present in the bot text.

An ensemble architecture is then trained with BERT Large Uncased word embeddings to beat the state-of-the-art classifiers over the newly generated dataset on the balanced dataset generated. BERT [14] or Bidirectional Encoder Representations from Transformers is a pre-trained transformer-based Language Model trained on a large corpus of unlabeled text. The BERT-large version has a vocabulary size of 32K and an embedding length of 1024. It consists of 24 hidden encoder layers stacked on top of each other, each with 16 attention heads, totaling 336M parameters. The proposed framework for detection can be divided into three phases:

1) Phase I: Preprocessing Training Data: The tweet dataset has to be cleaned before feeding it directly to the training pipeline. The major steps involved in cleaning the dataset are as follows:

- Remove Accented Characters like é.
- Expand Short words like ”ain’t” to ”is not”.
- Removal of tags, links, images, and duplicate words.
- Remove stop words like ‘an’ and ‘the’. And

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:39:28 UTC from IEEE Xplore. Restrictions apply.

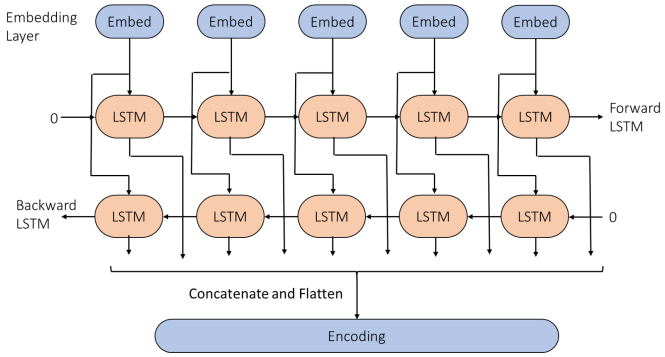


Fig. 2. The Bi-LSTM network used for handling the data imbalance

- Tokenization and stemming.

2) Phase II: Handling Data Imbalance(LMOTE): Once the data is pre-processed, we can build the bot language model and handle the data imbalance. The pseudocode for the same is described in Algorithm 1. *N_GRAMS* is selected to be five, and the max length of the tweets is taken as 70. A simple BiLSTM neural network is used for the *next_word_predictor* as shown in Fig. 2. Using the most frequent n-grams as a seed, the next word is repeatedly predicted from previous n-grams generated using *next_word_predictor*. This is done for every new sentence to be generated from a frequent n-gram.

Algorithm 1: Handling data imbalance

Input: Dataset of minority class(bot tweets) $train_x$, Size of N-GRAM N_GRAMS
Output: New generated dataset new_x

```
ngram ← []
next_word ← []
foreach sentence S_i ∈ train_x do
    N ← length of S_i - N_GRAMS
    for i ← 0 to N do
        Insert S_i[i...i + N_GRAMS] into ngram
        Insert S_i[i + N_GRAMS] into next_word
    end for
next_word_predictor ← BiLSTM model trained using ngram and next_word
popular_ngram ← most frequent n-grams from whole data set
new_x ← generate new sentences using next_word_predictor and popular_ngram
return new_x
```

3) Phase III: Training the Ensemble Classifier: The classification architecture will be an ensemble of Text CNN and LSTM with attention mechanism. Word vectors generated from the BERT embeddings are passed through the LSTM model and parallelly to CNN model, as shown in Fig.3. The CNN model consists of two Convolutions of kernel size 3 and 4. The results from the two models are finally concatenated for the classification output.

V. Results And Analysis

TABLE II
EXAMPLE OF SENTENCES GENERATED FROM DIFFERENT OVERSAMPLING TECHNIQUES

Technique	Example
SMOTE	west visited second rose system children than matches neighbourhood mother by within common it records where she
LMOTE	good morning all feed your mind something nutritious the brain muscle exercise daily

Table 2. shows an example of the bot tweets generated when SMOTE (Synthetic Minority Oversampling model) is used for handling the data imbalance and using LMOTE. SMOTE uses a KNN based model for developing new sentences using the older ones. But this model does not work well for the textual data. So, the sentences generated using the traditional SMOTE are less meaningful as compared to LMOTE used.

The results of the comparison with the state-of-the-art methods is shown in Table 3. The ensemble classifier proposed is able to separate the bot tweets with high accuracy, beating all the state-of-the-art classifiers proposed so far. It also performed equivalent to the conditional GANs used by Bin et al. [11] for handling the data imbalance. This is due to the ensemble classifier which captured the word relationships effectively and the attention mechanism of the transformer model that focused more on the important parts of the text.

BERT word embeddings effectively handles the out-of-vocabulary words and help create intelligent sentence embeddings. Simple Convolutional Neural Network can produce effective results in a wide range of machine learning problems. The convolutional layers are generally applied to convolve a vast representation, such as an image [15] to small feature maps. Here, the CNN model on the textual data captures the relationships among the inputs, i.e the n-grams and the LSTM (Long Short Term Memory) based attention model words to emphasize on the essential features in the sequential input.

TABLE III
CLASSIFICATION PERFORMANCE OF VARIOUS MODELS USED FOR BOT DETECTION

Classifier	Accuracy	Precision	Recall	F1-Score
Decorate Classifier	0.89	0.88	0.89	0.89
Contextual LSTM	0.96	0.96	0.96	0.96
Improved CGAN	**0.97**	**0.97**	**0.97**	**0.97**
Contextual Bi-LSTM	0.93	0.93	0.93	0.93
BERT with CNN & LSTM-Attention (Proposed Method)	**0.97**	**0.97**	**0.97**	**0.97**

VI. Conclusion

This research proposed an ensemble classifier for detecting bots after effectively handling the data imbalance present.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:39:28 UTC from IEEE Xplore. Restrictions apply.

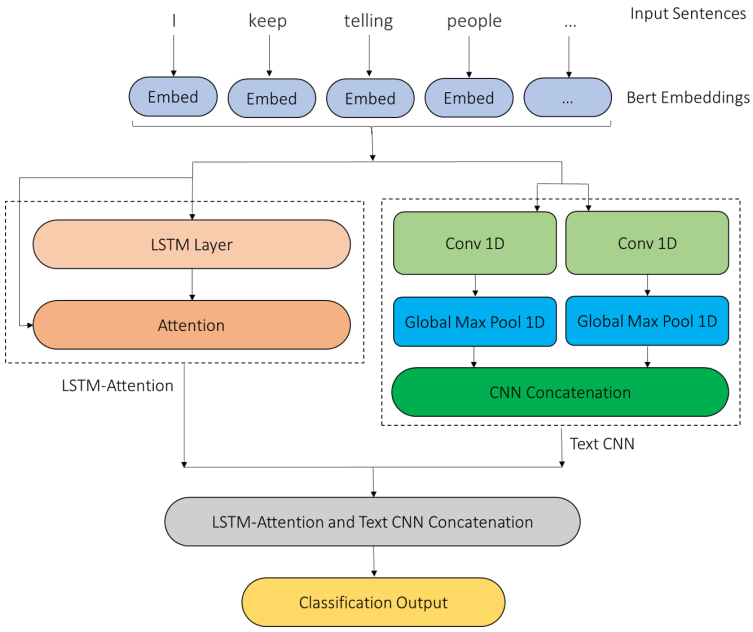


Fig. 3. Architecture of the Ensemble classifier used for bot detection.

Over the publicly available Cresci dataset, which contains 15 million tweets in total, the proposed model achieved a comparative performance to the state-of-the-art models after handling the data imbalance. Techniques such as SMOTE use KNN for generating new samples, which works well for less feature space. Still, for generic structured data like the text, the feature space is enormous and the samples generated are not semantically correct. The Bot model overcomes this problem and creates new tweets without losing the text semantics.

There are also some disadvantages of our proposed model. The foremost is that the sentences generated for handling the data imbalance are not new, but just slight modifications of the sentences already present in the dataset. The whole sentence is generated from the same limited dataset, and the frequent n-grams supplied. Also, handling the data imbalance leads to early overfit during the training process. Because of the same reason and so the use of dropouts is quite common.

REFERENCES

[1] J. Rodríguez-Ruiz, J. I. Mata-Sánchez, R. Monroy, O. Loyola-González, and A. López-Cuevas, "A one-class classification approach for bot detection on twitter," *Computers Security*, vol. 91, p. 101715, 2020.

[2] A. Bessi and E. Ferrara, "Social bots distort the 2016 us presidential election online discussion," *First Monday*, vol. 21, November 2016.

[3] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," *Commun. ACM*, vol. 59, p. 96–104, June 2016.

[4] K. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: Social honeypots + machine learning," in *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, (New York, NY, USA), p. 435–442, Association for Computing Machinery, 2010.

[5] M. S. Orabi, D. Mouheb, Z. Aghbari, and I. Kamel, "Detection of bots in social media: A systematic review," *Inf. Process. Manag.*, vol. 57, p. 102250, 2020.

[6] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race," in *Proceedings of the 26th International Conference on World Wide Web Companion*, WWW '17 Companion, (Republic and Canton of Geneva, CHE), p. 963–972, International World Wide Web Conferences Steering Committee, 2017.

[7] M. Jain, M. Leekha, and M. Goswami, "A multi-task approach to open domain suggestion mining (student abstract)," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 13817–13818, Apr. 2020.

[8] N. Chawla, K. Bowyer, L. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.

[9] O. Loyola-González, R. Monroy, J. Rodríguez, A. López-Cuevas, and J. I. Mata-Sánchez, "Contrast pattern-based classification for bot detection on twitter," *IEEE Access*, vol. 7, pp. 45800–45817, 2019.

[10] S. Kudugunta and E. Ferrara, "Deep neural networks for bot detection," *Information Sciences*, vol. 467, p. 312–322, Oct 2018.

[11] B. Wu, L. Liu, Y. Yang, K. Zheng, and X. Wang, "Using improved conditional generative adversarial networks to detect social bots on twitter," *IEEE Access*, vol. 8, pp. 36664–36680, 2020.

[12] M. Färber, A. Qurdina, and L. Ahmedi, "Identifying twitter bots using a convolutional neural network," in *Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum (CLEF'19)*, CEUR, 2019.

[13] F. Wei and U. T. Nguyen, "Twitter bot detection using bidirectional long short-term memory neural networks and word embeddings," in *2019 First IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA)*, 2019.

[14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, (Minneapolis, Minnesota), pp. 4171–4186, Association for Computational Linguistics, June 2019.

[15] A. Parihar, "A study on brain tumor segmentation using convolution neural network," *2017 International Conference on Inventive Computing and Informatics (ICICI)*, pp. 198–201, 2017.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:39:28 UTC from IEEE Xplore. Restrictions apply.

2021 2nd International Conference for Emerging Technology (INCET)
Belgaum, India. May 21-23, 2021

COVID 19, Pneumonia and Other Disease Classification using Chest X-Ray images

Abhishek Panwar
Department of Computer Science & Engineering
Delhi Technological University Shahbad
Daulatpur, Main Bawana Road, DL - 42, IN
abhishekpanwar971@gmail.com

Akshay Dagar
Department of Computer Science & Engineering
Delhi Technological University Shahbad
Daulatpur, Main Bawana Road, DL - 42, IN
akshaydagar98@gmail.com

Vishal Pal
Department of Computer Science & Engineering
Delhi Technological University Shahbad
Daulatpur, Main Bawana Road, DL - 42, IN
vishalpal1999@gmail.com

Dr. Vinod Kumar
Department of Computer Science & Engineering
Delhi Technological University Shahbad
Daulatpur, Main Bawana Road, DL - 42, IN
vinod_k@dtu.ac.in

2021 2nd International Conference for Emerging Technology (INCET) | 978-1-7281-7029-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/INCET51464.2021.9456192

***Abstract*— In this report, we propose an alternative way of detecting the Covid-19 disease using Convolutional Neural Networks based deep learning models which will classify between Covid-19 and other similar respiratory diseases such as other types of Viral Pneumonia or Bacterial Pneumonia, using images of Chest X-Rays. The models developed as part of this project achieved approximately 98% accuracy on a test dataset of 300 images.**

Keywords—Covid-19, Pneumonia, Disease Classification

I. INTRODUCTION

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome Corona virus 2 (SARS-CoV-2). It was first identified in December 2019 in Wuhan, Hubei, China, and has resulted in an ongoing pandemic. As per WHO on 21 March 2021, more than 123 million cases have been reported across 223 countries and territories with more than 2.7 million deaths: more than 69.5 million people have recovered Common symptoms include fever, cough, fatigue, shortness of breath or breathing difficulties, and loss of smell and taste. While most people have mild symptoms, some people develop acute respiratory distress syndrome (ARDS) possibly precipitated by cytokine storm, multi-organ failure, septic shock, and blood clots. The incubation period may range from one to fourteen days. This disease generally spreads through air during close proximity of an infected person, contaminated surfaces etc. Currently the standard method of diagnosis is by real-time reverse transcription polymerase chain reaction (rRT-PCR) from a nasopharyngeal swab. The Chest X-ray imaging may also be helpful for diagnosis in individuals where there is a high suspicion of infection based on symptoms and risk factors, however guidelines do not recommend using it for routine screening currently. In our study here, we trained our models on openly available chest X-ray dataset maintained by physicians and sources like Kaggle. To avoid any sort of overfitting due to shortage of data, we used various renowned image augmentation techniques. Our models had shown high accuracy on Testing dataset. However , we need more data and extensive testing to put it in production. In coming section, we will discuss about our literature overview, our methodology and results.

II. LITERATURE OVERVIEW

Mucahid Barstugan et. al. [5] study presents early phase detection of Coronavirus (COVID-19), which is named by the World Health Organization (WHO), by machine learning methods. The detection process was implemented on abdominal Computed Tomography (CT) images. The expert radiologists detected from CT images that COVID-19 shows different behaviours from other viral pneumonia. Therefore, the clinical experts specify that COVİD-19 virus needs to be diagnosed in early phase. For detection of the COVID-19, four different datasets were formed by taking patches sized as 16x16, 32x32, 48x48, 64x64 from 150 CT images. The feature extraction process was applied to patches to increase the classification performance. Grey Level Co-occurrence Matrix (GLCM), Local Directional Pattern (LDP), Grey Level Run Length Matrix (GLRLM), Grey-Level Size Zone Matrix (GLSZM), and Discrete Wavelet Transform (DWT) algorithms were used as feature extraction methods. Support Vector Machines (SVM) classified the extracted features. 2-fold, 5-fold and 10-fold cross-validations were implemented during the classification process. Sensitivity, specificity, accuracy, precision, and F-score metrics were used to evaluate the classification performance.

In Enzo Tartaglione et. al. [8] insights are provided and also warnings are raised on what is reasonable to expect by applying deep learning to COVID classification of CXR images. A methodological guide and critical reading of an extensive set of statistical results that can be obtained using currently available datasets is provided. In particular,the challenge posed by current small size COVID data and show how significant can be the bias introduced by transfer-learning using larger public non-COVID CXR datasets is taken. Contribution is also made by providing results on a medium size COVID CXR dataset, just collected by one of the major emergency hospitals in Northern Italy during the

© IEEE 2021. This article is free to access and download, along with rights for full text and data mining, re-use and analysis.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:30:10 UTC from IEEE Xplore. Restrictions apply.

peak of the COVID pandemic. These novel data allow for the contribution to validate the generalization capacity of preliminary results circulating in the scientific community. The conclusions shed some light into the possibility to effectively discriminate COVID using CXR.

However, in Gianluca Maguolo et. al. paper [9], different testing protocols used for automatic COVID-19 diagnosis from X-Ray images in the recent literature are compared and evaluated. It is shown that similar results can be obtained using X-Ray images that do not contain most of the lungs. It is possible to remove the lungs from the images by turning to black the center of the X-Ray scan and training their classifiers only on the outer part of the images. Hence, it is deduced that several testing protocols for the recognition are not fair and that the neural networks are learning patterns in the dataset that are not correlated to the presence of COVID-19. Finally, it is shown that creating a fair testing protocol is a challenging task, and they provide a method to measure how fair a specific testing protocol is. In the future research it is suggested to check the fairness of a testing protocol using their tools and they encourage researchers to look for better techniques than the ones that they propose.

This paper tries to incorporate better strategies in order to overcome aforementioned shortcomings. However, lack of ample sized dataset acts as a bottleneck in obtaining more accurate results.

III. METHODOLOGY

In this section, we have divided our work in four parts. Firstly, we will discuss about the collection of data, data sources and its relevance. Then, we will move to data processing and preparation part where we discuss about how we prepared the data for training. Then, we will see the different model architectures that we used for training the parameters and finally we will see the training and testing of the data for various parameters.

The workflow of our work throughout this process, can be summarised as shown in the figure down below.

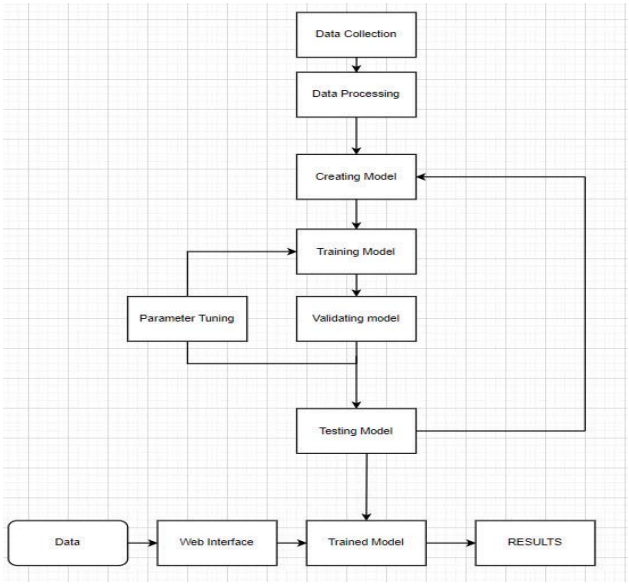


Fig. 1. Workflow Diagram

A. Data Collection

Data is a very important input for any machine learning or deep learning problem. A relevant collection of data helps in analysing a machine learning problem better. In our problem set, we have collected data majorly from two different and prominent sources present online. Firstly, we have data from Covid 19 chest X-ray dataset repository on Github maintained by Joseph Paul Cohen, Paul Morrison, Lan Dao, Karsten Roth, Tim Q Duong and Marzyeh Ghassemi. The data of this repository is collected from various hospitals and physicians. This data was collected from public sources like hospitals and maintained under the project approved by University of Montreal's Ethics Committee. We have chosen the 'AP' and 'PA' view images of chest X-ray for our identification. We further classified it in majorly three classes as covid affected, non-covid pneumonia affected and normal/ other disease affected. Further, we collected the data of Pneumonia and other diseases from Kaggle competition. Apart from these sources, we supplemented our data set from various online sources and studies done before. Finally, after combining these datasets, we were able to gather a total of 4563 images (478 Covid 19 or Covid 19 affected pneumonia images, 2647 other pneumonia images and the rest 1438 were the normal/ other disease images).

B. Data Preparation and Processing

From this combined dataset, we will create a testing dataset containing 100 images belonging to each of the 3 classes (Covid-19 Pneumonia images, Other Pneumonia Images, and Other Diseases/No disease images) and validation dataset containing 78 images belonging to each of the three classes. This will leave us with 300 Covid-19 images, 2469 other Pneumonia images and 1260 images belonging to other diseases/no diseases class

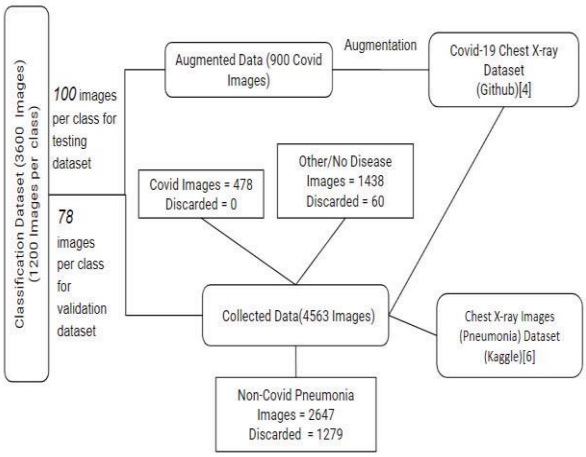


Fig. 2. Dataset Preparation

The leftover training dataset is highly imbalanced and if we use this dataset to train our models, there are high chances of overfitting as the model will be able to obtain a high accuracy by predicting the majority class(es). To balance the collected dataset, we used Image Augmentation techniques. The different augmentation techniques we used are as follows:

- Random Rotations
- Random Shifting (left, right, up, or down)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:30:10 UTC from IEEE Xplore. Restrictions apply.

- Random Gaussian Noise Adding
- Horizontal Flipping (mirror image)
- Random Cropping
- Random Shearing
- Random Scaling

For each Covid-19 Pneumonia Image, we will randomly select 3 of these 10 augmentation functions and apply them to the image to generate 3 augmented copies of the image and save the copies with the original image in our dataset.

After Augmenting, we have 1200 Covid-19 Pneumonia CXR images. However, the dataset is still imbalanced, and we need to reduce the no. of images of the other 2 classes to 1200 each to form a perfectly balanced dataset. Since the other 2 classes contain multiple images from a few persons, we will keep only 1 image per person thus reducing the no. of images belonging to these 2 classes. After that, we will keep only 1200 images from each class to form our final perfectly balanced training dataset.

IV. Model Architectures and classifiers

In this section, we discuss different model architectures we used. In our work, we mainly use two deep learning models which are 4 Layer CNN (Model 1) and AlexNet [10] (Model 2) Model 1 was further classified on the basis of our dataset. We used Model 1 with the original dataset initially collected once and then we used it with a modified dataset which was supplemented with our augmented dataset.

- Model 1 (4 Layer CNN)
- Model 2 (AlexNet)

The two models were trained with the gathered data as follows:

- Initially, we trained Model 1 with the original dataset. The training was done using Image Data Generators which will read images from the dataset folders and resize them to size (224,224) with 3 channels (RGB) and rescale the images to [0,1]. Data generators were also used for validation images. After training, we only want the model weights which give the best validation accuracy to be saved as our final model weights, hence we use the "Model Checkpoint" callback from the Keras library. We saved the model parameters that result in the best accuracy. Finally, we trained this model for 50 epochs with a default batch size of 32 train images per batch.
- Thereafter, We trained Model 2 with the original dataset. This model is based on AlexNet. We will take the Convolution layers of the AlexNet model and discard the Output layers and Dense (hidden) layers and instead attach our own hidden layers and output layer with Softmax activation to the output of the convolution layers. The complete architecture has already been defined in the previous section on "Model Architectures". The training of this model will be done similar to that of model1. The only exception being that we will fine tune the AlexNet weights while training instead of using randomly initialized weights as done for model 1.
- Finally, we used the same Model 1 however we trained it differently using Image Data generators with augmentation. Basically, we will randomly take images from the training dataset folder and apply random augmentations on them, thus generating augmented versions of those images, on the fly, while training. Note that we will not use the original image, only the augmented ones, while training. Since we have already augmented our original dataset, this will be kind of a double augmentation. Augmenting like this can help us reduce overfitting our model since the dataset contains augmented copies of the original 300 images and hence there will be a lot of similarities. The rest of the training was similar to other training above.

V. Experiment Results and Discussion

In this section, we will see the experiment results collection from testing.

We tested the trained models on the test dataset containing 100 samples for each of the 3 classes, the results are as follows:

TABLE I. No. of images of different classes used for training and testing dataset

Class Name	*No. of Images*	
	Training	***Testing***
Covid 19	1200	100
Non –Covid Pneumonia	1200	100
Normal/Other diseases	1200	100

(Note: 0 = Covid-19 Pneumonia, 1 = Other Diseases/No Disease, 2 = Other Pneumonia)

The results of the three models are as follows:

Finally, we will take the weighted average of the predictions of the 3 models to get final predictions. This weighted prediction function given below is the final classifier we used in our work. The weighted average is taken in the following manner:

Count_Class0 = (No. of Models with majority class prediction = 0)

Count_Class1 = (No. of Models with majority class prediction = 1)

Count_Class2 = (No. of Models with majority class prediction = 2)

Final Prediction = ((P1*Count_ClassP1) + (P2*Count_ClassP2) + (P3*Count_ClassP3)) / (Count_ClassP1 + Count_ClassP2 + Count_ClassP3)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:30:10 UTC from IEEE Xplore. Restrictions apply.

TABLE II. CLASSIFICATION ACCURACY, PRECISION AND RECALL (IN PERCENTAGES) FOR DIFFERENT DEEP LEARNING MODELS IN COVID-19, PNEUMONIA AND OTHER DISEASES/NORMAL CLASSIFICATION ON TRAINING AND TESTING DATASET

Model Name	***Closs***	**Testing/Training**		
		Accuracy	***Precision***	***Recall***
Model 1	Covid	98.67	98.00	98.00
	Pneumonia	97.00	95.00	95.96
	Normal/other Disease	97.67	97.00	96.04
Model 2	Covid	98.00	98.00	96.07
	Pneumonia	97.67	94.00	98.94
	Normal/Other Disease	98.33	99.00	96.11
Model 3	Covid	98.00	95.00	98.95
	Pneumonia	95.00	88.00	96.70
	Normal/Other Disease	95.67	100.00	88.49
Final Result	Covid	99.00	99.00	98.02
	Pneumonia	98.33	95.00	100.00
	Normal / Other Disease	98.67	100.00	96.15

Thus the overall accuracy for Model 1 is 96.67 %, for Model 2 is 97.00% , for Model 3 is 94.34%.The final results show that our classifier is performing well on a test dataset of 300 images and achieving 98% accuracy with 100% accuracy for “Other Diseases/No disease” class and 99% accuracy for “Covid-19 Pneumonia” class.

Due to lack of data for Covid-19, we cannot test it on a larger dataset, however testing the classifier on a dataset of 1000 images each of classes 1 and 2 and 100 images of class 0 (Covid-19) produced almost similar results with slightly reduced accuracy (97%) which is bound to happen on larger datasets).

VI. CONCLUSION AND FUTURE WORK

AlexNet and 4 layer Convolutional Neural Network have been proved to be really worthy and provided us the commendable results in predicting the Covid 19 cases using only the X ray images. Using them, we are able to achieve almost 98% accuracy which is comparatively higher than many other parallel research work going in the field. But since the data available is quite less right now, we cannot completely rely on these models and their results for actual implementation in the Medical industry. So, to conclude we can say that a more exhaustive availability of the required data is needed.

As future work, we plan to test our classifier on more data as it becomes available so as to produce more reliable results as per the medical industry acceptance standards. Till then, we are also looking forward to training some parallel CNN models on the available data in near future and testing parallel CNN models for the same in order to achieve better results.

REFERENCES

[1] R. M. Pereira, D. Bertolini, D. Bertolini, C. N. S. Jr. and Y. M. Costa, "COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios," *Computer Methods and Programs in Biomedicine,* vol. 194, 2020.

[2] A. I. F. AI, AI2, CZI, MSR, Georgetown, NIH and T. W. House, "COVID-19 Open Research Dataset Challenge (CORD-19)," 2020. [Online]. Available: https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge. [Accessed 2020].

[3] J. P. Cohen, P. Morrison and L. Dao, "COVID-19 Image Data Collection," *arXiv,* p. 4, 25 March 2020. K. Elissa, “Title of paper if known,” unpublished.

[4] J. P. Cohen, "covid-chestxray-dataset," 2020.

[5] S. ozturk, U. Ozkaya and M. Barstugan, "Classification of Coronavirus (COVID-19) from X-rayand CT images using shrunken features," *International Journal of Imaging Systems and Technology,* p. 11, 2020.

[6] Kaggle, "COVID-19 chest xray," 2020.

[7] Wikipedia, "Coronavirus disease 2019," 2019. [Online]. Available: https://en.wikipedia.org/wiki/Coronavirus_disease_2019. [Accessed 2020].

[8] E. Tartaglione, C. A. Barbano, C. Berzovini, M. Calandri, and M. Grangetto, “Unveiling COVID-19 from CHEST X-Ray with Deep Learning: A Hurdles Race with Small Data,” *International Journal of Environmental Research and Public Health*, vol. 17, no. 18, p. 6933, Sep. 2020.

[9] G. Maguolo and L. Nanni, "A Critic Evaluation of Methods for COVID-19 Automatic Detection from X-Ray Images," *ArXiv,* vol. abs/2004.12823, 2020.

[10] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," 25th International Conference on Neural Information Processing Systems, vol. 1, pp. 1097-1105, 2012.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:30:10 UTC from IEEE Xplore. Restrictions apply.

2021 2nd International Conference for Emerging Technology (INCET)
Belgaum, India. May 21-23, 2021

COVID-19 Severity Assessment from Chest X-rays using Attention-based Weakly-Supervised Learning

Kunal Chaturvedi
Department of Applied Mathematics
Delhi Technological University
New Delhi, India
kunalchaturvedi98@gmail.com

Tanmay Kansal
Department of Applied Mathematics
Delhi Technological University
New Delhi, India
tanmaykansal702@gmail.com

Saksham Gupta
Department of Applied Mathematics
Delhi Technological University
New Delhi, India
sakgupta55@gmail.com

Dinesh Kumar Vishwakarma
Department of Information Technology
Delhi Technological University
New Delhi, India
dvishwakarma@gmail.com

Naokant Deo
Department of Applied Mathematics
Delhi Technological University
New Delhi, India
naokantdeo@dce.ac.in

2021 2nd International Conference for Emerging Technology (INCET) | 978-1-7281-7029-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/INCET51464.2021.9456449

***Abstract*— The potential of automated severity assessment of COVID-19 pneumonia is immense due to its ability to facilitate clinical decision-making. It enables efficient escalation or de-escalation of COVID-19 care. In this work, we propose an efficient pipeline based on weakly-supervised learning for severity score prediction. In the first stage, Attention feature fusion (AFF)-ResNet-101 is trained on five large Chest X-ray (CXR) datasets. In the second stage, using transfer learning on 94 posteroanterior (PA) COVID-19 CXR images, refined feature maps are extracted. These feature descriptors are then mapped to severity ratings based on Geographic Extent score and lung opacity output using the linear regression model. Experimental results show the efficacy of our proposed architecture for the severity score prediction.**

Keywords*— *COVID-19, severity assessment, weakly-supervised learning, attention mechanism

I. INTRODUCTION

The novel coronavirus SARS-CoV-2, is a fatal respiratory illness that is highly contagious and can be transmitted through the respiratory droplets or by contact with a contaminated surface [1][2]. Cough, fever, loss of smell, shortness of breath, body aches, lung inflammation [3] are some of the common symptoms of this disease. Inflammation in the lungs, can also cause acute respiratory distress syndrome [4]. Compared to SARS, MERS, and other similar respiratory diseases, the COVID-19 spread rate is enormous [5][6] as it can also be spread by asymptomatic COVID-19 carriers. The pandemic has slowed down the world economy and deeply affected the healthcare infrastructure in all the countries. Early detection of this disease can limit the spread of the virus and help in saving lives by facilitating timely treatment.

COVID-19 disease diagnosis requires detection of SARS-CoV-2 through laboratory tests using CT scans, Chest X-rays (CXRs), RT-PCR testing, etc. RT-PCR test is less favourable due to its limited sensitivity, longer detection time, and inconsistent results. Also, due to the limited availability of testing kits and labs, it is quite difficult for every patient with respiratory symptoms to be tested by RT-PCR. Even though in the initial stages of diseases, CXRs are not very sensitive to accurate detection of pulmonary involvement but they can be very helpful in monitoring an increase in lung abnormalities over a period of time [7][8]. CXRs are preferred over CT-scans because the former is more efficient in terms of time and cost involved than the latter. CXR images of COVID-19 patients show specific characteristics which could help in diagnosing and monitoring such patients. Hence, it is widely used in hospitals and medical facilities. Given that the number of patients infected by this disease is far more than the number of radiologists present, there is a need to use an artificial intelligence-based algorithm to aid radiologists in analysing CXR images quickly and predicting COVID-19 severity within a short span of time.

The paper proposes a model which predicts COVID-19 pneumonia severity score. The model involves several steps - data pre-processing, Attention based feature extraction, and severity score prediction using Linear Regression. The first step, data pre-processing involves resizing, normalisation, and NSST [9]. The second step involves use of Attentional Feature Fusion [10] to ResNet-101. In most of the previous work [11][12], features were fused via short skip connection in ResNet which led to very deep network training. Attentional Feature Fusion (AFF) generalises the attention-based feature fusion and helps in providing steady direction, thereby enhancing ResNet performance. The addition operator in the ResNet block is replaced by the AFF module. The third step involves the use of Linear Regression to predict COVID-19 pneumonia severity score using a different set of features. This model can be used to monitor the COVID-19 progress and the response to treatment given to the patients.

The paper is organized as follows. Previous related work has been discussed in Section II below. The materials and the methodology are described in detail in Section III. Section IV provides the experimental results. Finally, in Section V the main conclusions are outlined.

II. RELATED WORKS

In this section, we briefly address the previously proposed approaches that diagnose COVID-19 using medical imaging. These techniques use computed tomography (CT) scans and chest X-ray (CXR) for the initial diagnosis of COVID-19. However, due to better financial viability and portability, CXRs are preferred over CT-scans for the initial imaging modality. For this reason, many healthcare facilities use CXRs as a first diagnostic imaging option for COVID-19 disease monitoring and patient management. Most of the recent works on COVID-19 diagnosis are focused on classification [13], and segmentation [14][15]. [13] developed COVIDX-Net for the classification purpose.

© IEEE 2021. This article is free to access and download, along with rights for full text and data mining, re-use and analysis.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:28:36 UTC from IEEE Xplore. Restrictions apply.

Moreover, deep learning-based models [15] are used to assess the impact of lung segmentation on COVID-19 detection.

However, a more informative visual assessment of CXRs is necessary for COVID-19 diagnosis. Severity assessment of CXR is used as an assistive tool for efficient patient management. It provides the extent of COVID-19 pulmonary abnormalities in the CXRs. The approach of severity analysis can be used for health care management, i.e., escalation or de-escalation of care [16]. Such assessment is done using semiquantitative scoring systems [17][18] that quantitatively estimate the pulmonary involvement to numerical scales. [17]uses scoring systems for CT-scans to quantify the involvement. Yoon et al. [18] performed CXR analysis using a 5-point scale-based chest radiograph scoring system [19] to grade abnormalities. So far, limited work has been done with the use of deep learning approaches in this direction. Recently, Signoroni et al. [20] developed an end-to-end deep learning architecture to predict Brixia score [7]. To obtain the Brixia score, the lung region is subdivided into six zones, and based on the extent of lung abnormalities, a score (range, 0-3) is assigned to each region.

Cohen et al. [21] proposed two COVID-19 severity scores for the assessment: geographic extent score and lung opacity score. Based on the ground glass opacity, a score (range, 0-4; 0 = no lung involvement, 1 $\leq$ 25%, 2 = 25-50%, 3 = 50-75%, 4 $\geq$ 75% lung involvement) is assigned and the sum of score of each lung is calculated which is termed as geometric extent score (range, 0-8). Similarly, lung opacity score (range, 0-6) is the sum of assigned score for each lung depending on the degree of opacity (0 = no opacity; 1 = ground-glass opacity; 2 = consolidation; 3 = white-out). Due to a limited annotated dataset for severity assessment, a CNN model was pre-trained on seven non-COVID-19 CXR datasets consisting of a total of 18 classes to be used as a feature extractor on a small COVID-19 dataset. However, the use of non-COVID-19 datasets and extraneous classes for training significantly affected the performance of the deep learning model.

This work uses Attention Feature Fusion (AFF) based ResNet-101 that is trained using weakly supervised learning to assess the degree of opacity and the extent of lung involvement on posteroanterior (PA) COVID-19 CXR images.

III. Materials and Methods

In this section, we study in detail, the datasets used and the proposed methodology for the severity assessment.

A. Datasets

A COVID-19 image data collection as shown in Fig. 1, was proposed by Cohen et al. [22] consisting of 94 posteroanterior (PA) CXR images (average age; 56±14.8). The images were sourced within the duration of December 2019 to March 2020 and manually annotated by radiologists for Geographic Extent (GE) score and Lung opacity (LO) score. However, due to the limited availability of annotated dataset for this task, we use the pre-training step. The pre-training step avoids overfitting of deep learning model on a very small dataset. For this purpose, we used five publicly available chest X-ray datasets including the Brixia COVID-19 dataset [20](see Table 1). These five datasets were aligned to each other and four classes (COVID-19 pneumonia; Lung Opacity; Consolidation; Infiltration) from the datasets were used for the pre-training step.

TABLE I. Different Public Datasets Used in the Experimentation

Dataset Used	Classes Used
Brixia COVID-19 [20]	COVID-19 pneumonia
CheXpert - Stanford University [23]	Lung Opacity, Consolidation
MIMIC-CXR – MIT [24]	Lung Opacity, Consolidation
OpenI [25]	Lung Opacity
ChestX-ray8 - National Institutes of Health (NIH) [26]	Infiltration

B. Severity Assessment

We introduced a multi-stage framework that assesses the severity of the infection in the CXR region of COVID-19 patients. The framework is designed to learn and perform different tasks such as: data pre-processing, feature extraction, and the severity score prediction: geography extent score and lung opacity score. The proposed methodology is discussed in detail below.

a. Data Pre-processing

Most of the research conducted on the CXR use entire raw data for the training purpose. However, there is a major drawback of using untreated data: the existence of irrelevant objects including such as symbols, bright texts, varying spatial resolutions in the CXR images. This adds to the unnecessary noise that limits the performance of the deep learning networks. Below are some data pre-processing techniques to alleviate some of the aforementioned issues.

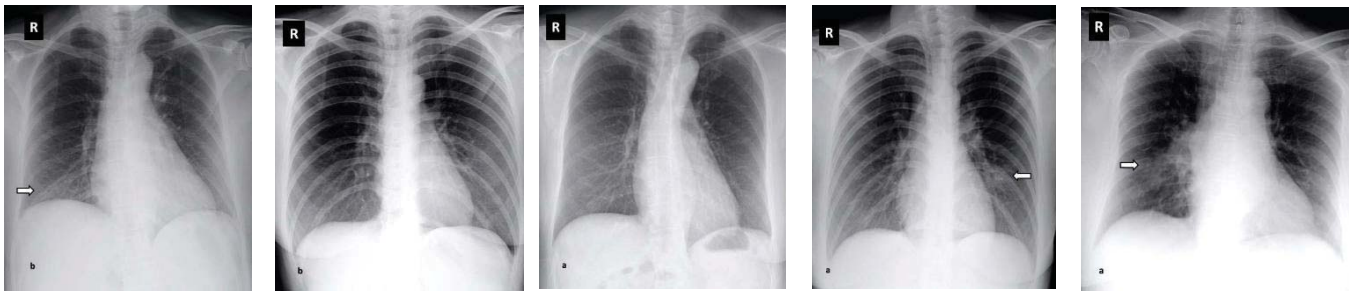


Fig. 1. The illustration of some example chest X-rays from the posteroanterior (PA) COVID-19 CXR dataset [32]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:28:36 UTC from IEEE Xplore. Restrictions apply.

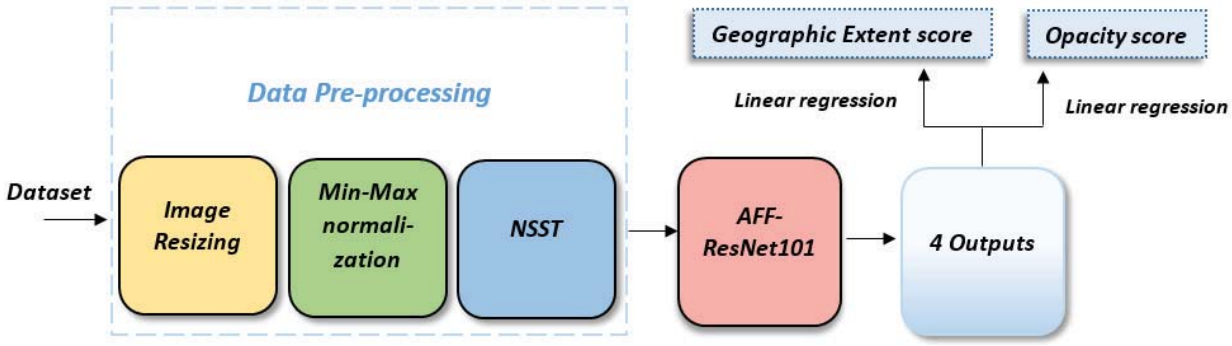


Fig. 2. Block diagram of methodology for severity assessment

- **Image resizing** and **min-max normalization** is applied on X-ray images to ensure uniformity across different datasets.
- **Non-subsample shearlet transform** (NSST) [9] consists of Laplacian pyramid filters and shift-invariant filters that are the extension of the wavelet transform. It decomposes the noisy X-ray image into multi-scale sub-band and multi-directional images. Further, these images are de-noised with the help of filters and combined to get an output X-ray image using an inverse shearlet transform.

b. Attention-based Feature Extraction

Here, we apply Attentional Feature Fusion (AFF) [10], a multi-scale attention-based mechanism to ResNet-101 [12]. The feature fusion operator i.e., the addition operator in ResNet block (ResBlock) is replaced with the attention mechanism. The module addresses the issue of contextual aggregation and enhances contextual awareness by feature-wise modulation and refinement. As it is challenging to train a CNN model on small datasets, we adopt two-stage training to address the issue: Initially, ResNet-101 with Attentional Feature Fusion (AFF) module is trained on the six datasets as mentioned in Table 1. Further we perform supervised fine-tuning on the smaller dataset using the initially trained AFF-ResNet101.

IV. Experiments

After applying the aforementioned pre-processing steps on the five chest X-ray datasets, weakly-supervised pretraining was performed using Attentional Feature Fusion (AFF) based ResNet-101 and further, transfer learning is applied to 94 posteroanterior (PA) CXR images of COVID-19 patients. In the experimentation, five-fold cross-validation is adopted to train the deep learning model with RMSprop optimizer and the initial learning rate is chosen at 0.0001. The implementation of the proposed technique is done using the Pytorch framework on two Nvidia RTX 2080Ti GPU. As per evaluation strategies used in the evaluation of previous literature work, linear regression is trained on the different sets of features to predict the lung involvement and the degree of opacity. For the purpose, the training sample of 94 posteroanterior (PA) CXR images is divided into training set (50%) and test set (50%). Further, performance metrics including R Square (R^2), Pearson correlation, and mean squared error (MSE) are adopted for the model evaluation. Block diagram representation of the methodology is shown in Fig. 2. The comparison between model's prediction and the ground truth for geographic extent and opacity (as illustrated in Fig. 3) shows that the most of the points lie near

TABLE II. Comparison of the different variants of our weakly supervised approach with other methodologies on Posteroanterior (PA) CXR images

	Task	Opacity Score			Geographical Extent		
Method		"lung opacity" output	4 outputs	Intermediate network features	"lung opacity" output	4 outputs	Intermediate network features
Cohen et al. [21]	Pearson Correlation	0.78±0.04	0.78±0.04	0.66±0.08	0.80±0.05	0.79±0.05	0.74±0.08
	R square	0.58±0.09	0.58±0.09	0.25±0.21	0.60±0.09	0.57±0.10	0.43±0.16
	MSE	0.86±0.11	0.87±0.12	1.54±0.28	2.06±0.34	2.17±0.37	2.88±0.58
ResNet-101 + Regression (Ours)	Pearson Correlation	0.79±0.04	0.79±0.04	0.67±0.09	0.81±0.06	0.80±0.05	0.75±0.09
	R square	0.61±0.09	0.58±0.09	0.26±0.21	0.63±0.09	0.57±0.11	0.46±0.15
	MSE	0.88±0.12	0.86±0.11	1.56±0.27	2.09±0.37	2.19±0.36	2.90±0.57
AFF-ResNet-101 + Regression (Ours)	Pearson Correlation	0.85±0.05	0.86±0.05	0.71±0.10	0.86±0.05	0.86±0.05	0.71±0.10
	R square	0.66±0.11	0.64±0.10	0.32±0.23	0.66±0.10	0.62±0.10	0.50±0.17
	MSE	0.79±0.09	0.81±0.12	1.47±0.29	1.98±0.32	2.06±0.35	2.69±0.59

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:28:36 UTC from IEEE Xplore. Restrictions apply.

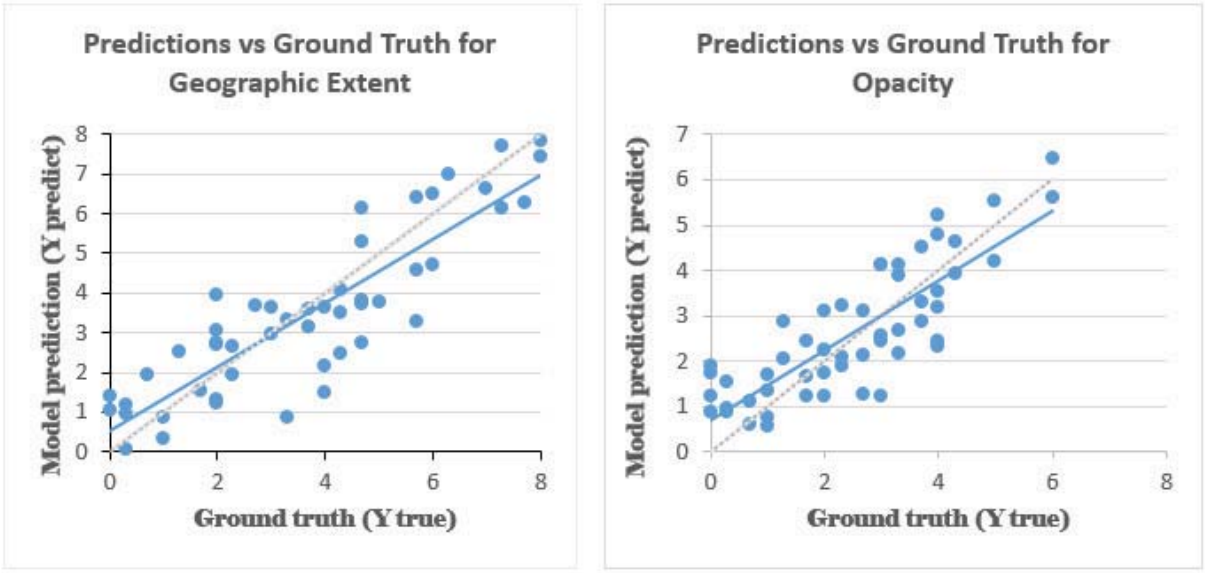


Fig. 3. Two scattered plots representing the score prediction made by model vs the ground truth (manual annotation by radiologist) for geographic extent and opacity respectively. In this graph, blue line represents linear regression line while the dotted grey line represents the exact prediction.

the grey line representing perfect prediction.

As observed in Table 2, AFF-ResNet-101 is able to achieve improved performance over the ResNet-101 based weakly-supervised approach and Cohen et al. [1] that offers no such attention mechanism. It is attributed to the fact that Attention Feature Fusion (AFF) module replaces the conventional addition operation. It enhances the feature fusion in skip connections of ResNet-101 and boosts the model's ability to identify regions of utmost importance. Moreover, Cohen et al. [21] trained DenseNet-101 on a large set of non-COVID-19 datasets that were aligned for 18 classes. The use of non-COVID-19 datasets and irrelevant classes create unnecessary confusion in the feature extraction process. However, our approach enables the initial and target dataset to be similar and relevant for the first round of training i.e., the training on five datasets. It provides an unbiased approach for an efficient feature extractor that further, allows better learning for the linear regression model.

V. Conclusions

The present paper proposes a pipeline that uses Attention-based weakly supervised learning for automatic severity assessment based on the degree of opacity and geographical extent. To the best of our knowledge, this is the first study that uses an attention mechanism to gauge the level of severity of infection in COVID-19 positive CXR. Moreover, our framework was able to outperform the existing state-of-the-art method on the posteroanterior (PA) COVID-19 CXR dataset. In the future, our aim would be to extend the scope of the proposed system to the other pulmonary diseases and investigate further improvements for better efficiency and robustness.

References

[1] The Lancet Respiratory Medicine, "COVID-19 transmission—up in the air," Lancet Respir. Med., vol. 8, no. 12, p. 1159, Dec. 2020, doi: 10.1016/S2213-2600(20)30514-2.

[2] K. Chaturvedi, D. K. Vishwakarma, and N. Singh, "COVID-19 and its impact on education, social life and mental health of students: A survey," Child. Youth Serv. Rev., vol. 121, p. 105866, Feb. 2021, doi: 10.1016/j.childyouth.2020.105866.

[3] [C. Huang et al., "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," Lancet, vol. 395, no. 10223, pp. 497–506, Feb. 2020, doi: 10.1016/S0140-6736(20)30183-5.

[4] N. Zhu et al., "A Novel Coronavirus from Patients with Pneumonia in China, 2019," N. Engl. J. Med., vol. 382, no. 8, pp. 727–733, Feb. 2020, doi: 10.1056/NEJMoa2001017.

[5] E. Mahase, "Coronavirus: covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate," BMJ, p. m641, Feb. 2020, doi: 10.1136/bmj.m641.

[6] D. Wang et al., "Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China," JAMA, vol. 323, no. 11, p. 1061, Mar. 2020, doi: 10.1001/jama.2020.1585.

[7] A. Borghesi and R. Maroldi, "COVID-19 outbreak in Italy: experimental chest X-ray scoring system for quantifying and monitoring disease progression," Radiol. Med., vol. 125, no. 5, pp. 509–513, May 2020, doi: 10.1007/s11547-020-01200-3.

[8] R. Yasin and W. Gouda, "Chest X-ray findings monitoring COVID-19 disease course and severity," Egypt. J. Radiol. Nucl. Med., vol. 51, no. 1, p. 193, Dec. 2020, doi: 10.1186/s43055-020-00296-x.

[9] A. Zhang, H. Jiang, L. Ma, Y. Liu, and X. Yang, "A Shearlet-based algorithm for quantum noise removal in low-dose CT images," Mar. 2016, p. 97843O, doi: 10.1117/12.2216562.

[10] Y. Dai, F. Gieseke, S. Oehmcke, Y. Wu, and K. Barnard, "Attentional Feature Fusion," Sep. 2020, [Online]. Available: http://arxiv.org/abs/2009.14082.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Identity Mappings in Deep Residual Networks," 2016, pp. 630–645.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[13] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images," Mar. 2020, [Online]. Available: http://arxiv.org/abs/2003.11055.

[14] A. Degerli et al., "COVID-19 Infection Map Generation and Detection from Chest X-Ray Images," Sep. 2020, [Online]. Available: http://arxiv.org/abs/2009.12698.

[15] L. O. Teixeira et al., "Impact of lung segmentation on the diagnosis and explanation of COVID-19 in chest X-ray images," Sep. 2020, [Online]. Available: http://arxiv.org/abs/2009.09780.

[16] J. M. Courtney and A. Bax, "Hydrating the Respiratory Tract: An Alternative Explanation Why Masks Lower Severity of COVID-19 Disease.," medRxiv Prepr. Serv. Heal. Sci., Dec. 2020, doi: 10.1101/2020.12.23.20248671.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:28:36 UTC from IEEE Xplore. Restrictions apply.

[17] F. Pan et al., “Time Course of Lung Changes at Chest CT during Recovery from Coronavirus Disease 2019 (COVID-19),” Radiology, vol. 295, no. 3, pp. 715–721, Jun. 2020, doi: 10.1148/radiol.2020200370.

[18] J. Zhu, B. Shen, A. Abbasi, M. Hoshmand-Kochi, H. Li, and T. Q. Duong, “Deep transfer learning artificial intelligence accurately stages COVID-19 lung disease severity on portable chest radiographs,” PLoS One, vol. 15, no. 7, p. e0236621, Jul. 2020, doi: 10.1371/journal.pone.0236621.

[19] E. Taylor et al., “A chest radiograph scoring system in patients with severe acute respiratory infection: a validation study,” BMC Med. Imaging, vol. 15, no. 1, p. 61, Dec. 2015, doi: 10.1186/s12880-015-0103-y.

[20] A. Signoroni et al., “End-to-end learning for semiquantitative rating of COVID-19 severity on Chest X-rays,” Jun. 2020, [Online]. Available: http://arxiv.org/abs/2006.04603.

[21] J. P. Cohen et al., “Predicting COVID-19 Pneumonia Severity on Chest X-ray With Deep Learning.,” Cureus, vol. 12, no. 7, p. e9448, Jul. 2020, doi: 10.7759/cureus.9448.

[22] J. P. Cohen, P. Morrison, and L. Dao, “COVID-19 Image Data Collection,” Mar. 2020, [Online]. Available: http://arxiv.org/abs/2003.11597.

[23] J. Irvin et al., “CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison,” Jan. 2019, [Online]. Available: http://arxiv.org/abs/1901.07031.

[24] A. E. W. Johnson et al., “MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs,” Jan. 2019, [Online]. Available: http://arxiv.org/abs/1901.07042.

[25] D. Demner-Fushman et al., “Preparing a collection of radiology examinations for distribution and retrieval.,” J. Am. Med. Inform. Assoc., vol. 23, no. 2, pp. 304–10, Mar. 2016, doi: 10.1093/jamia/ocv080.

[26] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul. 2017, pp. 3462–3471, doi: 10.1109/CVPR.2017.369.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:28:36 UTC from IEEE Xplore. Restrictions apply.

Check for updates

Design and Analysis of Nonlinear Controller for a Standalone Photovoltaic System Using Lyapunov Stability Theory

Narendra Kumar
Department of Electrical Engineering,
Delhi Technological University,
Delhi 110042, Delhi, India
e-mail: narendrakumar@dtu.ac.in

Aman Sharma
Department of Electrical Engineering,
Delhi Technological University,
Delhi 110042, Delhi, India
e-mail: aman_mt2k18@dtu.ac.in

This paper presents the Lyapunov stability theory-based nonlinear controller design for a standalone photovoltaic (PV) system. The comparative analysis of different nonlinear controllers is also carried out. Due to the nonlinear I-V characteristics of photovoltaic systems, conventional hill-climbing methods like Perturbate and Observe and Incremental Conductance do not show reliable tracking of maximum power under various uncertainties. Therefore, these methods require nonlinear tools to meet control objectives and design specifications. Out of various nonlinear controlling techniques, the one presented in this paper is the sliding mode approach for maximum power point tracking (MPPT). In the context of Lyapunov stability theory, the sliding mode approach uses a switching manifold. In this approach, the system trajectories are made to follow the sliding surface and remain there forever to ensure the stability of equilibrium points. Two types of sliding mode controllers have been simulated, namely, conventional—sliding mode controller (CSMC) and terminal—sliding mode controller (TSMC). The results are analyzed and compared scientifically on various performance parameters including duty cycle ratio, ideal and PV output power, and time taken for error convergence, under varying dynamic conditions. All the control algorithms are developed in MATLAB/SIMULINK. [DOI: 10.1115/1.4051584]

Keywords: efficiency, measurement, photovoltaics, MPPT, renewable energy, simulation, nonlinear control

1 Introduction

Per capita, energy consumption is one of the key indexes to assess the progress of a nation. Since the primary source for energy generation is mainly dependent on fossil fuels, which are responsible for the emission of harmful greenhouse gases. Renewable energy is one of the present solutions to climatic change issues which provide energy for our regular necessities with almost zero-carbon emissions. Despite the Indian government's commitment to the solar sector, the amount of solar energy generation in India remains only 7.9% of the total (as of March 2019) [1]. The rooftop solar segment being the sluggish sector with numerous problems comprising only 15% of cumulative solar installation in India, reaching 4375 MW. Given the above, rooftop or standalone photovoltaic (PV) systems' efficiency can be improved by using improved and robust controllers.

Maximum power point tracking algorithms are essential in PV system applications because a solar panel's maximum power point varies continuously with the irradiation, temperature, and shading conditions. Maximum power point tracking (MPPT) techniques try to maintain the PV array's operating point at its maximum power point (MPP) under varying climatic conditions. Maximum power transfer between the panels and the load is achieved by varying the duty ratio of the converter, matching the intrinsic impedance of panels with the load impedance.

Many research papers have already been published in the field of MPPT algorithms, such as conventional hill-climbing methods that include Perturb and Observe Technique and Incremental conductance, also MPPT algorithms based on ANFIS and Fuzzy Logic theory have already been published. In literature [2–5] P&O MPP tracking algorithm for standalone PV modules has been studied. In Refs. [6–9], incremental conductance MPP tracking algorithm for photovoltaic modules shows various advantages over the P&O algorithm in terms of response and tracking efficiency. However, in works of literature, it was observed in plots of PV output voltage that whenever there is a change in climatic conditions the tracking of MPP is not achieved properly. In continuance of the literature work, Refs. [10–14] describe a comparative analysis between the ANN-based MPP tracking algorithm and hill-climbing methods.

There exists a logarithmic relationship between the output voltage and current of the PV array. Analysis of PV systems requires solving Nonlinear differential equations, and frequency domain transformation does not apply. So tedious efforts have to be put in to develop appropriate tools for finding the solution of systems equations. Lyapunov stability based nonlinear controller can be used to meet the control objectives and design specifications. Using Lyapunov stability theory for nonlinear systems, stability analysis can be achieved without actually solving the nonlinear differential equations. Out of various nonlinear tools in literature [15–23], the sliding mode approach for MPPT is studied. Further, in these literatures, the robustness of control algorithms and tracking efficiency with other conventional hill-climbing methods are carried out.

This paper finally makes a comparative analysis between the two sliding mode controllers TSMC and CSMC and their performance parameters are further discussed in Secs. 3 and 4 of the paper.

1.1 Photovoltaic Effect and Solar Cell Model.

A solar cell represented by an electrical equivalent one-diode model is presented in Fig. 1

$$I_{ph} = I_{PV} - \mathrm{Io}\left\{\exp\left[\frac{e(V_{PV} + I_{PV} * R_s)}{\eta KT}\right] - 1\right\} - \frac{V_{PV} + I_{PV} * R_s}{R_{sh}} \tag{1}$$

Contributed by the Solar Energy Division of ASME for publication in the JOURNAL OF SOLAR ENERGY ENGINEERING: INCLUDING WIND ENERGY AND BUILDING ENERGY CONSERVATION. MANUSCRIPT received August 7, 2020; final manuscript received June 21, 2021; published online July 16, 2021. Assoc. Editor: Brendan O'Connor.

 Copyright © 2021 by ASME

Downloaded from http://asmedigitalcollection.asme.org/solarenergyengineering/article-pdf/144/1/011003/6730015/sol_144_1_011003.pdf by Delhi Technological University user on 28 July 2021

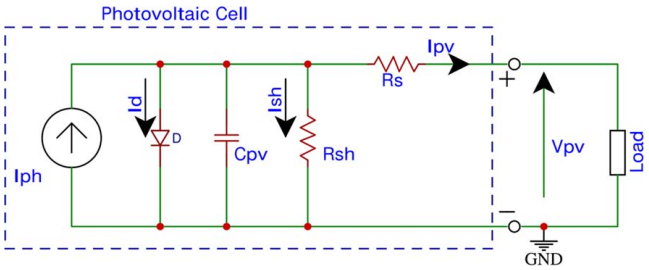


Fig. 1 Equivalent circuit of solar cell

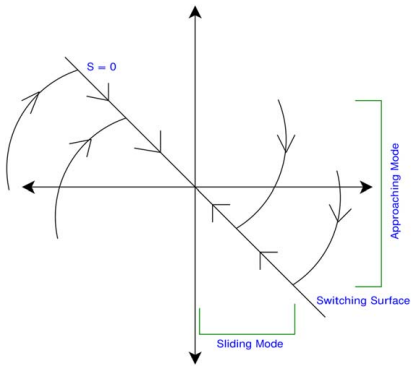


Fig. 2 Modes of operation of sliding mode control

For a typical solar cell, the value of internal shunt resistance (R_{sh}) is very large as compared with series resistance (R_s).

Neglecting the current through R_{sh}, the net current equation, which is the difference between the photocurrent (I_{ph}) and the normal diode current (I_d), can be modified as

$$I_{PV} = I_{ph} - I_D = I_{ph} - \text{Io}\left\{\exp\left[\frac{e(V_{PV} + I_{PV} * R_s)}{\eta K_b T}\right] - 1\right\} \quad (2)$$

The PV panel is not like an ordinary classical source. The output voltage is coupled and is affected by the output current expressed by a nonlinear I-V relationship. The output voltage is coupled with the current by a logarithmic relation given by

$$V_{PV} = V_{\text{Thermal}} * \ln(I_{ph} - I/I_{\text{saturation}}) \quad (3)$$

where V_{Thermal} and $I_{\text{saturation}}$ are the thermal equivalent voltage and dark saturation current, respectively.

Table 1 Boost converter and PV module parameters

Parameters	Value	Parameters	Value
C_{pv}	2000 μF	Maximum power (MP)	305.226 W
L	10 mH	Cells per module (n_s)	96
C_o	2000 μF	V_{OC}	64.2 V
R	70 Ω	I_{SC}	5.96 A
F_{sw} (switching freq.)	50 KHz	V_{mp}	54.7 V
V_d (diode vol. drop)	0.82 V	I_{mp}	5.58 A
A (diode ideality factor)	0.94497	K_i	0.061/°C

2 Lyapunov Stability Theory

According to theory, the equilibrium points are said to be stable if all the system solutions beginning at neighborhood points stay nearby, otherwise, they are unstable. Those points are asymptotically stable if all the solutions beginning at neighborhood points stay nearby and be likely to move toward the equilibrium point as the time approaches infinity. It uses the idea for constructing a scalar energy-like function also known as Lyapunov function of the system state and then choosing an appropriate control law to make this function decrease, i.e., its first derivative should be negative

$$\dot{V} \leq 0 \quad (4)$$

2.1 Sliding Mode Methodology for MPPT. Sliding mode control [24] is a nonlinear control tool where trajectories of the system are forced to follow or reach a sliding manifold, also known as switching surface in a finite duration of time. These trajectories are made to stay on switching manifolds for all future time, as shown in Fig. 2. It is a variable structure control method that can switch from one continuous structure to another depending on the present position in the state space.

It has two modes of operation:

(1) Approaching mode—The system states converges to a predefined switching manifold in finite time duration.
(2) Sliding mode—System states are confined on the sliding surface and are forced toward the desired state, which is the MPP on PV characteristics.

2.2 System Design and Modeling. A Boost type converter is connected to the PV Module with a resistive load as shown in Fig. 3. Table 1 shows the specifications of the converter and PV module used.

Mathematical modeling of the system can be carried out using the state-space averaging technique. The system can be expressed in two sets of state-space equations depending upon the position of the switch S for the Boost converter as shown in Fig. 4.

Based on the concept of pulse-width modulation (PWM), the ratio of switch position at point "a" to the total time period is defined as duty ratio.

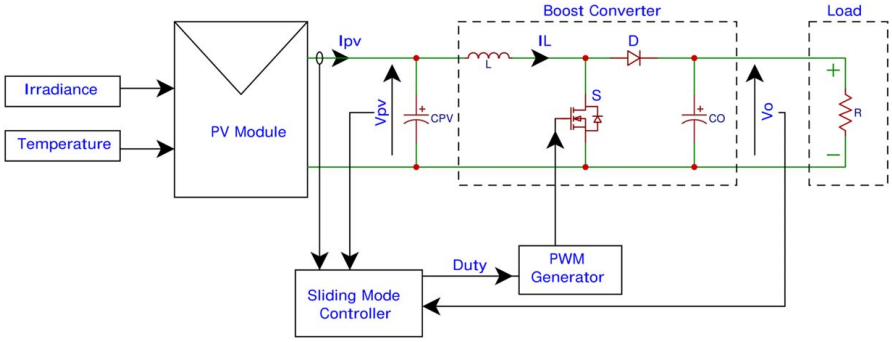


Fig. 3 PV system with boost converter schematic

Downloaded from http://asmedigitalcollection.asme.org/solarenergyengineering/article-pdf/144/1/011003/6730015/sol_144_1_011003.pdf by Delhi Technological University user on 28 July 2021

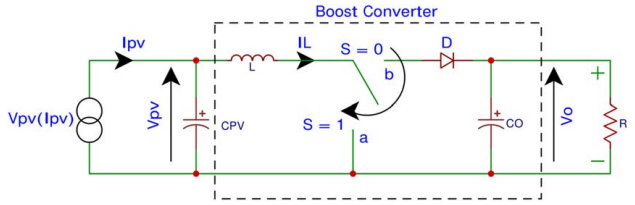


Fig. 4 Boost converter operation

Considering "d" as the duty ratio for the control signal provided to switching MOSFET (S), the dynamics of the system can be rewritten as follows:

$$\frac{dI_L}{dt} = \frac{V_{PV}}{L} - (1-d)\frac{V_o}{L} \tag{5}$$

$$\frac{dV_o}{dt} = \frac{I_L}{C_o}(1-d) - \frac{V_o}{C_o R} \tag{6}$$

where $0 < d < 1$.

3 Conventional—Sliding Mode Control

At first, it involves selecting a switching surface that will guarantee the system trajectories to hit the surface and produces maximum output power repeatedly. Selecting the sliding surface corresponding to the point where Maximum power occurs. So, the switching manifold can be mathematically expressed as

$$\frac{\partial P_{pv}}{\partial I_{pv}} = \frac{\partial (I_{pv}^2 * R_{pv})}{\partial I_{pv}} = I_{pv}\left(2R_{pv} + I_{pv}\frac{\partial R_{pv}}{\partial I_{pv}}\right) = 0 \tag{7}$$

where the term $R_{pv} = V_{pv}/I_{pv}$ refers to the equivalent load connected to the Photovoltaic system. So, the sliding surface is deduced as

$$\sigma = 2R_{pv} + I_L\frac{\partial R_{pv}}{\partial I_L} \tag{8}$$

have assumed that $I_{pv} \cong I_L$.

The updation in duty ratio provided to the control the operation as depicted in Fig. 5 can be expressed as

$$d_{\mathbf{update}} = \begin{cases} d - \Delta d, & \sigma < 0,\ d \text{ decrease} \\ d + \Delta d, & \sigma > 0,\ d \text{ increase} \end{cases} \tag{9}$$

The equivalent control can be determined from the following condition:

$$\sigma^{\bullet} = \frac{d\sigma}{dt} = \frac{d\sigma}{dI_{pv}}\frac{dI_{pv}}{dt} = 0 \tag{10}$$

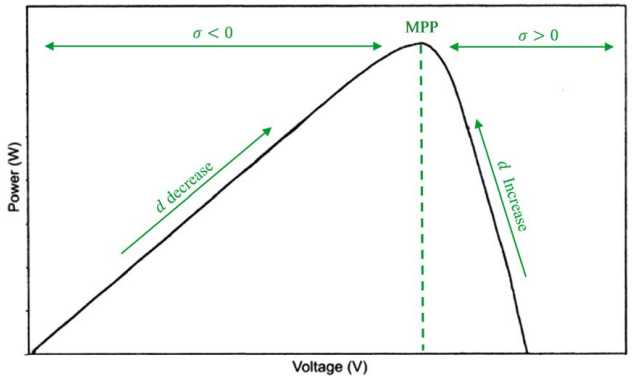


Fig. 5 Updation in duty cycle provided to control the operation as expressed from Eq. (9) by assuming $I_{pv} \cong I_L$

The trivial solution for the above condition can be obtained by

$$\left[\frac{V_{pv}}{L} - (1 - d_{\mathbf{equ}})\frac{V_o}{L}\right] = 0 \quad d_{\mathbf{equ}} = 1 - \frac{V_{pv}}{V_o} \tag{11}$$

The real control signal can be given as

$$d = \begin{cases} 1, & d_{\mathbf{equ}} + K * \text{sign}(\sigma) \geq 1 \\ d_{\mathbf{equ}} + K * \text{sign}(\sigma), & 0 < d_{\mathbf{equ}} + K * \text{sign}(\sigma) < 1 \\ 0, & d_{\mathbf{equ}} + K * \text{sign}(\sigma) \leq 0 \end{cases}$$

where K is a positive constant.

Evaluating system performance under the following varying irradiance conditions shown in Fig. 6. Starting with a minimum irradiance of 200 W/m^2, applied to the PV system and later as the simulation progresses the irradiance value is changed for analyzing the behavior of the system's operating point. The MATLAB design model for the system is shown in Fig. 7.

It is evident from the plot shown in Fig. 8 that the duty ratio changes with the irradiance, suggesting that updation in the duty ratio is achieved by the MPP tracking algorithm. However, it can be observed from the plot that duty ratio profile is not flat/linear and suffers from oscillations due to the chattering phenomena.

3.1 Chattering Phenomenon. It is a phenomenon that occurs when the system operating trajectories heading toward the switching manifold tries to slide on the surface. When the trajectories strike the surface at point "a," ideally it should glide on the surface but due to delay in switching devices it crosses the surface and enters the region $s < 0$ shown in Fig. 9. When the control switches, the trajectories reverse their direction and moves toward the manifold. The repetition of this process creates the "zig-zag" motion known as chattering.

Chattering affects the following parameters of the system:

- Duty ratio profile
- Mismatch in ideal and PV outpower (MPPT tracking efficiency)

The plot for the ideal power and PV output power under the varying irradiance is shown in Fig. 10.

It is evident from the subplot that the approaching mode of sliding mode control is achieved in a finite duration. However, it suffers from overshoots and undershoots in reaching the switching surface. Further analysis for the approaching mode is shown is Fig. 11 and Table 2. The primary function of MPPT algorithm is to extract the maximum output power from PV module under all irradiance conditions.

Figure 12 shows that due to the chattering, the converter's output power suffers from oscillations, leading to more losses in switching circuitry, which may lead to unwanted high-frequency dynamics.

From Fig. 13, it can be observed that during the initial irradiance conditions of 200 W/m^2 the converter efficiency is indeterministic because of the chattering phenomenon.

From Fig. 14, it can be observed that the condition $\sigma\dot{\sigma} < 0$ is satisfied under all regions of the irradiance curve hence proves for the existence of CSMC based controller based on Lyapunov's stability theory.

4 Terminal—Sliding Mode Control

The terminal sliding mode takes into consideration both the voltage and current tracking errors and a sliding surface is designed to converge these errors to zero in finite time duration. The convergence of these errors to 0 will guarantee the existence of controller action. TSMC more robustly drives the system trajectories toward the switching manifold described by $\sigma(t)$

$$e_1 = V_{\text{PV}} - V_{\text{PV}}^* \tag{12}$$

Downloaded from http://asmedigitalcollection.asme.org/solarenergyengineering/article-pdf/144/1/011003/6730015/sol_144_1_011003.pdf by Delhi Technological University user on 28 July 2021

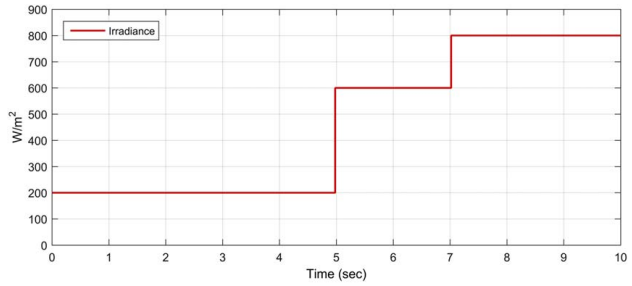


Fig. 6 Irradiance curve generated using the signal builder

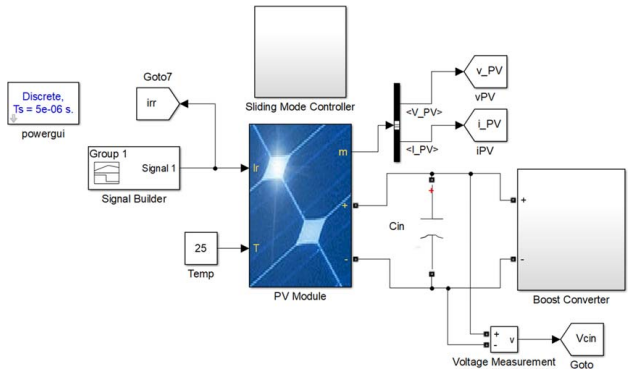


Fig. 7 Complete design model for the system on MATLAB. Temperature conditions are fixed at $T = 25$ °C. PV module (SunPower SPR-305-WHT-U) is used, parallel strings = 1 and series strings = 1.

$$e_2 = I_L - I_L^* \tag{13}$$

where V_{PV}^* and I_L^* are reference voltage and current, respectively.

The schematic for the terminal sliding mode controller is deduced in Fig. 15.

4.1 MPV Searching Algorithm. A numerical method based on the linear dependence among open circuit voltage and reference voltage (optimal voltage) can be used to determine the system operating point corresponding to maximum power under varying irradiance and temperature conditions. The equation is given as

$$V_{PV}^* = K_v V_{oc} \tag{14}$$

where V_{oc} is the open circuit voltage of the PV system and K_v is voltage proportionality constant. Its value depends on the material and structure of the PV module, generally lying between 0.72 and 0.83. The open circuit voltage can be calculated by open circuiting the load cyclically and measuring the value; however, it will interrupt the power supply to the load. So, it is preferred to calculate directly from the electrical characteristics of the PV module.

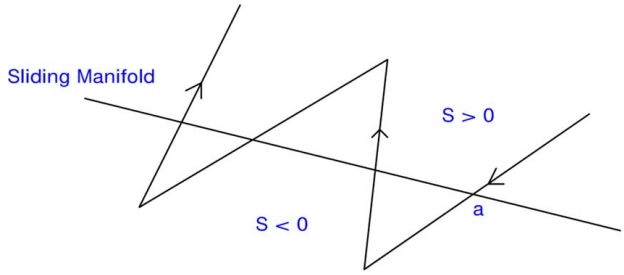


Fig. 9 Chattering phenomenon

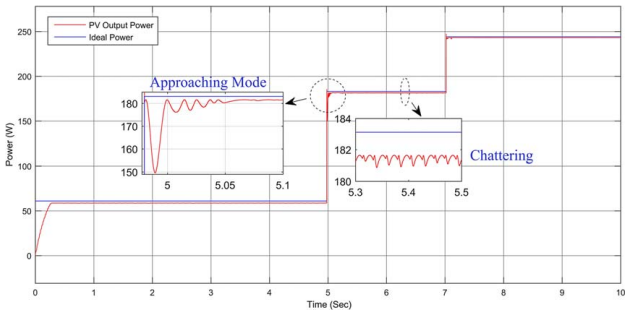


Fig. 10 Ideal power (obtained directly from PV module—SunPower SPR-305-WHT-U data under varying irradiance conditions) and PV output power (obtained by the product of V_{PV} and I_{PV}) observed under varying irradiance conditions. Temperature conditions are fixed at 25 °C.

On neglecting the shunt current under open circuit conditions, Eq. (2) can be written as

$$V_{oc} = n_s V_t \ln\left\{\frac{I_{ph} + I_o}{I_o}\right\} \tag{15}$$

where

$$V_t = \frac{AK_bT}{e}$$
$$I_o = I_{or}\left(\frac{T}{T_{ref}}\right)^3 \exp\left\{\frac{qE_g}{n_s K_b T}\left(\frac{1}{T_{ref}} - \frac{1}{T}\right)\right\} \tag{16}$$

4.1.1 Accuracy of MPV Searching Algorithm. The next step involves calculating the reference current I_L^*. The reference current can be calculated simply by using equation $I_L^* = I_{PV} - C_{PV}V_{PV}^*$ (Figs. 16 and 17).

Figure 18 depicts the reference current corresponding to the reference voltage obtained from MPV searching algorithm (Tables 3 and 4).

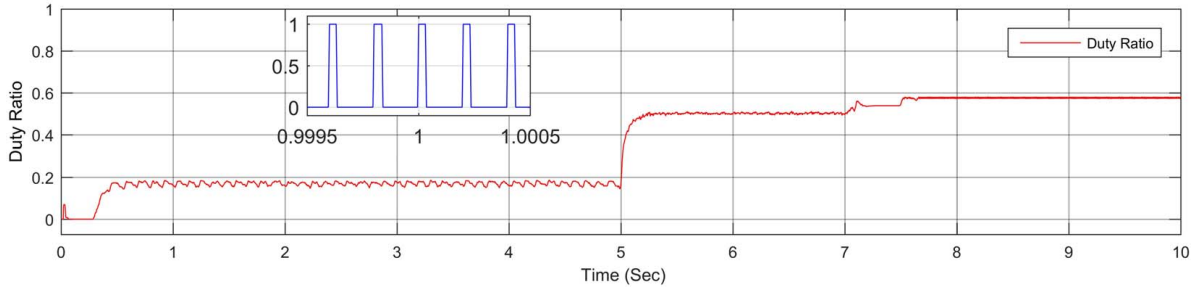


Fig. 8 Variations in duty cycle (d) observed with the changes irradiance in CSMC. Subplot shows a portion PWM pulses applied to the MOSFET(S) using PWM generator (switching freq. = 5000 and sampling time = 5×10^{-6}).

Downloaded from http://asmedigitalcollection.asme.org/solarenergyengineering/article-pdf/144/1/011003/6730015/sol_144_1_011003.pdf by Delhi Technological University user on 28 July 2021

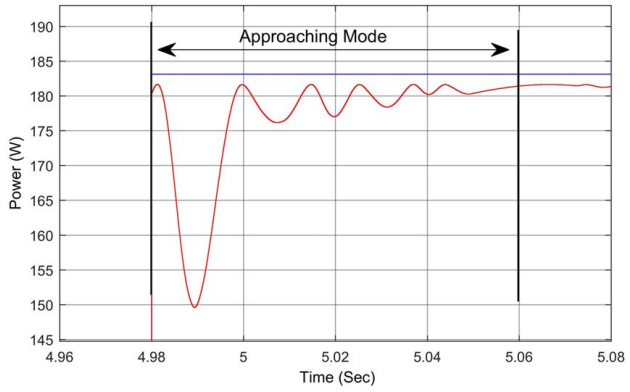


Fig. 11 Approaching mode parameters (Table 2) calculated using the cursor measurement feature of the scope

Table 2 Approaching mode performance parameters in CSMC

Parameters	Measurements
Overshoot	3.646%
Undershoot	−3.646%
Approaching time	83.116 ms

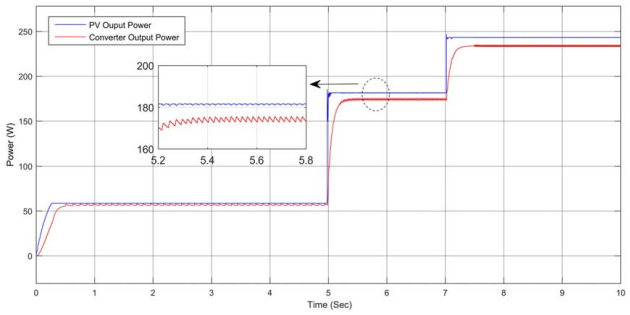


Fig. 12 PV module power (obtained as product of V_{PV} and I_{PV}) and converter output power (obtained as a product of I_{out} and V_o) under varying irradiance conditions. Temperature conditions are fixed at 25 °C.

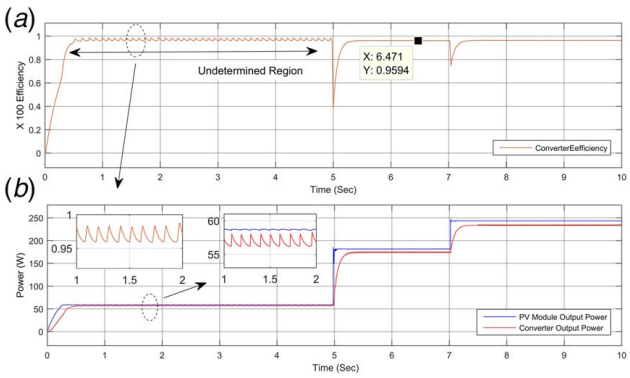


Fig. 13 (*a*) Converter efficiency (obtained by dividing the converter power by PV module power) under varying irradiance conditions, assuming there are no losses in the converter and (*b*) PV module and converter output power. Temperature conditions are fixed at 25 °C.

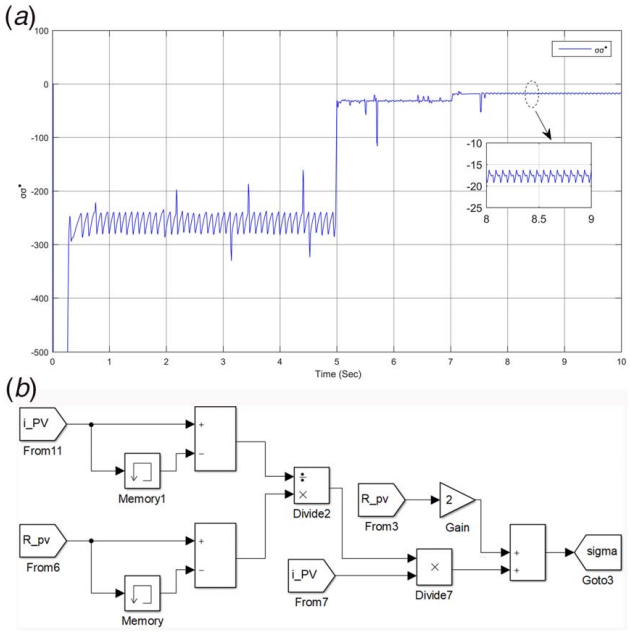


Fig. 14 (*a*) Plot for $\sigma\dot{\sigma}$ (obtained using Eqs. (8) and (10)) under varying irradiance conditions and (*b*) schematic model for obtaining σ

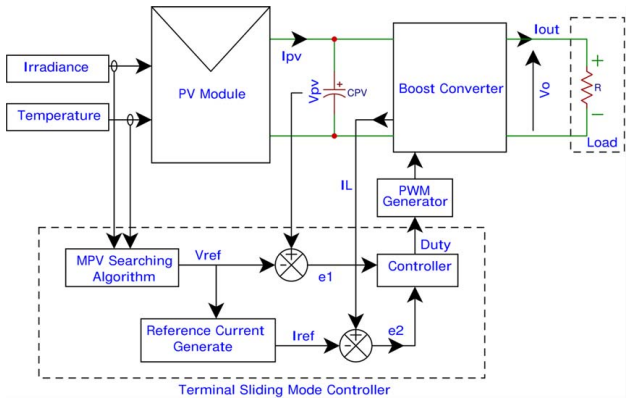


Fig. 15 Terminal sliding mode control. PV module (SunPower SPR-305-WHT-U) is used, parallel strings = 1 and series strings = 1. Keeping the converter parameters same as mentioned in Table 1.

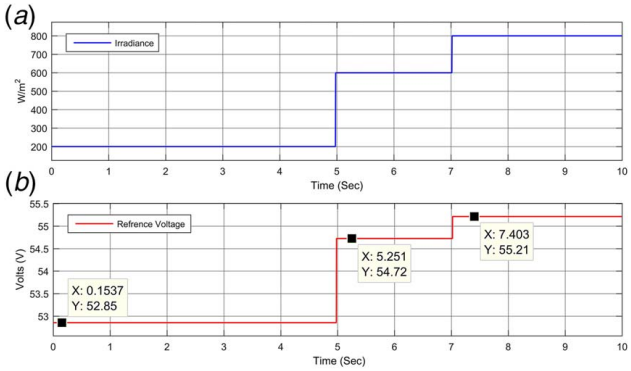


Fig. 16 (*a*) Irradiance curve using signal builder (same as used in CSMC) and (*b*) MPV values obtained under the varying irradiance condition from algorithm (keeping $K_v = 0.73$)

Downloaded from http://asmedigitalcollection.asme.org/solarenergyengineering/article-pdf/144/1/011003/6730015/sol_144_1_011003.pdf by Delhi Technological University user on 28 July 2021

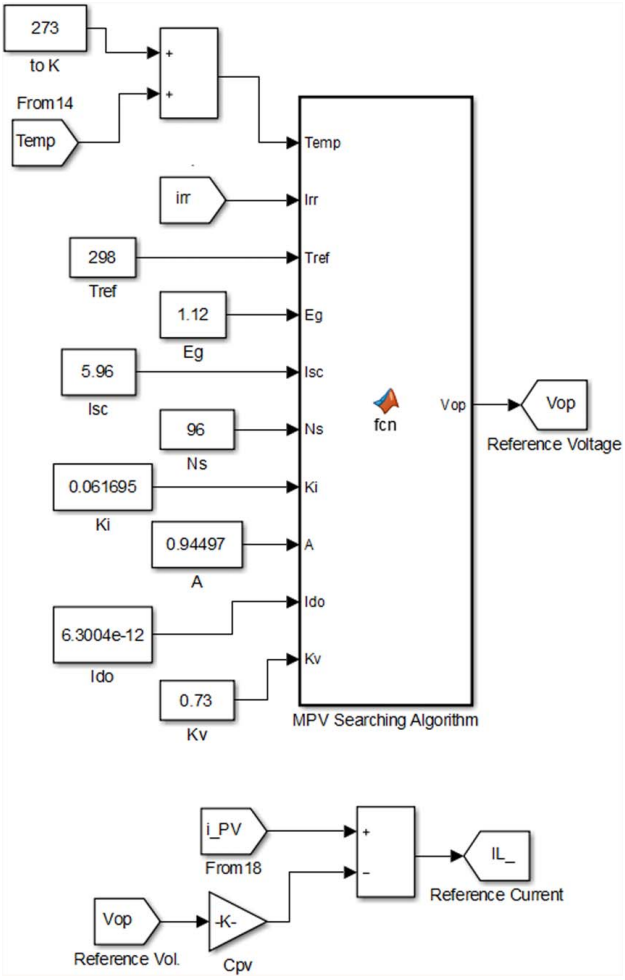


Fig. 17 Schematic for obtaining reference voltage V^*_{PV} and reference current I^*_L

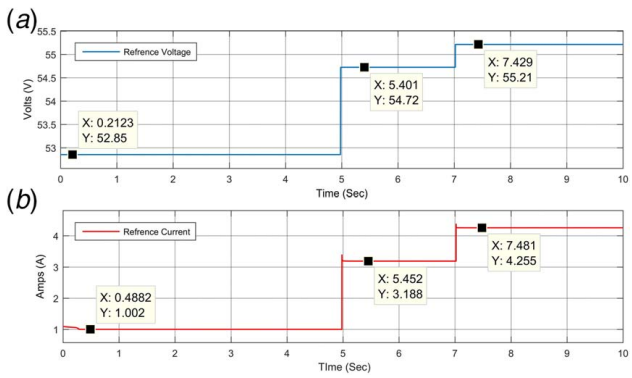


Fig. 18 (*a*) Reference voltage and (*b*) reference current under varying irradiance conditions

4.2 Sliding Surface Buildup. Till now, both the reference voltage V^*_{PV} and reference current I^*_L are calculated. The next step involves calculating the tracking errors defined as

$$e_1 = V_{PV} - V^*_{PV}$$

$$e_2 = I_L - I^*_L$$

where e_1 represents the tracking for voltage and e_2 represents the tracking error for current. The aim is to design an appropriate

Table 3 % Error between MPV values obtained from PV module characteristics and algorithm

Irradiance value (W/m^2)	MPV values obtained from PV module characteristics	MPV values obtained from algorithm	% Error
200	52.61 V	52.85 V	0.45
600	54.35 V	54.72 V	0.68
800	54.45 V	55.21 V	1.02

Table 4 Reference voltage and current

Irradiance value (W/m^2)	Reference voltage	Reference current
200	52.61 V	1.002 A
600	54.35 V	3.188 A
800	54.45 V	4.255 A

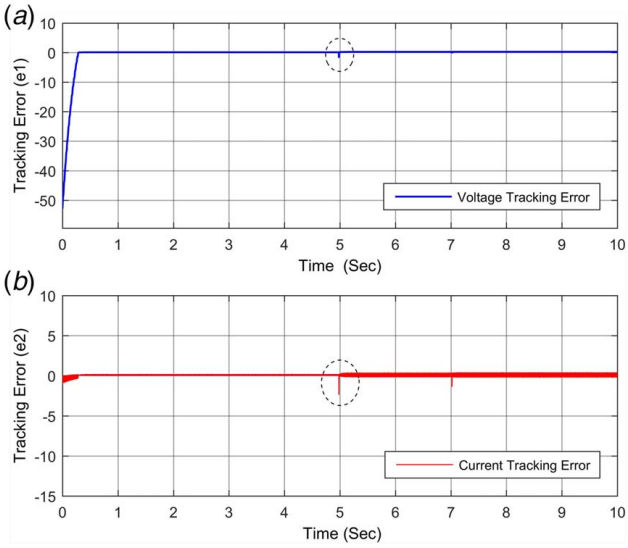


Fig. 19 Convergence of tracking errors: (*a*) e_1 and (*b*) e_2

sliding or switching manifold $\sigma(t)$, to drive these errors to 0 in finite time duration. Terminal sliding mode controller will cause all the system trajectories to approach toward the switching manifold and further stay or glide over the surface forever by the action of control law, i.e., by varying the duty cycle ratio of the converter.

The sliding surface that will be used for providing the control input is chosen as

$$\sigma(t) = \frac{1}{\beta} e_2^r - e_1 \tag{17}$$

where $\beta > 0$, $r = p/q$ with $1 < r < 2$ and p and q are odd integers.

The error dynamics can be described by taking the first derivate of the tracking errors:

$$\dot{e_1} = \dot{V_{PV}} - \dot{V^*_{PV}} = -\frac{e_2}{C_{PV}} \tag{18}$$

$$\dot{e_2} = \dot{I_L} - \dot{I^*_L} = \frac{V_{pv}}{L} - (1-d)\frac{V_o}{L} - \dot{I^*_L} \tag{19}$$

Downloaded from http://asmedigitalcollection.asme.org/solarenergyengineering/article-pdf/144/1/011003/6730015/sol_144_1_011003.pdf by Delhi Technological University user on 28 July 2021

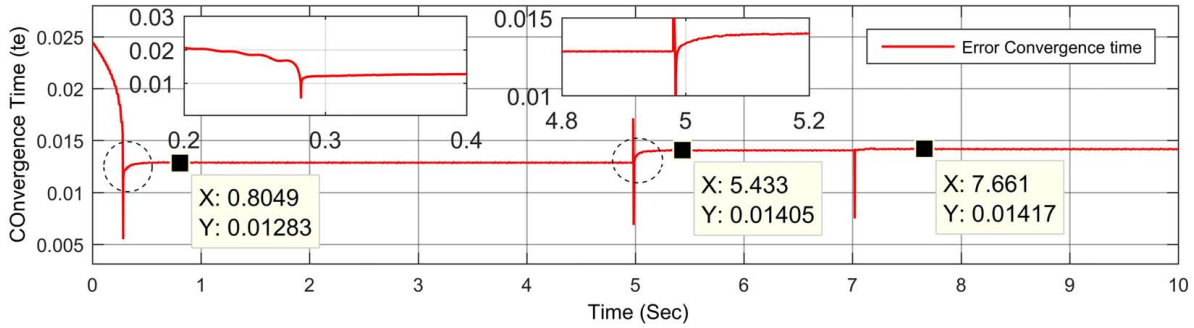


Fig. 20 Convergence time t_e (obtained by using Eq. (21))

Table 5 Convergence time and required parameters

Irradiance value (W/m^2)	Convergence time t_e	Δt_e change in time	Parameters
200	12.83 mS	14.05 – 12.83 = 1.22 mS	$\beta = 1.2$
600	14.05 mS	14.17 – 14.05 = 0.12 mS	$p = 19$
800	14.17 mS		$q = 17$

The sliding controller is satisfied when $\sigma(t) = 0$ then

$$e_1 = \frac{1}{\beta} e_2^r$$

$$e_2 = e_1^{1/r} \beta^{1/r}$$

Consequently the $\dot{e_1}$ becomes

$$\dot{e_1} = -\frac{\beta^{1/r}}{C_{PV}} e_1^{1/r} \tag{20}$$

This represents the dynamics of voltage tracking error and for satisfying the sliding condition it is required that e_1 should converge to zero, which will guarantee the convergence of tracking error e_2 to zero as shown in Fig. 19.

The dynamics of the voltage tracking error is given as

$$\dot{e_1} = -\frac{\beta^{1/r}}{C_{PV}} e_1^{1/r}$$

Assuming that system trajectories acquires the switching manifold in a finite duration of time t_o, i.e., $\sigma(t = t_o) = 0$ and let t_e be the time taken to reach the equilibrium point, i.e., convergence of $e_1(t_o + t_e) = 0$. The time t_e can be calculated by integrating the above equation on both sides

$$t_e = \frac{C_{PV}}{\beta^{1/r}\left(1 - \frac{1}{r}\right)} |e_1|^{1-1/r} \tag{21}$$

From Fig. 20, it can be concluded that tracking error converges in a finite time duration however, due to the variation in solar irradiance the value t_e changes as depicted in sub-plots (Table 5).

4.3 Development of Control Law. The final step involves developing an appropriate control law based on the switching manifold that will cause the tracking errors e_1 and e_2 to converge zero. The control law is providing the duty ratio to the switching Mosfet for guiding the system trajectories toward switching manifold to satisfy the condition $\sigma(t) = 0$

$$d = \frac{-L}{V_o}\left[\frac{1}{L}(V_{PV} - V_o) - (\dot{e_2} + \dot{I_L^*})\right] \tag{22}$$

The sliding controller is satisfied when $\sigma(t) = 0$ then

$$e_2 = e_1^{1/r} \beta^{1/r}$$

$$\dot{e_2} = -\frac{\beta e_2^{2-r}}{r C_{PV}} \tag{23}$$

Putting Eq. (23) in Eq. (22), we get

$$d = \frac{-L}{V_o}\left[\frac{1}{L}(V_{PV} - V_o) + \frac{\beta e_2^{2-r}}{r C_{PV}} - \dot{I_L^*}\right]$$

This is the extra term that was missing in the control law of the conventional sliding mode controller. So, the final control law

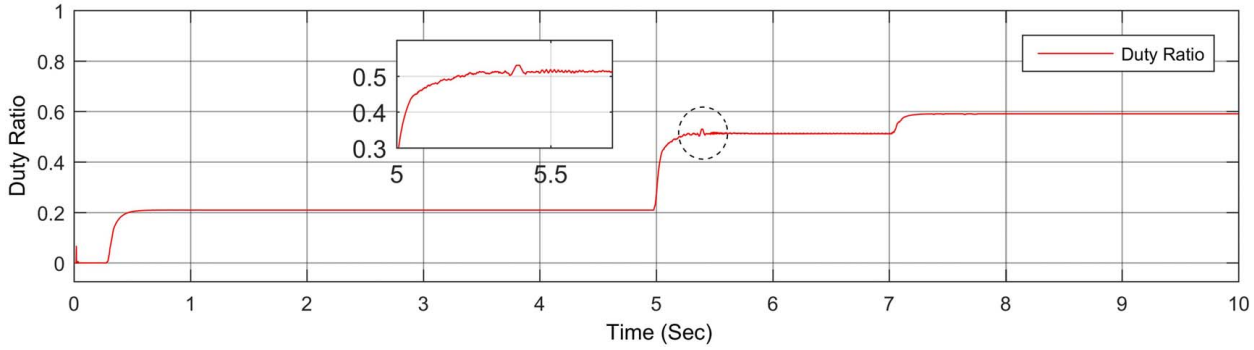


Fig. 21 Variations in duty cycle (d) observed with the changing irradiance in TSMC. PWM pulses can be obtained using PWM generator (switching freq. = 5000 and sampling time = 5×10^{-6}).

Downloaded from http://asmedigitalcollection.asme.org/solarenergyengineering/article-pdf/144/1/011003/6730015/sol_144_1_011003.pdf by Delhi Technological University user on 28 July 2021

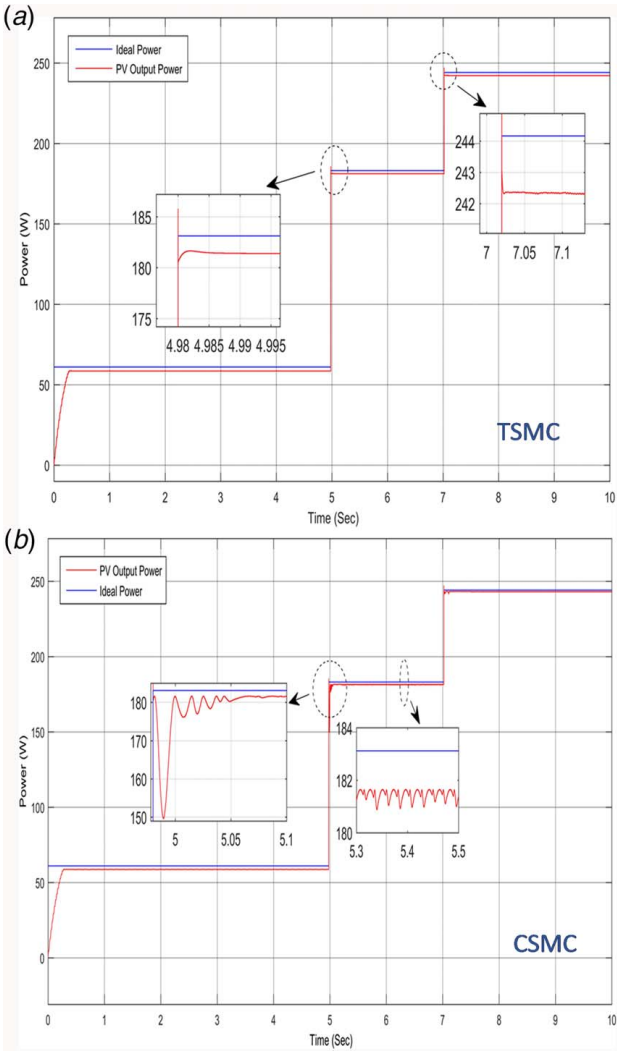


Fig. 22 Side by side analysis of ideal and PV output power in both (*a*) TSMC and (*b*) CSMC. Keeping the converter parameters same as mentioned in Table 1.

can be deduced as

$$d = \frac{-L}{V_o}\left[\frac{1}{L}\left(V_{\mathrm{PV}} - V_o\right) + \frac{\beta e_2^{2-r}}{rC_{\mathrm{PV}}} - I_L^{*\cdot} + K * \mathrm{sign}(\sigma)\right] \tag{24}$$

The plot in Fig. 21 shows that the duty cycle variation under the varying irradiance conditions as per the control action law to converge the errors to zero. On further comparing the duty cycle variation with that of conventional sliding mode controller, it can be observed that chattering phenomena is reduced considerably and almost a flat/linear duty ratio profile can be seen in TSMC.

On making a side-by-side analysis of approaching mode in both TSMC and CSMC as in Fig. 22, it can be concluded that overshoots and undershoots are reduced considerably in TSMC and the approaching time is also reduced. This is due to the reduction in chattering phenomenon. Analysis for the approaching mode in case of TSMC is shown in Fig. 23 and Table 6.

Chattering is reduced considerably by modifying the control law. The plot for the PV and converter output power is obtained in Fig. 24.

On making a side-by-side analysis of output powers of converter and PV system in TSMC and CSMC, it can be concluded that oscillations in the output power of converter are considerably reduced in TSMC as compared with that obtained in CSMC. This is due to a reduction in chattering phenomena. Also, the converter efficiency

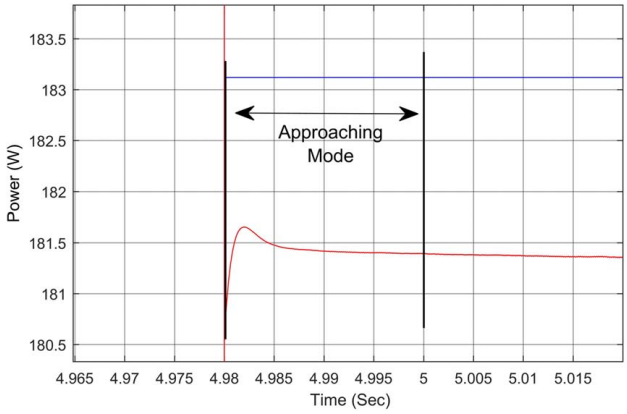


Fig. 23 Approaching mode in TSMC parameters (Table 6) calculated using the cursor measurement feature of the scope

Table 6 Approaching mode performance parmeters in TSMC

Parameters	Measurements
Overshoot	1.73%
Undershoot	−1.73%
Approaching time	17.096 ms

can be obtained under varying irradiance conditions as shown in Fig. 25 and Table 7.

From Fig. 26, it can be observed that the condition $\sigma\dot{\sigma} < 0$ is satisfied under all regions of irradiance curve hence proves for the existence of TSMC based controller based on Lyapunov's stability theory.

5 Results and Discussion

Two different types of sliding mode controllers are developed and studied in MATLAB/Simulink. After a brief analysis between TSMC and CSMC, it can be concluded that CSMC suffers from the chattering phenomenon due to which the oscillations were obtained in the output power of the converter.

Ideal PV characteristics of the module under varying irradiance are shown in Fig. 27. From the figure, it can be deduced that as the irradiance changes the sliding surface changes corresponding to new MPP.

5.1 Movement of Operating Point on Sliding Surface With the Change in Irradiance. PV output voltage and output power values were sampled when the system was simulated. From these sampled values, a PV curve is obtained, given in Fig. 28. It shows the movement of the system operating point when the irradiance changes and the chattering phenomena can also be observed significantly during change in switching surface. The closeness between the MPP points obtained from the sampled values and the ideal characteristics MPP points can be obtained by combining both the plots.

Similarly, PV output voltage and output power values are sampled in the case of TSMC when the system was simulated. From obtained values, a PV curve is obtained as given in Fig. 29. On making a side-by-side comparison with the CSMC, it can be concluded that the chattering phenomenon is reduced considerably in the case of TSMC. The closeness between the MPP points obtained from the sampled values with the ideal characteristics MPP points can be obtained by combining both the plots.

5.2 Effect of Temperature Variation. A side-by-side analysis of PV output power in both the TSMC and CSMC based

Downloaded from http://asmedigitalcollection.asme.org/solarenergyengineering/article-pdf/144/1/011003/6730015/sol_144_1_011003.pdf by Delhi Technological University user on 28 July 2021

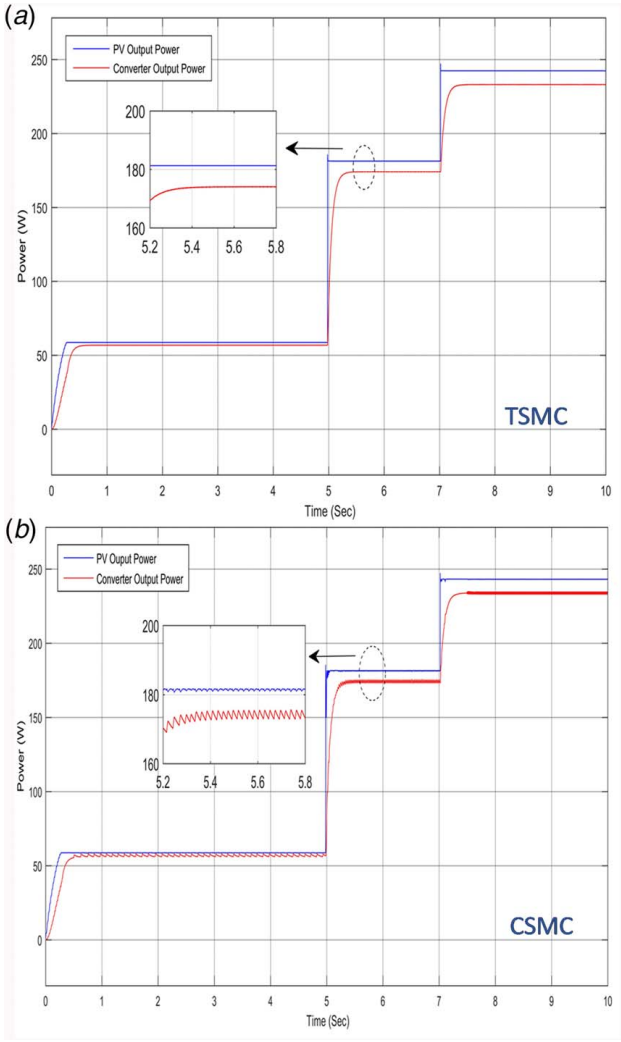


Fig. 24 Side by side analysis of PV and converter output power in both (*a*) TSMC (Fig. 15) and (*b*) CSMC (Fig. 3). Keeping the converter parameters same as mentioned in Table 1. Temperature conditions are fixed at 25 °C.

controllers have been carried out. The varying temperature conditions at a particular irradiance for which the controllers are analyzed are shown in Fig. 30.

From the above plot, it can be concluded that as the temperature increases the PV output power decreases. In the case of TSMC

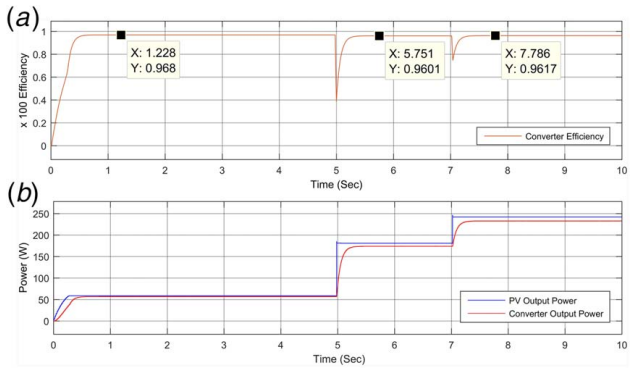


Fig. 25 (*a*) Converter efficiency (obtained by dividing the converter power by PV module power, Fig. 15) under varying irradiance conditions, assuming there are no losses in the converter and (*b*) PV output power and converter output power. Temperature conditions are fixed at 25 °C.

Table 7 Converter mean efficiency

Irradiance value (W/m^2)	Converter efficiency (%)	Mean efficiency
200	96.80	96.32%
600	96.01	
800	96.17	

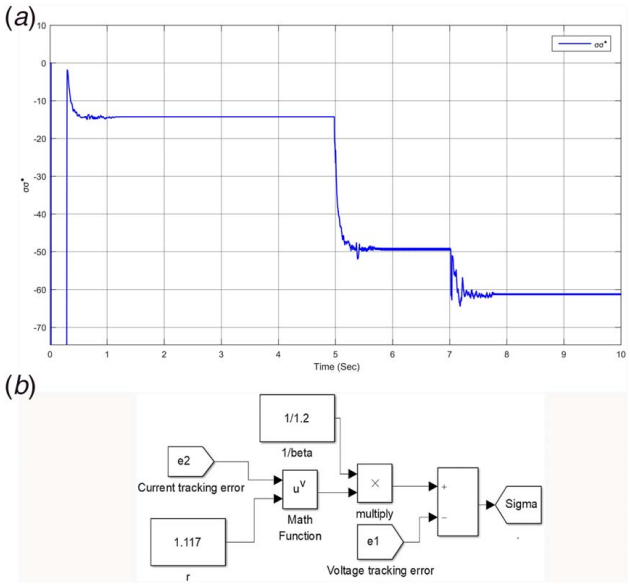


Fig. 26 (*a*) Plot for $\sigma\dot{\sigma}$ under varying irradiance conditions and (*b*) schematic model for obtaining σ

tracking algorithm, the deviation in the PV output power from the ideal power is more as compared with the CSMC tracking algorithm. However, in CSMC accurate tracking of photovoltaic power is obtained at the cost of more oscillations.

5.2.1 Dynamic Behavior of Operating Point With Temperature Variation. To visualize the effect of temperature variation on the movement of system's operating point, the system is subjected to varying temperature conditions as shown in Fig. 30.

From Fig. 31, it can be observed that as the temperature varies the system operating point tends to move downward which causes more deviations in the PV output power.

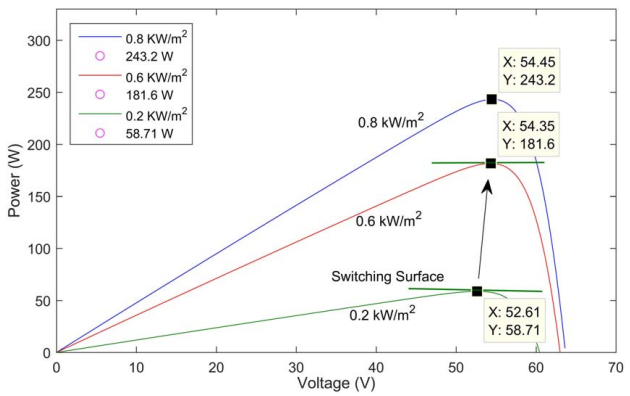


Fig. 27 Ideal PV characteristics of module under varying irradiance conditions. Temperature conditions are fixed at 25 °C.

Downloaded from http://asmedigitalcollection.asme.org/solarenergyengineering/article-pdf/144/1/011003/6730015/sol_144_1_011003.pdf by Delhi Technological University user on 28 July 2021

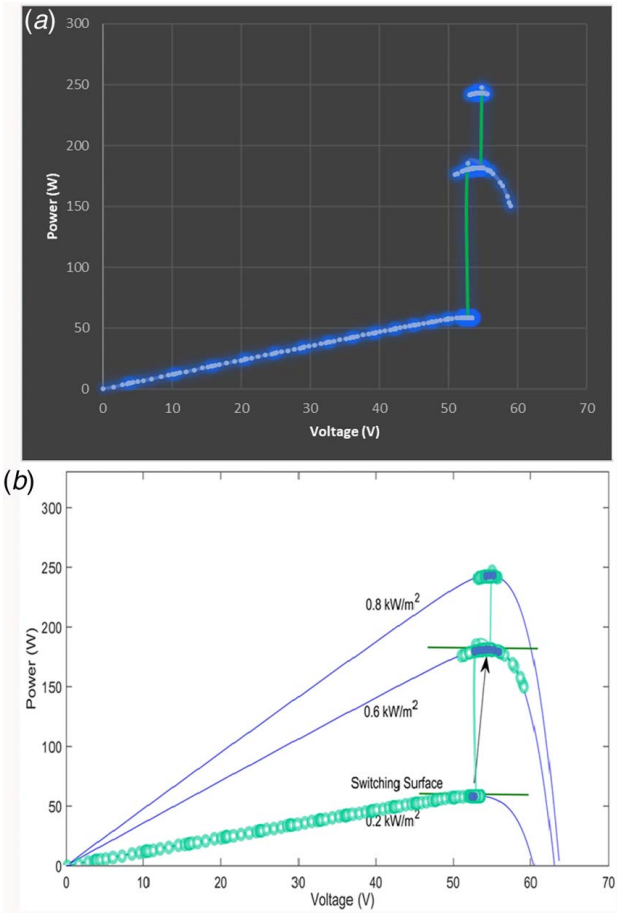


Fig. 28 (*a*) Movement of operating point with varying irradiance conditions in CSMC obtained by sampling (sampling time = 0.0025) the values (V_{PV} and P_{PV}) during simulation run time and (*b*) closeness of MPP points obtained theoretically and experimentally in CSMC

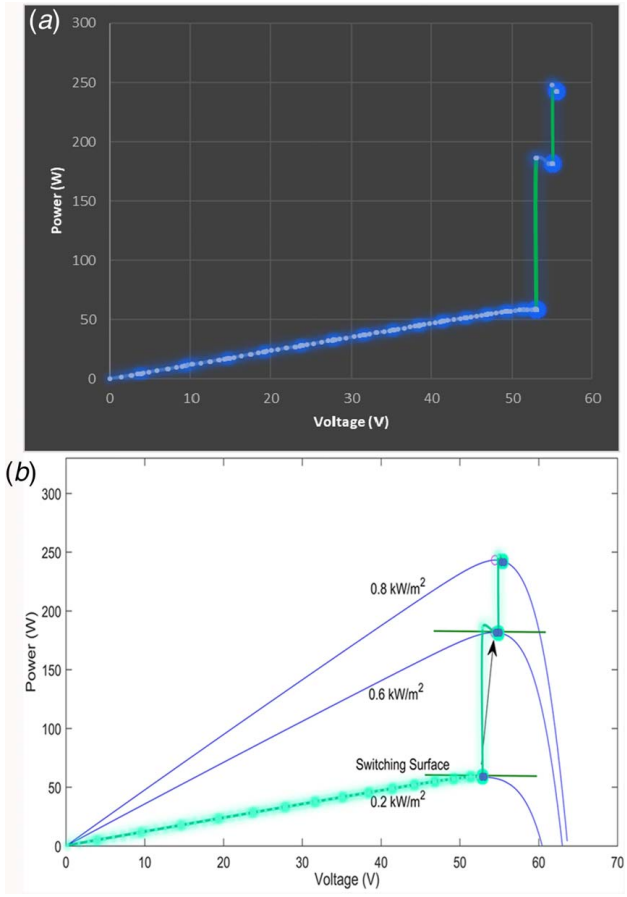


Fig. 29 (*a*) Movement of operating point with varying irradiance conditions in TSMC obtained by sampling (sampling time = 0.0025) the values (V_{PV} and P_{PV}) during simulation run time and (*b*) closeness of MPP points obtained theoretically and experimentally in TSMC

From Fig. 32, it can be observed that as the temperature varies the system operating point tends to move backward however, it stays on the same tracking track. It causes fewer deviations in the PV output power as compared with TSMC which can also be seen in Fig. 30.

5.3 Comparison With Other Hill-Climbing Method. Comparative analysis has been carried out between the sliding mode and conventional hill-climbing tracking algorithm, which is Perturbate and Observe Method. Similar work has been done by Saad et al. in the literature [19].

The plot for PV voltage and PV output power is shown in Fig. 33, under the conditions of same varying irradiance as shown in Fig. 6. From the plots, it can be observed that the conventional P&O tracking algorithm suffers from oscillations around the MPP. Similarly, plots under varying temperature conditions are shown in Fig. 34.

The controllers are further compared based on the following different performance parameters in Table 8.

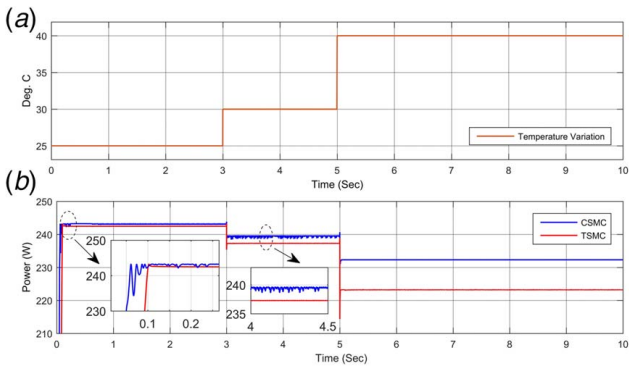


Fig. 30 (*a*) Temperature variations used for analysis of system operating point obtained by using signal builder and (*b*) PV output power in TSMC and CSMC obtained under varying temperature conditions with the irradiance condition fixed at 800 W/m^2

Downloaded from http://asmedigitalcollection.asme.org/solarenergyengineering/article-pdf/144/1/011003/6730015/sol_144_1_011003.pdf by Delhi Technological University user on 28 July 2021

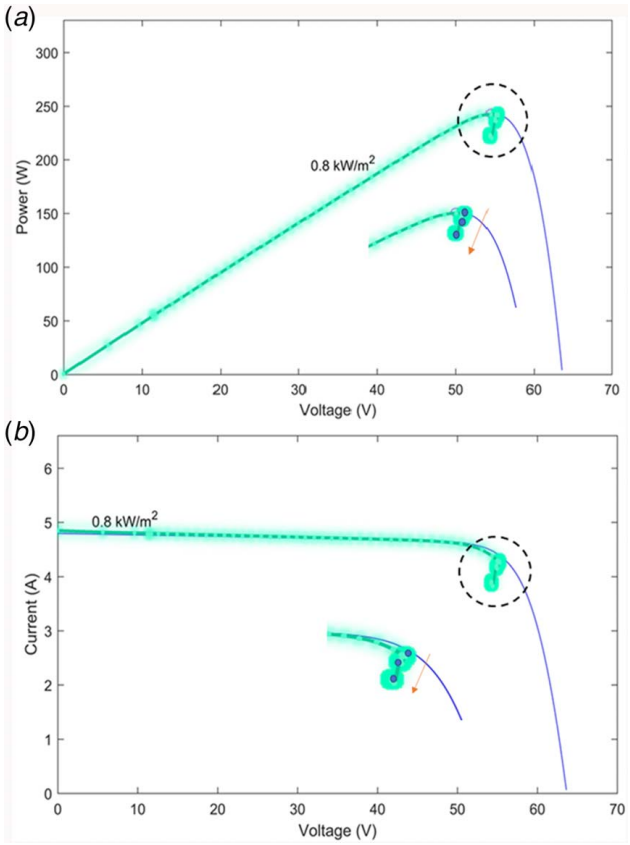


Fig. 31 (*a*) PV plot and (*b*) IV plot both displaying the behavior of operating point with temperature variation in TSMC with the Irradiance conditions fixed at 800 W/m^2. Temperature variations used as shown in Fig. 30. Keeping the converter parameters same as mentioned in Table 1.

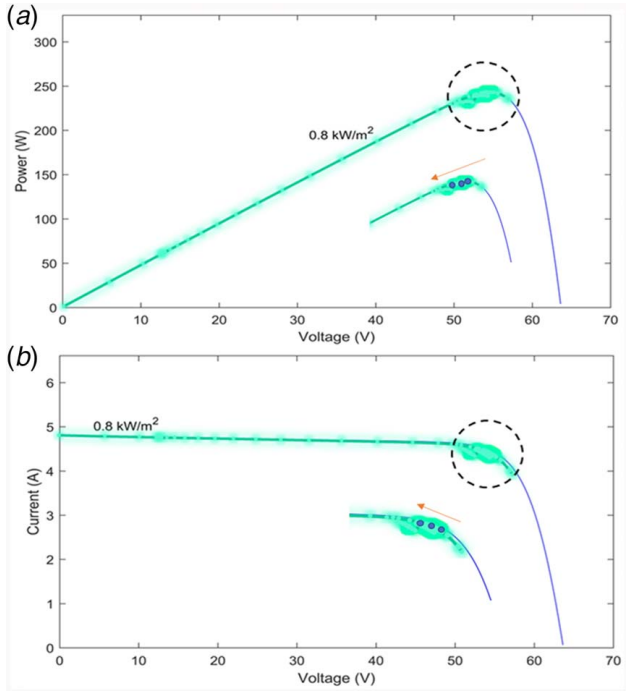


Fig. 32 (*a*) PV plot and (*b*) IV plot both displaying the behavior of operating point with temperature variation in CSMC with the Irradiance conditions are fixed at 800 W/m^2. Temperature variations used as shown in Fig. 30. Keeping the converter parameters same as mentioned in Table 1.

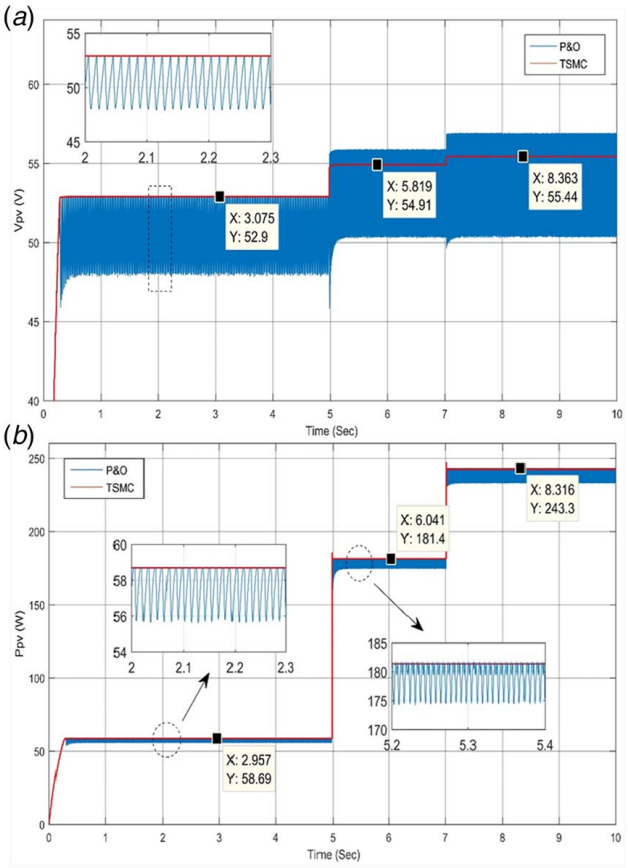


Fig. 33 (*a*) PV output voltage and (*b*) PV output power under the same varying irradiance conditions (Fig. 6). At a fixed temperature of 25 °C, keeping the converter parameters same as mentioned in Table 1.

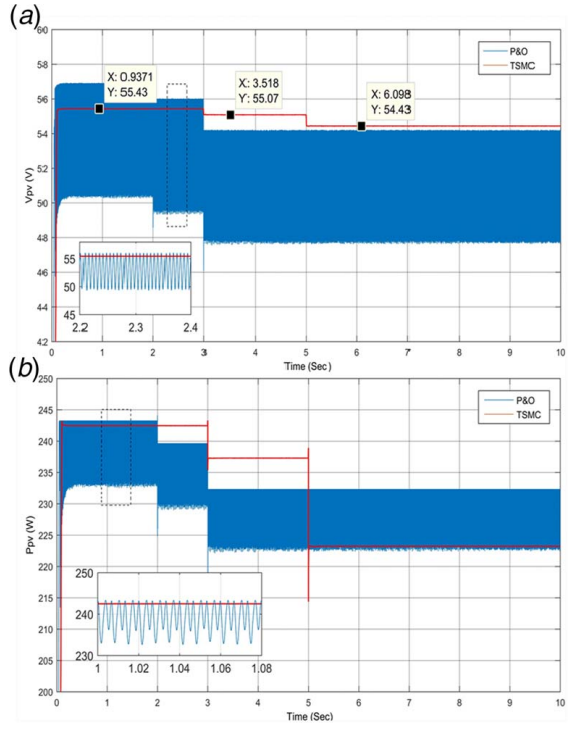


Fig. 34 (*a*) PV output voltage and (*b*) PV output power under the same varying temperature conditions (Fig. 30). At a fixed irradiance of 800 W/m^2, keeping the converter parameters same as mentioned in Table 1.

Downloaded from http://asmedigitalcollection.asme.org/solarenergyengineering/article-pdf/144/1/011003/6730015/sol_144_1_011003.pdf by Delhi Technological University user on 28 July 2021

Table 8 Performance parameters comparison between controllers

Performance parameters \ Controller type	Conventional sliding mode controller	Terminal sliding mode controller
Tracking accuracy	Low; Affected by the oscillations in the approaching mode and suffers from chattering phenomena	High; Modified sliding surface reduces the oscillations and removes the chattering
Duty cycle profile	Nonlinear; Suffers from oscillations due to chattering phenomena	Flat/linear; No chattering, have a better profile
Design complexity	Simple; Requires simple updation in duty ratio from the control law	Moderate; Requires formation of MPV searching algorithm and modified sliding surface with a control law
Convergence speed	Moderate; Tracking speed is moderate as it suffers from oscillations when system switches sliding surface and chattering phenomena exist	Fast; Modified sliding surface reduces the oscillations and chattering thereby improving the tracking speed
No. of prerequisite sensing of parameters	Less; It requires sensing of only 3 parameters, namely: V_{pv}, I_{pv}, V_o	Moderate; It requires sensing of 5 parameters, namely: G, T, V^*_{PV}, V_{PV}, I_L
PV and converter output power (accuracy of MPPT)	Low; Affected by the oscillations in the approaching mode and suffers from chattering phenomena	High; Modified sliding surface reduces the oscillations and removes the chattering

6 Conclusion

Two different types of sliding mode nonlinear controllers, CSMC and TSMC, are developed and studied in MATLAB/SIMULINK. Their comparative analyses have been carried out under dynamic climatic conditions. Based upon the results obtained, it is concluded that CSMC suffers from the chattering phenomenon which results in oscillations in the output power of the converter. The converter efficiency was indeterministic during the initial low irradiance regions of the CSMC converter. After that, modifications have been made in the switching surface in TSMC which resulted in the reduced chattering phenomenon. This paper presents an exhaustive performance analysis of nonlinear controllers for a standalone PV system with more parameters compared with the earlier work done by Saad et al. [19] in the literature.

Conflict of Interest

There are no conflicts of interest.

Nomenclature

d = duty ratio, dimensionless
A = ideality factor, dimensionless
G = irradiance, W/m^2
K = tuning parameter (positive constant), dimensionless
L = inductor, Henry (H)
R = load resistance, Ohm (Ω)
T = temperature, Kelvin (K)
I_{PP} = peak point current, Amps (A)
I_{SC} = short circuit current, Amps (A)
V_{OC} = open circuit voltage, Volts (V)
V_{PP} = peak point voltage, Volts (V)
e_1 = voltage tracking error, Volts (V)
e_2 = current tracking error, Amps (A)
n_s = number of series cells, dimensionless
t_e = convergence time, seconds (s)
C_o = capacitor (converter side), Farad (F)
C_{PV} = capacitor (PV side), Farad (F)
E_g = bandgap energy, eV
I_d = diode current, Amps (A)
I_L = inductor current, Amps (A)
I_{or} = diode dark saturation current (STP), Amps (A)
I_{ph} = photogenerated current, Amps (A)
I_{PV} = PV module current, Amps (A)
K_b = Boltzmann constant, dimensionless
K_i = short circuit coefficient of temperature, %/°C
K_v = voltage proportionality constant, dimensionless
R_{pv} = equivalent load connected, Ohm (Ω)
R_s = series resistance of PV module, Ohm (Ω)
R_{sh} = shunt resistance of PV module, Ohm (Ω)
T_{ref} = reference temperature, Kelvin (K)
V_o = output voltage, Volts (V)
V_{PV} = PV module voltage, Volts (V)
V_t = temperature equivalent voltage, Volts (V)
$q(e)$ = charge of electron, Coulomb

Greek Symbols

β = tuning parameter, dimensionless
$\eta(A)$ = ideality factor, dimensionless
σ = sliding surface, dimensionless

References

[1] Irfan, M., Zhao, Z. Y., Ikram, M., Gilal, N. G., Li, H., and Rehman, A., 2020, "Assessment of India's Energy Dynamics: Prospects of Solar Energy," J. Renewable Sustainable Energy, **12**(5), p. 053701.
[2] Piegari, L., and Rizzo, R., 2010, "Adaptive Perturb and Observe Algorithm for Photovoltaic Maximum Power Point Tracking," IET Renewable Power Gener., **4**(4), pp. 317–328.
[3] Elgendy, M., Zahawi, B., and Atkinson, D. J., 2012, "Evaluation of Perturb and Observe MPPT Algorithm Implementation Techniques," 6th IET International Conference on Power Electronics, Machines and Drives (PEMD 2012), Bristol, UK, Mar. 27–29, 2012, pp. 110–116.
[4] Ballaji, A., Hediyal, N., Mandi, P. R., and Swamy, N. K., 2018, "Design and Implementation of Perturb and Observation Maximum Power Point Transfer (MPPT) Algorithm for Photovoltaic System," Int. J. Eng. Res. Appl., **8**(5), pp. 16–23.
[5] Al-Diab, A., and Sourkounis, C., 2010, "Variable Step Size P&O MPPT Algorithm for PV Systems," 12th International Conference on Optimization of Electrical and Electronic Equipment, Basov, Romania, May 20–22, 2010, pp. 1097–1102.
[6] Qin, S., Wang, M., Chen, T., and Yao, X., 2011, "Comparative Analysis of Incremental Conductance and Perturb-and-Observation Methods to Implement MPPT in Photovoltaic System," International Conference on Electrical and Control Engineering, Yichang, China, Sept. 16–18, 2011, pp. 5792–5795.
[7] Harini, K., and Syama, S., 2015, "Simulation and Analysis of Incremental Conductance and Perturb and Observe MPPT With DC-DC Converter Topology for PV Array," IEEE International Conference on Electrical,

Downloaded from http://asmedigitalcollection.asme.org/solarenergyengineering/article-pdf/144/1/011003/6730015/sol_144_1_011003.pdf by Delhi Technological University user on 28 July 2021

Computer and Communication Technologies (ICECCT), Coimbatore, India, Mar. 5–7, 2015, pp. 1–5.
[8] Kjær, S. B., 2012, "Evaluation of the "Hill Climbing" and the "Incremental Conductance" Maximum Power Point Trackers for Photovoltaic Power Systems," IEEE Trans. Energy Convers., **27**(4), pp. 922–929.
[9] Putri, R., Wibowo, S., and Rifa'I, M., 2015, "Maximum Power Point Tracking for Photovoltaic Using Incremental Conductance Method," Energy Procedia, **68**(5), pp. 22–30.
[10] Zainuri, M. A. A. M., Radzi, M. A. M., Che Soh, A., and Rahim, N. A., 2014, "Development of Adaptive Perturb and Observe-Fuzzy Control Maximum Power Point Tracking for Photovoltaic Boost dc-dc Converter," IET Renewable Power Gener., **8**(2), pp. 183–194.
[11] Abdelkader, H. I., Hatata, A. Y., and Hasan, M. S., 2015, "Developing Intelligent MPPT for PV Systems Based on ANN and P&O Algorithms," Int. J. Sci. Eng. Res., **6**(2), pp. 367–373.
[12] Blange, R., Mahanta, C., and Gogoi, A. K., 2015, "MPPT of Solar Photovoltaic Cell Using Perturb & Observe and Fuzzy Logic Controller Algorithm for Buck-Boost DC-DC Converter," International Conference on Energy, Power and Environment: Towards Sustainable Growth (ICEPE), Shillong, India, June 12–13, 2015, pp. 1–5.
[13] Selman, N. H., and Mahmood, J. R., 2016, ""Comparison Between Perturb & Observe, Incremental Conductance and Fuzzy Logic MPPT Techniques at Different Weather Conditions," Int. J. Innovative Res. Sci. Eng. Technol., **5**(7), pp. 12556–12569 .
[14] Khateb, A. E., Rahim, N., Selvaraj, J., and Uddin, M. N., 2014, "Fuzzy-Logic-Controller-Based SEPIC Converter for Maximum Power Point Tracking," IEEE Trans. Ind. Appl., **50**(4), pp. 2349–2358.
[15] Farhat, M., Barambones, O., and Sbita, L., 2017, "A new Maximum Power Point Method Based on a Sliding Mode Approach for Solar Energy Harvesting," Appl. Energy, **185**(2), pp. 1185–1198.
[16] Cabal, C., Salamero, L. M., Seguier, L., Alonso, C., and Guinjoan, F., 2014, "'Maximum Power Point Tracking Based on Sliding Mode Control for Output Series Connected Converters in Photovoltaic Systems'," IET Power Electron., **7**(4), pp. 914–923.
[17] Chiu, C. S., Ouyang, Y. L., and Ku, C. Y., 2012, "'Terminal Sliding Mode Control for Maximum Power Point Tracking of Photovoltaic Power Generation Systems'," Sol. Energy, **86**(10), pp. 2986–2995.
[18] Vega, P. V., and Hirzinger, G., 2001, "Chattering-free Sliding Mode Control for a Class of Nonlinear Mechanical Systems," J. Robust Nonlinear Control, **11**(12), pp. 1161–1178.
[19] Saad, W., Sellami, A., and Garcia, G., 2018, "Terminal Sliding Mode Control-Based MPPT for a Photovoltaic System With Uncertainties," Int. J. Modell. Identif. Control, **29**(2), pp. 118–126.
[20] Afghoul, H., Chikouche, D., Krim, F., and Beddar, A., 2013, "A Comparative Study Between Sliding Mode Controller and P and O Controller Applied to MPPT," Proceedings of 2013 International Renewable and Sustainable Energy Conference (IRSEC), Ouarzazate, Morocco, Mar. 7–9, pp. 112–117.
[21] Kchaou, A., Naamane, A., Koubaa, Y., and M'Sirdi, N., 2017, "Second Order Sliding Mode-Based MPPT Control for Photovoltaic Applications," Sol. Energy, **155**(15), pp. 758–769.
[22] Zhu, Y., and Fei, J., 2017, "Adaptive Global Fast Terminal Sliding Mode Control of Grid-Connected Photovoltaic System Using Fuzzy Neural Network Approach," IEEE Access, **5**, pp. 9476–9484.
[23] Azevedo, G. M. S., Cavalcanti, M. C., Oliveira, K. C., Neves, F. A. S., and Lins, Z. D., 2009, "Comparative Evaluation of Maximum Power Point Tracking Methods for Photovoltaic Systems," ASME J. Sol. Energy Eng., **131**(3), p. 031006.
[24] Koshkouei, A. J., and Zinober, A. S. I., 2000, "Sliding Mode Control of Discrete-Time Systems," ASME J. Dyn. Syst. Meas. Control, **122**(4), pp. 793–802.

Downloaded from http://asmedigitalcollection.asme.org/solarenergyengineering/article-pdf/144/1/011003/6730015/sol_144_1_011003.pdf by Delhi Technological University user on 28 July 2021

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/353170180

Design and Simulation of Connecting Rod using Finite Element Analysis

Research · July 2021

DOI: 10.13140/RG.2.2.10672.07689

CITATIONS
0

READS
3

1 author:

Suraj Kumar
Delhi Technological University
3 PUBLICATIONS **0** CITATIONS

SEE PROFILE

All content following this page was uploaded by Suraj Kumar on 12 July 2021.

The user has requested enhancement of the downloaded file.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 07 | July 2021 www.irjet.net p-ISSN: 2395-0072

Design and Simulation of Connecting Rod using Finite Element Analysis

Suraj Kumar[1]

[1]*Student of Mechanical Engineering, Delhi Technological University, Delhi, India*

Abstract – *A connecting rod is the core component of an I.C. engine that has a function to transmit power released from combustion by the piston to the crankshaft. During the combustion process, deformation is the most commonly occurring damage in the connecting rods. The connecting rods made of aluminium alloy and alloy steel are designed & assembled using PTC CREO 7.0 software and analysed using the Finite Element Analysis (FEA) method in ANSYS. The simulation has shown the result of stress, strain, and total deformation in the connecting rod occur during the operating load condition. This article compares and discusses the various obtained results and selects the best material for the connecting rod.*

Keywords: **– Connecting Rod, Calculations, PTC CREO, Modelling, Assembly, Simulation, Finite Element Analysis.**

1. INTRODUCTION

Internal Combustion engine has many parts like cylinder block, piston, connecting rod, crank, and crankshaft. The connecting rod is very important part of an I.C. engine. It comprises of an eye at the small end to assemble with a piston with the help of a piston pin, a long shank, and a big end opening split into two parts to assemble with the crank pin. The main function of the connecting rod is to transmit power from the piston pin to the crank pin. The connecting rod converts the reciprocation motion developed by the piston to the rotary motion of the crankshaft. It also carries the lubrication oil from the oil sump (pan) to the piston assembly through the crank pin.

The drop forging process is used to make the connecting rod of an I.C. engine and its external surfaces are omitted unfinished. Mostly split type or two-piece connecting rod is used in internal combustion engines. The eye and shank are forged in a single piece; the rod cap is cut off, the outer surface and bolt portion left for the machining process of the big end. The small end of the connecting rod is usually made in the form of a solid eye and then machined.

The connecting rod is subjected to gas pressure by the force of combustion fuel and inertia force by its reciprocation parts. It is one of the most heavily stressed components of I.C. engines. The length of the connecting rod is a primary consideration. When the connecting rod is smaller than the crank radius, it has maximum angular swing, consequently greater side thrust on the piston. In low-speed engines, the length of connecting rod and radius of crank (L/r) is usually 4 or less than 4. In high-speed engines, the (L/r) ratio differ from 4 to 5.

In high-speed engines, mostly I-section connecting rods are used and it helps to reduce the weight as well as inertia forces. The I-section will also be easier for the forging process. For circulating lubricant oil, most of the connecting rod has a rifled-drilled hole in the whole length from the big end to the small end for transfer oil from the rod cap to the piston pin. While circular cross-section is used in low-speed engines [1].

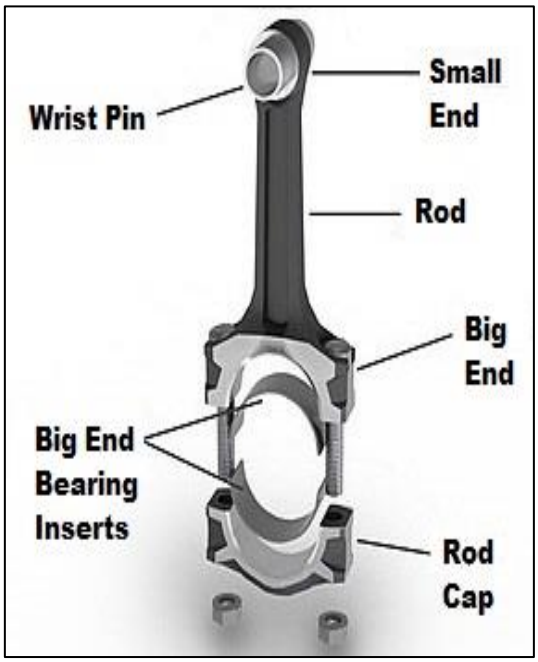


Fig -1: Connecting Rod Nomenclature [2]
Source: Wikipedia

1.1. Introduction to PTC CREO Software

CREO is a family or suite of Computer-Aided Design (CAD) apps supporting product design for discrete manufacturers and is developed by Parametric Technology Corporation (PTC). It is upgraded version for Pro/ENGINEER (Pro/E) software used for modelling, analysis and simulation. It is widely used for product design and development in automobile firms and manufacturing plants. It differs from AutoCAD in that it is based on

parametric, this means all dimensions are parametric which affects the actual portion [3].

1.2. Introduction to Ansys Software

ANSYS is a general-purpose finite element modelling packages for numerically different types of solutions to mechanical problems. Simulation of a system is the operation of a model in terms of time or space, which helps analyse the performance of an existing or a proposed system. In other words, simulation is the process of using a model to study the performance of a system. It is an act of using a model for simulation [4,5].

2. OBJECTIVES

i. To learn and get familiar with the PTC CREO and Finite Element Analysis.
ii. To create a 3D model of the connecting rod in PTC CREO 7.0 and imported into ANSYS software for further analysis.
iii. Study the effects of the loads acting on the connecting rod under the considered loading conditions.
iv. To evaluation of the comparison between results.

3. DESIGN OF CONNECTING ROD

Taking the following data of four-stroke internal combustion engine for designing a connecting rod.

Bore x Stroke	-	100mm x 190mm
The compression ratio of the engine	-	6:1
Engine speed	-	1800 RPM
Maximum pressure (Pmax.)	-	3.15 N/mm^2
Mass of reciprocating parts	-	2.25 kg
Length of connecting rod	-	380 mm
Factor of safety (FOS)	-	6
Length of dia. ratio for Big End	-	1.3
Length of dia. ratio for Small End	-	2
Bearing pressure at Big End	-	10 N/mm^2
Bearing pressure at Small End	-	15 N/mm^2
Allowable stress in Bolts	-	60 N/mm^2
Allowable stress in Cap	-	80 N/mm^2
Elastic limit in compression	-	320 N/mm^2
Constant (a)	-	1/7500

3.1. Dimensions of Cross-Section of the Connecting Rod

The connecting rod is a component that is subjected to the axial compressive force which is equal to the maximum gas force acting on the piston. The compressive stress is of significant magnitude. Hence, the design of the connecting rod is as a column or a strut [1].

- For buckling, the connecting rod is four times stronger about *YY*-axis than the about *XX*-axis.
- If a rod is designed just like that it is equal resistant to buckling on both planes, then,

$$4I_{yy} = I_{xx}$$

We know that,

$$k = \sqrt{\frac{I}{A}}$$

$$k^2 = \frac{I}{A}$$

$$\boldsymbol{I = Ak^2}$$

So that,

$$4\,k_{yy}^2 = k_{xx}^2$$

$$\boldsymbol{k_{yy}^2 = \frac{1}{4}\,k_{xx}^2}$$

- The above relation justifies that the I-section of the connecting rod is ideally useful. While a circular cross-section for connecting rod is not suitable for buckling about the *YY*-axis [1].

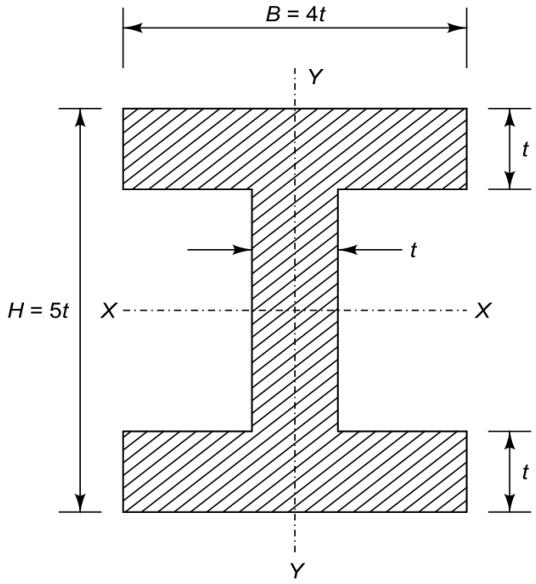


Fig -2: I – Section
Source: Design of machine elements, third edition, page no. 869, fig. 25.14

- **Moment of Inertia about *XX*-axis:**

$$I_{xx} = \frac{1}{12}(4t) \times (5t)^3 - \frac{1}{12}(4t - t) \times (5t - 2t)^3$$
$$= \frac{1}{12}(500t^4 - 81t^4)$$
$$= \left(\frac{419}{12}\right) t^4$$
$$\boldsymbol{I_{xx} = 34.91\, t^4}$$

And,

$$k_{xx}^2 = \frac{I_{xx}}{A}$$
$$= \left(\frac{419\, t^4}{12}\right) \times \left(\frac{1}{11\, t^2}\right)$$
$$k_{xx}^2 = 3.17\, t^2$$
$$\boldsymbol{k_{xx} = 1.78\, t}$$

- **Moment of Inertia about *YY*-axis:**

$$I_{yy} = 2\left[\frac{1}{12}(t) \times (4t)^3\right] + \frac{1}{12}(5t - 2t) \times (t)^3$$
$$= \frac{1}{12}(128t^4 + 3t^4)$$
$$= \left(\frac{131}{12}\right) t^4$$
$$\boldsymbol{I_{yy} = 10.91\, t^4}$$

And,

$$k_{yy}^2 = \frac{I_{yy}}{A}$$
$$= \left(\frac{131\, t^4}{12}\right) \times \left(\frac{1}{11\, t^2}\right)$$
$$\boldsymbol{k_{yy}^2 = 0.992\, t^2}$$

Therefore,

$$\frac{I_{xx}}{I_{yy}} = \frac{\left(\frac{419}{12}\right) t^4}{\left(\frac{131}{12}\right) t^4}$$
$$= \frac{419}{131}$$
$$\frac{I_{xx}}{I_{yy}} = 3.2$$
$$\boldsymbol{3.2\, I_{yy} = I_{xx}}$$

Chosen of I-section is quite satisfactory because the values of I_{xx}/I_{yy} is coming between 3 and 3.5.

- **Area of the Cross-section:**

$$A = 2(4t \times t) + (5t - 2t) \times t$$
$$= 8t^2 + 3t^2$$
$$\boldsymbol{A = 11t^2}$$

➢ Load Acting on Connecting Rod by Gases

A mechanism diagram of the connecting rod and crank are shown below:

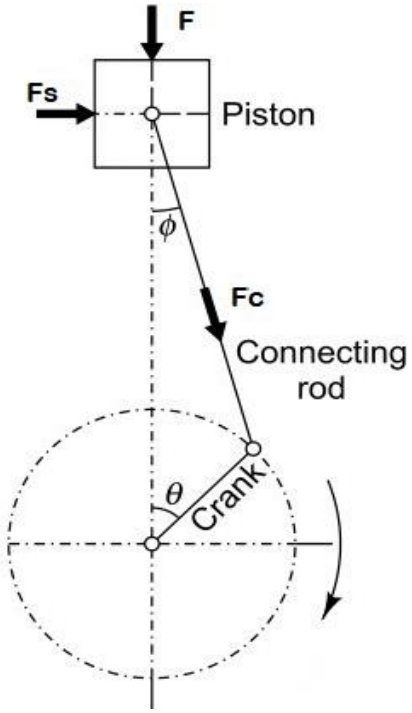


Fig -3: Crank and Connecting Rod Mechanism
Source: Design of machine elements, third edition, page no. 870, fig. 25.15

$$F_C = \frac{\pi}{4} \times D^2 \times P_{max.}$$
$$= \frac{\pi}{4} \times (100)^2 \times 3.15$$
$$\boldsymbol{F_C = 24740\ N}$$

➢ Critical Buckling Load on Connecting Rod

$$F_{CB} = F_C \times FOS$$
$$= 24740 \times 6$$
$$\boldsymbol{F_{CB} = 148440\ N}$$

➢ Dimensions of Cross- section

By Rankine Formula,

$$F_{CB} = \frac{\sigma_c A}{1 + a\left(\frac{L}{k_{xx}}\right)^2}$$

∴ Due to the both ends hinged, L = l

$$148440 = \frac{320 \times 11t^2}{1 + \frac{1}{7500} \times \left(\frac{380}{1.78\, t}\right)^2}$$

$$148440\left[1+\frac{1}{7500}\times\left(\frac{380}{1.78\,t}\right)^2\right]=320\times11t^2$$

$$148440+\frac{148440\times(380)^2}{7500\times(1.87)^2\times t^2}=320\times11t^2$$

$$148440+\frac{902021.5}{t^2}=320\times11t^2$$

$$148440\,t^2+902021.5=320\times11t^4$$

$$\frac{148440}{320\times11}t^2+\frac{902021.5}{320\times11}=t^4$$

$$42.17\,t^2+256.26=t^4$$
$$t^4-42.17\,t^2-256.26=0$$

The above function is a quadratic equation in terms of (t^2).

$$x=\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

$$t^2=\frac{42.17\pm\sqrt{(-42.17)^2+4\times1\times256.26}}{2\times1}$$

$$=\frac{42.17\pm52.95}{2}$$
$$t^2=47.56$$
$$t=6.9\text{ mm OR}$$
$$\boldsymbol{t=7\text{ mm}}$$

- The width B of the connecting rod is kept constant throughout its length.

Width (B) $=4\,t=4\times7=$ **28 mm**

- The height H varies from the big end to the small end in the in the following way:

i. At the middle section, Height (H) = 5 t
H $=5\,t=5\times7=$ **35 mm**
ii. At the small end, H1 = 0.75 H to 0.9 H
H1 $=0.75$ H $=0.85\times35=29.75$ mm OR
H1 = **30 mm**
iii. At the big end, H2 = 1.1 H to 1.25 H
H2 $=1.1$ H $=1.2\times35=$ **42 mm**

3.2. Dimensions of the Big End and Small End Bearing of Connecting Rod

➢ Crank Pin Bearing

$$F_C=d_C\times l_C\times(P_b)_C$$

Relation, $l_C=1.25\,d_C$ to $1.5\,d_C$

$$F_C=d_C\times1.3\,d_C\times(P_b)_C$$
$$24740=d_C\times1.3\,d_C\times10$$
$$d_C^2=1903.08$$
$$d_C=43.63\text{ OR}$$
$$\boldsymbol{d_C=44\,mm}$$

$$l_C=1.3\,d_C=1.3\times44$$
$$l_C=57.2\,mm\text{ OR}$$
$$\boldsymbol{l_C=57\,mm}$$

➢ Piston Pin Bearing

$$F_C=d_P\times l_P\times(P_b)_P$$

Relation, $l_P=1.5\,d_P$ to $2\,d_P$

$$F_C=d_P\times2\,d_P\times(P_b)_P$$
$$24740=d_P\times2\,d_P\times15$$
$$d_P^2=824.67$$
$$d_P=28.71\,mm\text{ OR}$$
$$\boldsymbol{d_P=29\,mm}$$

$$l_P=2\,d_P=2\times29$$
$$\boldsymbol{l_P=58\,mm}$$

3.3. Size of Bolts for Securing the Big End Cap

The maximum inertia force acting on the cap and bolts at the top dead centre during exhaust stroke. The inertia force action on cap and bolt which is given by,

$$F_I=m_r r\omega^2\times\left[Cos\theta+\frac{Cos2\theta}{n}\right]$$

At the top dead centre position, the inertia force will be maximum. where ($\theta=0$).
When ($\theta=0$),

$$Cos\theta=1\text{ and }Cos2\theta=1$$

So,

$$=m_r r\omega^2\times\left[1+\frac{1}{n}\right]$$

$$F_I=m_r r\omega^2\times\left[1+\frac{r}{l}\right]\quad\therefore\quad\text{Here, n = l/r}$$

As we know, the crank radius is found by,

$$r=\frac{l}{2}=\frac{190}{2}$$
$$\boldsymbol{r=95\,mm}$$

And then,

$$F_I=2.25\times0.095\times\left(\frac{2\times\pi\times1800}{60}\right)^2\times\left(1+\frac{0.095}{0.37}\right)$$

$$=2.25\times0.095\times35530.58\times1.257$$
$$\boldsymbol{F_I=9546.49\,N}$$

➢ **Diameter of Bolts**

$$F_I = \frac{\pi}{4} \times (d_{CB})^2 \times n_s \times \sigma_t$$
$$9546.49 = \frac{\pi}{4} \times (d_{CB})^2 \times 2 \times 60$$
$$(d_{CB})^2 = \frac{9546.49 \times 4}{2 \times \pi \times 60}$$
$$(d_{CB})^2 = 101.29$$
$$d_{CB} = 10.064 \text{ OR}$$
$$\boldsymbol{d_{CB} = 10\ mm}$$

The nominal diameter (d) of the bolt is calculated by,

$$d = \frac{d_{CB}}{0.8} = \frac{10}{0.8}$$
$$d = 11.94\ mm \text{ OR}$$
$$\boldsymbol{d = 12\ mm}$$

3.4. Thickness of the Big End Cap

The rod cap is subjected to inertia force (F_I) as mentioned in figure 4 (b).

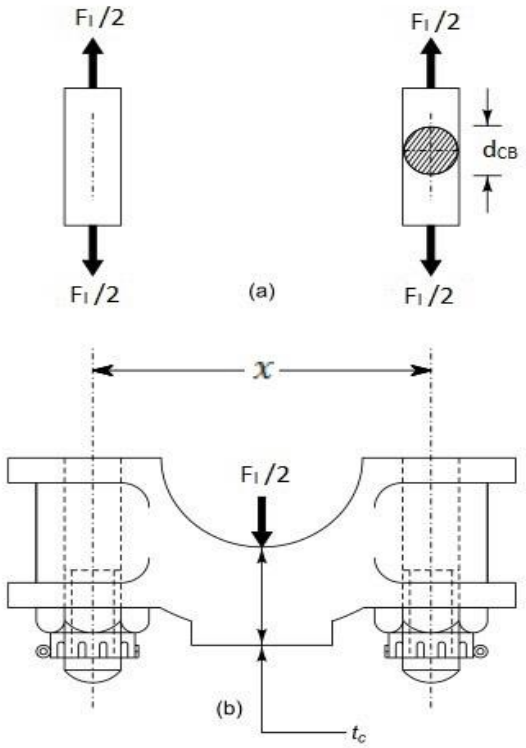


Fig -4: Forces acting on the Cap and Bolts
Source: Design of machine elements, third edition, page no. 874, fig. 25.18

It is considered as a freely supported beam at the bolt centres and loaded in an intermediate manner between centrally concentrated load and a uniformly distributed load that have a bending moment (Wl/6) [1].

Therefore,

$$M_b = \frac{F_I \times x}{6}$$

let's find x,

$x =$ dia. of crank pin (d_C) $+ 2$ times of thickness of bearing (3mm) $+$ nominal dia. of bolt (d) $+$ Clearance (3mm)

$$x = 44 + 2 \times 3 + 12 + 3$$
$$\boldsymbol{x = 65\ mm}$$

So,

$$M_b = \frac{9546.49 \times 65}{6}$$
$$\boldsymbol{M_b = 103420.3\ N\text{-}mm}$$

$$I = \frac{b_C \times t_C^3}{12} = \frac{57 \times t_C^3}{12}$$
$$I = 4.75\ t_C^3$$

And,

$$y = \frac{t_C}{2}$$
$$y = 0.5\ t_C$$

We know that,

$$Z_C = \frac{I}{y} = \frac{4.75\ t_C^3}{0.5\ t_C}$$
$$\boldsymbol{Z_C = 9.5\ t_C^2}$$

The thickness of cap (t_C) is obtained by,

$$\sigma_b = \frac{M_b}{Z_C}$$
$$80 = \frac{103420.3}{9.5\ t_C^2}$$
$$t_C^2 = \frac{103420.3}{9.5 \times 80}$$
$$t_C^2 = 136.08$$
$$t_C = 11.67\ mm \text{ OR}$$
$$\boldsymbol{t_C = 12\ mm}$$

4. PROPERTIES OF MATERIALS

4.1. Aluminium

Aluminium is obtained from mineral bauxite from which it is extracted by Bayer's process. Aluminium material can be used for manufacturing of connecting rod due to its light weight (approx. 1/3 of steel) preferred over steel material and also has high shocking absorbing capacity. It possesses superior strength ratio than steel. Aluminium alloys like aluminium 6061, aluminium 7075 can also be used instead of steel. Connecting rods made up of aluminium are also preferred in drag cars because of lightness and cheap coasting than steel.

Aluminium alloy 7075 is composed of 90.0% Al, 5.6% Zn, 2.5% Mg, 0.23% Cr, and 1.6% Cu, though these numbers nominally fluctuate depending upon manufacturing factors. That is relatively lighter than other metals [6,7].

4.2. Steel

Steel is obtained by alloying of iron with other elements like tungsten, chromium, manganese, nickel, cobalt, titanium etc. The strength of steel depends upon the chemical composition of the alloying elements and the percentages of carbon content in it. Comparing the properties of aluminium and steel, steel is less compatible than aluminium in cold regions. Steel also corrosion resistance and the ability of steel to resist corrosion also depends upon the percentage of the alloying elements which are resistant to corrosion. Steel is advantageous for vehicles during driving, manufacturing and can be recycled. Pollution cause by steel products is comparatively low than other materials and hance environmentally effective.

Grade 4340 is one of the alloy steel-grades that name 'Nickel-Chromium-Molybdenum Steel'. Nickel helps to Increases strength & toughness and improves corrosion resistance. Chromium adds hardness in steel, increased toughness and wear resistance. Molybdenum is a major factor which is increases strength and improves resistance to shock and heat. The combination of these materials makes alloy steel 4340. This alloy steel can obtain much deeper harden-ability than 4100 series steel. High strength in heavy sections is possible with maximum toughness and ductility. This material has a high fatigue/tensile ratio and maintains strength, ductility, and toughness at relatively high temperatures. Also has excellent non-distorting properties for alloy steel.

These two materials i.e., Aluminium Alloy 7075 and Alloy Steel 4340 grade have been selected for the study of analysis. The results of the analysis of stress, strain, and total deformation will be obtained using the tools of ANSYS [8,9].

Table -1: Material Properties [10,11]

Properties	Materials	
	Aluminium Alloy 7075	Alloy Steel 4340
Ultimate Tensile Strength	572 MPa	745 MPa
Tensile Yield Strength	503 MPa	470 MPa
Shear Strength	331 MPa	430 MPa
Fatigue Strength	159 MPa	330 MPa
Young's Modulus	71.7 GPa	200 GPa
Shear Modulus	26.9 GPa	80 GPa
Machinability	70 %	50 %
Density	2.81 g/cm^3	7.85 g/cm^3
Poisson's Ratio	0.33	0.29

5. MODELLING OF CONNECTING ROD

PTC CREO 7.0 software has been used for modelling a
connecting rod with the calculated dimensions.

Table -2: Parameters of Connecting Rod

Parameters	Dimensions (in MM)
Outer dia. of big end	Inner dia. of big end + t_C + 2 times of thickness of bearing (3mm) = 62
Inner dia. of big end	44
Length of big end	57
Length of connecting rod	380
Outer dia. of small end	Inner dia. of small end + t_C + 2 times of thickness of bush (2mm) = 45
Inner dia. of small end	29
Length of small end	58
Width of middle section	28
Height of middle section	35
Height of middle section at big end	42
Height of middle section at small end	30
Diameter of bolt	12
Length of bolt	70
Radius of bolt's head	12
Height of bolt's head	10

Radius of nut	12
Height of nut	10

5.1. Connecting Rod

Step 1. Started PTC CREO 7.0 Software.

Step 2. Clicked on "New" then NEW dialog box open.

Step 3. Selected the PART option in Type and SOLID option in Sub-Type for creating a new model.

Step 4. To specify the sketch plane in the Model Tree, selected the datum plane TOP.

Step 5. On the Model Tab, clicked sketch in the Shapes group. Then opened an Extrude or Sketch tab for creating a sketch on a planar reference.

Step 6. Selected the CIRCLE option and drew a circle (Circle 1). (Inner dia. of big end – 44).

Step 7. Again, selected the CIRCLE option and drew a circle (Circle 2) for same centre as that of previous circle. (Outer dia. of big end – 62).

Step 8. Selected the LINE option and drew a line in the middle of the circles and also drew bolted section and then selected the DELETE SEGMENT option and trimmed the left half circle and unwanted lines of the sketch.

Step 9. Selected the EXTRUDE option and selected the remaining sketch and extruded it symmetrically. (Length of big end – 57).

Step 10. Selected the datum plane RIGHT and drew a circle on the bolt portion and then extruded it symmetrically. Then clicked on the REMOVE MATERIAL option to create a bolt's hole. (Bolt hole dia. – 12).

Step 11. Selected the previous created bolt hole in MODEL TREE and clicked the MIRROR option for creating to be the same hole on other side by selecting the FRONT plane.

Step 12. Again, selected the datum plane TOP for drew the remaining portion. Then selected the CENTERLINE option and drew a centerline from the center of the previous circle.

Step 13. Again, selected the CIRCLE option and drew a circle (Circle 3) at distance (380) from the previous circles at the same centerline. (Inner dia. of small end – 29).

Step 14. Drew another circle (Circle 4) on the centre of previous circle (Circle 3). (Outer dia. of small end – 45).

Step 15. Selected the EXTRUDE option and selected circle 3 and circle 4 and extruded it symmetrically. (Length of small end – 58).

Step 16. Similarly, again selected the datum plane TOP for drew the Rod portion of the connecting rod.

Step 17. Selected the LINE option and drew two parallel lines on both sides of the centerline and connected to circles with help of the REFERENCES option. Then drew middle section of connecting rod. (Height of middle section at big end and at small end – 42 & 30).

Step 18. Selected the DELETE SEGMENT option and trimmed the unwanted lines.

Step 19. Selected the EXTRUDE option and extruded the Rod portion symmetrically. (Width of middle section – 28).

Step 20. Selected the ROUND option to give the shapes and finishing on the edges of the connecting rod.

Step 21. Saved the model as a name "Connecting Rod" using the SAVE option.

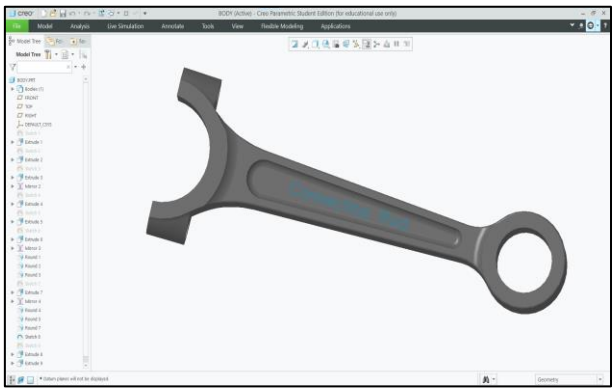


Fig -5: Connecting Rod

5.2. Rod Cap

Step 1. Clicked on "New" then NEW dialog box open.

Step 2. Selected the PART option in Type and SOLID option in Sub-Type for creating a new model.

Step 3. To specify the sketch plane in the Model Tree, selected the datum plane TOP.

Step 4. On the Model Tab, clicked sketch in the Shapes group. Then opened an Extrude or Sketch tab for creating a sketch on a planar reference.

Step 5. Selected the CIRCLE option and drew a circle (Circle 1). (Inner dia. of big end – 44).

Step 6. Again, selected the CIRCLE option and drew a circle (Circle 2) for same centre as that of previous circle. (Outer dia. of big end – 62).

Step 7. Selected the LINE option and drew a line in the middle of the circles and also drew bolted section and then selected the DELETE SEGMENT option and trimmed the right half circle and unwanted lines of the sketch.

Step 8. Selected the EXTRUDE option and selected the remaining sketch and extruded it symmetrically. (Length of big end – 57).

Step 9. Selected the datum plane RIGHT and drew a circle on the bolt portion and then extruded it symmetrically. Then clicked on the REMOVE MATERIAL option to create a bolt's hole. (Bolt hole dia. – 12).

Step 10. Selected the previous created bolt hole in MODEL TREE and clicked the MIRROR option for creating to be the same hole on other side by selecting the FRONT plane.

Step 11. Selected the ROUND option to give the shapes and finishing on the edges of the rod cap.

Step 12. Saved the model as a name "Rod Cap" using the SAVE option.

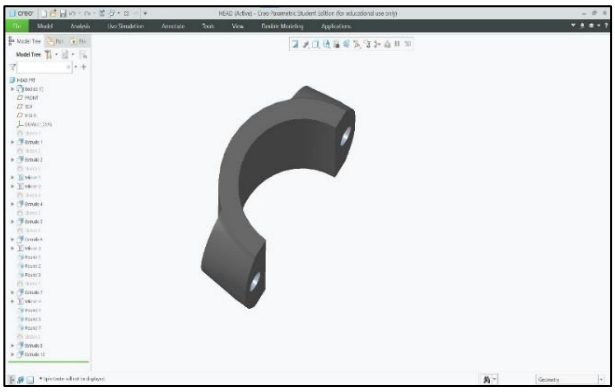


Fig -6: Rod Cap

5.3. Bolt

Step 1. Clicked on “New” then NEW dialog box open.

Step 2. Selected the PART option in Type and SOLID option in Sub-Type for creating a new model.

Step 3. To specify the sketch plane in the Model Tree, selected the datum plane RIGHT. Which will make it easy to assembly.

Step 4. On the Model Tab, clicked sketch in the Shapes group. Then opened an Extrude or Sketch tab for creating a sketch on a planar reference.

Step 5. Selected the CIRCLE option and drew a circle (Circle 1). (Circle dia. – 12).

Step 6. Selected the EXTRUDE option and extruded the circle symmetrically. (Length of the bolt – 70).

Step 7. Selected the extruded surface at the right end and clicked on the SKETCH to draw the bolt's head.

Step 8. Imported pre-built "Octagon" shape using PALETTE and dragged and dropped at the centre of the extruded surface and gave the radius. (Bolt's head radius – 12)

Step 9. Selected the EXTRUDE option and extruded the Octagon sketch towards the bolt. (Height of bolt’s head – 10)

Step 10. Created outside thread on the bolt surface by using the HELICAL SWEEP option.

Step 11. Selected the ROUND option to give the shapes and finishing on the edges of the bolt.

Step 12. Saved the model as a name "Bolt" using the SAVE option.

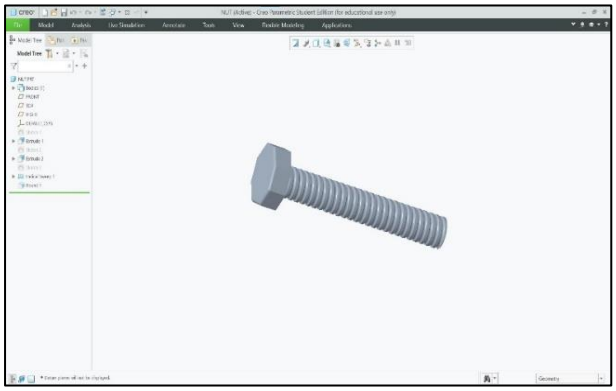


Fig -7: Bolt

5.4. Nut

Step 1. Clicked on “New” then NEW dialog box open.

Step 2. Selected the PART option in Type and SOLID option in Sub-Type for creating a new model.

Step 3. To specify the sketch plane in the Model Tree, selected the datum plane RIGHT.

Step 4. On the Model Tab, clicked sketch in the Shapes group. Then opened an Extrude or Sketch tab for creating a sketch on a planar reference.

Step 5. Imported pre-built "Octagon" shape using PALETTE and dragged and dropped to

Graphics window and gave the radius. (Nut radius – 12)

Step 6. Selected the EXTRUDE option and extruded the Octagon sketch. (Height of Nut – 10)

Step 7. Selected the extruded surface at the right end and clicked on the SKETCH to draw a circle. (Dia. of circle – 12)

Step 8. Selected the EXTRUDE option and extruded the circle sketch symmetrically throughout the Nut, and created a hole inside the Nut by using the REMOVE MATERIAL option.

Step 9. Created inside thread in Nut by using the HELICAL SWEEP option.

Step 10. Selected the ROUND option to give the shapes and finishing on the edges of the Nut.

Step 11. Saved the model as a name "Nut" using the SAVE option.

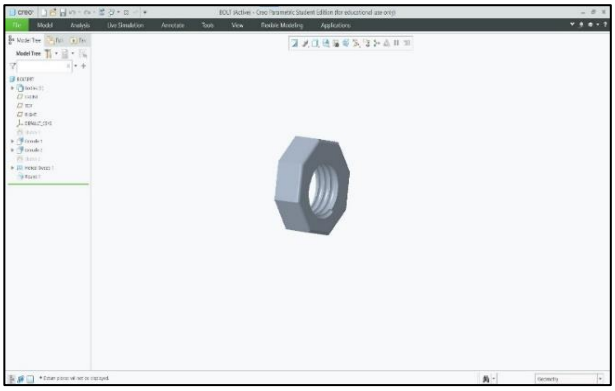


Fig -8: Nut

6. ASSEMBLY

Step 1. Clicked on “New” then NEW dialog box open.

Step 2. Selected the ASSEMBLY under Type and DESIGN under Sub-Type for assembly the connecting rod.

Step 3. In Model Tab, Selected the ASSEMBLY option from the Component Group, then the OPEN dialog box was open.

Step 4. Imported the “Connecting Rod” part from the saved file onto the graphics window and then selected DEFAULT CONSTRAINT.

Step 5. Again, Imported the "Rod Cap" part from the saved file onto the graphics window and then coincided by selecting the front planes of both parts. After that, selected the Connecting Rod's surface and then the Rod Cap's surface for matching coincident. Its assembly completed now.

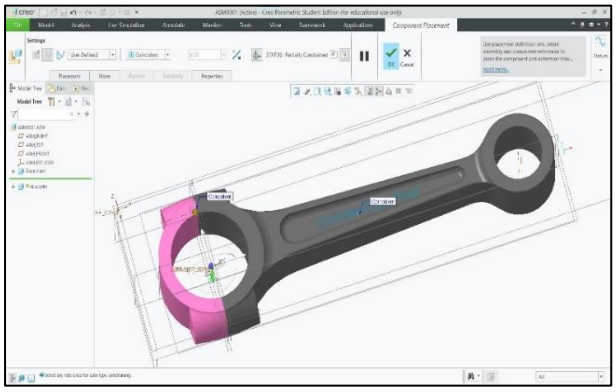


Fig -9: Assembly of Cap through matching coincident

Step 6. Imported its "Bolt" part then selected the centerlines of Bolt and Bolt's hole then it became align, after that selected the Rod Cap's surface and then Bolt's inside surface for matching coincident.

Step 7. Now, imported its "Nut" part then selected the centerlines of Bolt & Nut and it became align, after that selected the Connecting Rod's surface and then Nut's surface for matching coincident. One side Bolt assembly completed now.

Step 8. Similarly, assembled its other side's Bolt and Nut. (Repeat steps - 6 and 7)

Step 9. Saved the model as a name "Connecting Rod" using the SAVE option.

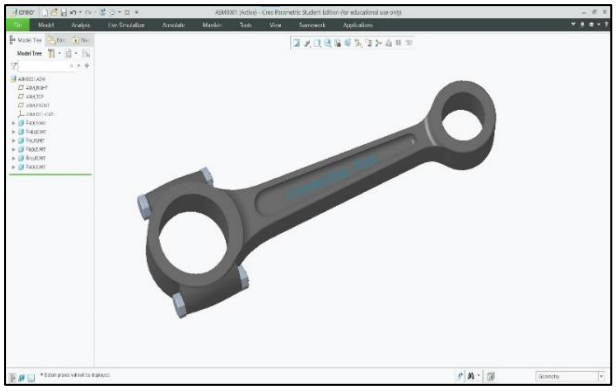


Fig -10: Assembled of Connecting Rod in PTC Creo

Step 10. Selected the SAVE AS option and save the model in ACIS FILE format for further analysis on ANSYS Software.

7. SIMULATION

Step 1. Started Ansys Workbench.

Step 2. Double clicked on Static Structural in the toolbox window to add it to the Project Schematic window.

Step 3. Renamed the Static Structural analysis system as "Connecting Rod".

Step 4. In the connecting rod analysis system, we double clicked the "Geometry" cell; the "Design Modeler" window along with the ANSYS Workbench dialog box is displayed.

Step 5. Set the units in milimetres and imported the external geometry file of the connecting rod which designed in PTC Creo 7.0 software (As per the steps mentioned above).

Step 6. Exit the "Design Modeler" window.

Step 7. Then we defined "Engineering Data". In this section material for the current project is assigned. In ANSYS by default structural steel is selected. To assign new material we turned on the engineering data source. Then we generated new library for connecting rod and made a tick mark on the new generated library named connecting rod to convert it into edit mode. After this, we added different two materials which are kept under study. After giving all required properties for connecting rod we saved them into the Library. In the same way we saved the information about all consideration and then added all the two materials in the content by clicking on the plus sign.

Step 8. After importing the external geometry, we meshed the connecting rod. Triangle (default) meshing has been done for quick result. The numbers of nodes are 41141 and the numbers of elements are 22114.

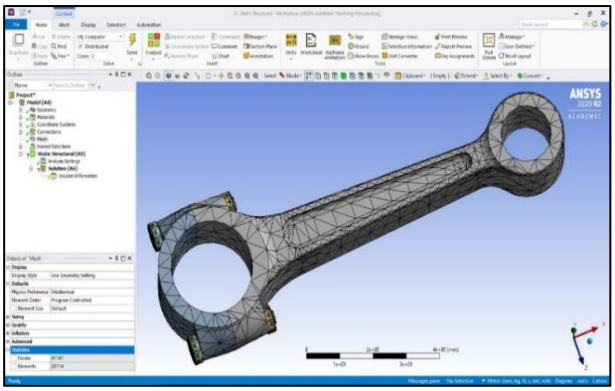


Fig -11: Mesh Generation

Step 9. After Meshing we double clicked the "Setup" cell, where we provided fixed support and Load (Force) on connecting rod. A fixed support has applied to the big end and a bearing load of 24740N has applied to the small end of the connecting rod for both materials condition.

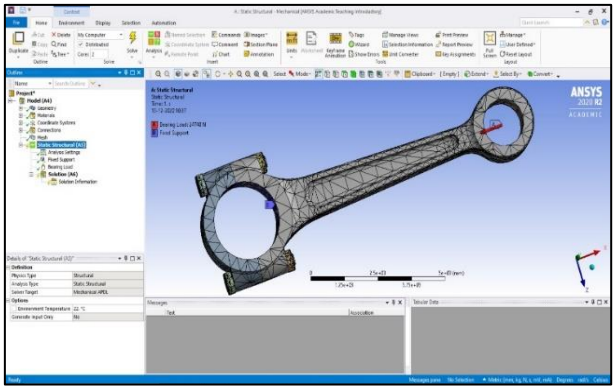


Fig -12: Applied bearing load and fixed support to the connecting rod

Step 10. After providing force and fixed support to the connecting rod, we double clicked the "Solution" cell; where the solutions to be obtained from ANSYS Workbench were found.

Step 11. Then the various results such as equivalent stress, equivalent strain, and deformation were obtained.

8. RESULTS

8.1. Aluminium Alloy 7075

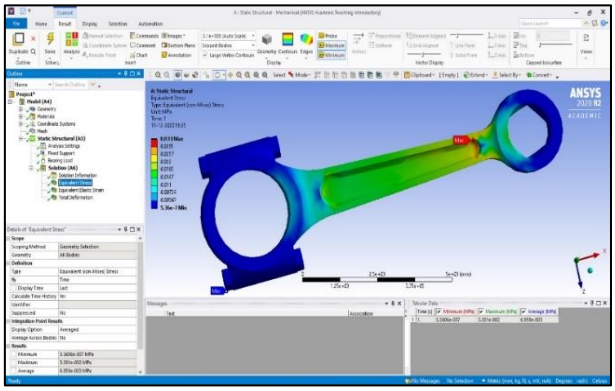


Fig -13: Equivalent (Von-Mises) Stress

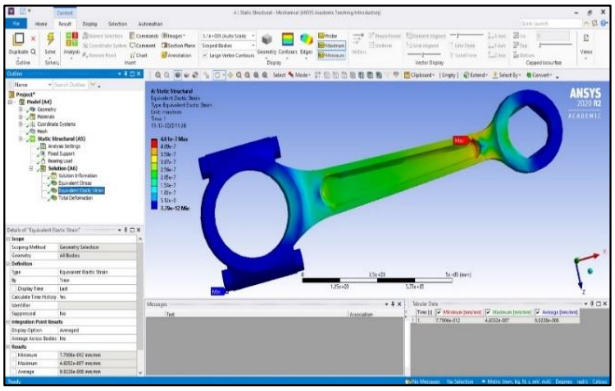


Fig -14: Equivalent (Von-Mises) Elastic Strain

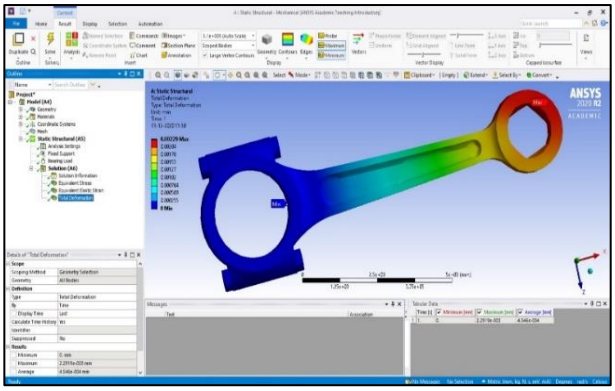


Fig -15: Total Deformation

Table -3: Result of Aluminium Alloy 7075

Load applied = 24740 N		
Name	**Maximum**	**Minimum**
Stress	0.0330 MPa	5.36×10^{-7} MPa
Elastic Strain	4.61×10^{-7}	7.79×10^{-12}
Total Deformation	0.00229 mm	0

8.2. Alloy Steel 4340

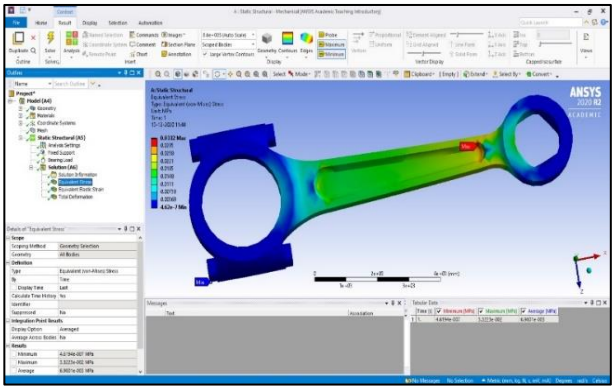


Fig -16: Equivalent (Von-Mises) Stress

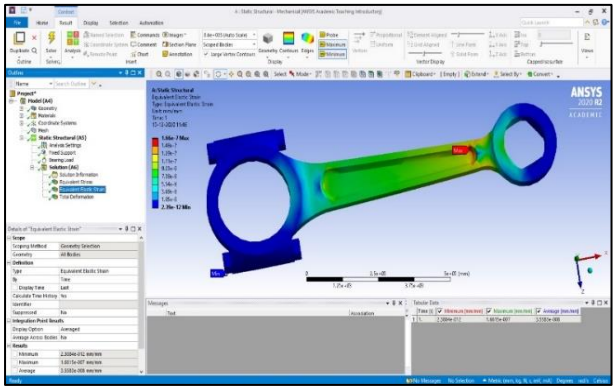


Fig -17: Equivalent (Von-Mises) Elastic Strain

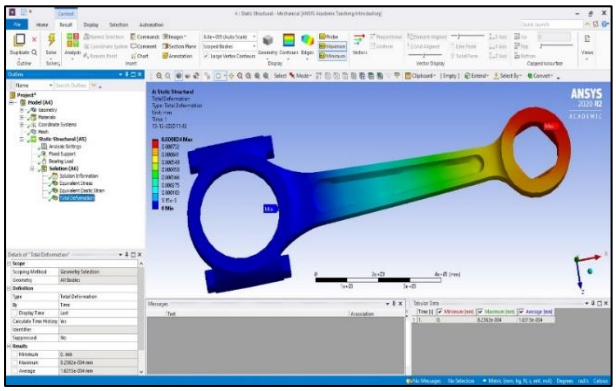


Fig -18: Total Deformation

Table 4: Result of Alloy Steel 4340

Load applied = 24740 N		
Name	**Maximum**	**Minimum**
Stress	0.0332 MPa	4.36×10^{-7} MPa
Elastic Strain	1.66×10^{-7}	2.39×10^{-12}
Total Deformation	0.000824 mm	0

9. CONCLUSION

A connecting rod is subjected to large and repetitive axial as well as compressive load by the gas pressure on the piston, due to this the failure takes place in the connecting rod. Most commonly damage occurring in the connecting rod is deformation. The results of the simulation of connecting rod for two different materials is shown [12]:

Table -5: Comparison of Simulation Results

Material	Stress (MPa)		Strain		Total Deformation (mm)	
	Max	Min	Max	Min	Max	Min
Aluminium Alloy 7075	0.0330	5.36×10^{-7}	4.61×10^{-7}	7.79×10^{-12}	0.00229	0
Alloy Steel 4340	0.0332	4.36×10^{-7}	1.66×10^{-7}	2.39×10^{-12}	0.000824	0

The connecting rod made of alloy steel 4340 has a slightly higher maximum stress of 0.0332MPa as compared to an aluminium alloy 7075 that is 0.0330MPa. The connecting rod made of Aluminium Alloy 7075 has a higher maximum deformation of 0.00229mm as compared to an alloy steel 4340 that is 0.000824mm. In addition, the connecting rod made of aluminium alloy 7075 has a higher maximum strain of 4.61 x 10-7 as compared to an alloy steel 4340 that is 1.66 x 10-7.

We have noticed from the above discussion that if the stress is the main factor in the material, then I will prefer Aluminium Alloy 7075 because in this material stress is produced slightly less as compared to the alloy steel 4340. If we have neglect stress and the main factor is deformation then alloy steel 4340 is the best material for the connecting rod. In addition, if slightly stresses difference is neglected, and assume stress in alloy steel 4340 is almost equal to the stress in aluminium alloy 7075, then the overall comparison shows that alloy steel 4340 is the best material for connecting rod [13,14].

10. FUTURE SCOPE

i. The whole analysis can be repeated with more material used for connecting rods.
ii. The vibration analysis of connecting rod can be helpful for further study of the vibrations produced in the engine.
iii. Transient structure can be done instead of the static structure of the connecting rod.
iv. Determine dynamic analysis and buckling load of the connecting rod.

REFERENCES

[1] Bhandari, V. B. (2012). *Design of Machine Element* (Vol. 3). New Delhi: Tata McGraw Hill Education Private Limited.
[2] Wiki (2020, November 7). Connecting Rod.
[3] Wiki (2020, September 30). PTC Creo.
[4] A Muhammad, M. A. (2020). Finite Element Analysis of a connecting rod in ANSYS: An overview. *Materials Science and Engineering.* IOP SCINENCE LTD. doi:https://doi.org/10.1088/1757- 899X/736/2/ 022119
[5] Modelling & Simulation - Quick Guide. (2020). Retrieved November 27, 2020, from tutorialspoint
[6] Ashish Somani, P. G. (2016, February 3). Comparative Study on Different Materials for Connecting Rod. *International Journal of Innovations in Engineering and Technology, 6*(3), 183-187.
[7] Cavallo, C. (2020, December 4). All About 7075 Aluminum (Properties, Strength and Uses). Retrieved December 5, 2020, from THOMAS
[8] METAL supermarkets. (2015, March 23). ALLOY STEEL | THE FOUR TYPES OF STEEL (PART 2). Retrieved December 03, 2020, from metalsupermarkets
[9] ASM Aerospace Specification Metals Inc. (1978). Aerospace Specification Metals Inc. Retrieved December 3, 2020, from aerospacemetals
[10] AZoM. (2020, September 13). AISI 4340 Alloy Steel (UNS G43400). Retrieved December 3, 2020, from AZO MATERIALS
[11] MakeItFrom. (2020, May 30). SAE-AISI 4340 (SNCM439, G43400) Ni-Cr-Mo Steel. Retrieved December 3, 2020, from MakeItFrom
[12] Mr. H D. Nitturkar, M. S. (2020, March). Design and Analysis of Connecting Rod using Different Materials. *International Research Journal of Engineering and Technology, 7*(3), 1011-1017.
[13] Andoko, & Saputro, N. E. (2018). Failure simulation of connecting rods without pistons using finite element method. *MATEC Web of Conferences.* EDP Sciences.
[14] Manpreet Singh, D. P. (2019). Materials Comparison for Crankshaft in ANSYS. *PAIDEUMA JOURNAL OF RESEARCH, 12*(7), 32-39.

ABBREVIATION

I	Moment of inertia of cross-section
k	Radius of gyration
t	Thickness of connecting rod
A	Cross-section area of connecting rod
F	Force acting on piston due to pressure
F_S	Side thrust on the cylinder wall
F_C	Force acting on connecting rod
$\emptyset$	Angle of inclination of connecting rod with line of stroke
θ	Angle of inclination of crank from top dead centre position
F_{CB}	Critical buckling load

σ_c	Compressive yield stress
L	Length of connecting rod
k_{xx}	Radius of gyration about XX axis
d_C	Diameter of the crank pin
l_C	Length of the crank pin
$(P_b)_C$	Allowable bearing pressure for the crank pin bush
d_P	Diameter of the piston pin
l_P	Length of the piston pin
$(P_b)_P$	Allowable bearing pressure for the piston pin bush
F_I	Inertia force on the cap or bolts
m_r	Mass of reciprocating parts
r	Crank radius
ω	Angular speed of the engine
n	Ratio of length of the connecting rod to the crank radius
l	Length of stroke
d_{CB}	Core diameter of bolts
n_s	Number of bolts
σ_t	Permissible tensile stress for bolt
d	Nominal diameter of bolts
M_b	Bending moment acting on cap
x	Span length or distance between the bolt centres
b_C	Width of cap; equal to the length of big end bearing (l_C)
t_C	Thickness of big end cap
Z_C	Section modulus
σ_b	Whipping stress

BIOGRAPHIES

SURAJ KUMAR,
B.Tech in Mechanical Engineering from Delhi Technological University, India. He has experience in the field of LPG bottling plant at Bharat Petroleum Corporation Limited.

View publication stats

2021 2nd International Conference for Emerging Technology (INCET)
Belgaum, India. May 21-23, 2021

Design and Study of Dadda Multiplier by using 4:2 Compressors and Parallel Prefix Adders for VLSI Circuit Designs

Ayush Kumar
Department of ECE
Delhi Technological University
New Delhi, India
ayushkumardtu@gmail.com

Ankit Kumar
Department of ECE
Delhi Technological University
New Delhi, India
akankitkmr53@gmail.com

Dr. Deva Nand
Department of ECE
Delhi Technological University
New Delhi, India
devkamboj@dce.ac.in

2021 2nd International Conference for Emerging Technology (INCET) | 978-1-7281-7029-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/INCET51464.2021.9456283

***Abstract*—In this paper, a new Dadda Multiplier is designed using 4:2 compressors and parallel prefix adders. From study and analysis in the literature reported in the recent past it is well known that a significant functional component in any processing module is a multiplier. In the processor, the multiplier takes up the majority of the hardware consumed as compared to any other arithmetic operations. Among the existing multiplication methods, Dadda multiplication method is advantageous in terms of reduction in delay.**

The main aim of a quality multiplier is to be as fast and consume as low area as possible. To decrease the delay of Dadda multiplier we have employed 4:2 compressor in the reduction stage. In the proposed system, parallel prefix adders are used to add the final stage of the partial product. In this paper, 4 Dadda multipliers employing 4:2 compressor block are proposed using Sklansky adder (SA), Kogge-Stone adder (KSA), Brent-Kung adder (BKA), and Ladner-Fischer adder (LFA). The proposed multiplier designs and conventional design are simulated on Xilinx Vivado 2018.2. The proposed designs are analyzed concerning conventional multiplier structure in terms of parameters delay (ns) and area (No. of LUTs).

Keywords— Sklansky Adder, Dadda, Kogge-Stone Adder, Parallel Prefix Adder, Brent-Kung Adder, Ladner-Fischer Adder

I. INTRODUCTION

In DSP (Digital Signal Processing) various essential functions are performed in which multiplication is the most important one from all of the other fundamental functions like arithmetic addition and subtraction, these operations can be performed much easily as compared to the requirements for multiplication. It has been observed that the delay in DSP (Digital Signal Processing) system majorly depends on the delay time of the multiplier circuit. The demand has been increasing for High-speed processors as a result of expanding computers and signal processing applications causes a need for high-speed multiplier.

In designing a multiplier, low power consumption is also a crucial issue, so in order to reduce significant power consumption, it is better to minimize the number of operations thereby decreasing dynamic power which is an important part of total power consumption so the demand for low power and high-speed multiplier has increased.

At present, technology is moving along very quickly in a small period of time. The circuits which are manufactured consist of large number of components with high speed, low area, and also low power consumption. Hence, In the design of a circuit, area, speed, and power plays a crucial role [1], [2]. Which leads to the demand of a multiplier circuit which meets these current trends.

In the literature survey of multiplier design, we have gone through different multiplication algorithms like Array multiplier, Binary multiplier, Wallace Tree multiplier, Dadda multiplier.[3] Dadda multiplier algorithm is dominant in different types of multipliers [4]. Dadda multiplier works as same as Wallace Tree Multiplier [5] but it has been observed that Dadda Multiplier is a bit faster in nature. Another advantage of Dadda Multiplier is that the number of components are less as compared to Wallace tree structure.[4]

Prefix adders are derived from parallel prefix circuit concept which comes up with a strong theoretical basis for a broad range of design trade-offs between area, delay and writing complexity.[9]

In the proposed design of Dadda multiplier, the operation of the multiplier is enhanced by employing the 4:2 compressor, due to which the number of levels in partial product reduction stage gets minimized.

In this paper, a different design of Dadda multiplier is proposed in which parallel prefix adders are employed to calculate the final two rows of partial products which are made after reduction from reduction tree stage with the help of 4:2 compressors, full adders and half adders to produce final product terms. All the PPAs are based on carry look-ahead adder theory of generating and propagating carry bits.[10] Kogge Stone adder[11], Sklansky adder[12], Brent-Kung adder[13], Ladner-Fischer adder [14]are some of the adders utilized in the design of Dadda multiplier.

This paper is arranged into various sections. Section I is the Introduction. Section II throws light on the comprehensive functionality of the Dadda multiplier and gives brief information about 4:2 compressor. Section III contains numerous subdivisions which express the in-depth functionality of different parallel prefix adders. In section IV the outcome analysis is distinctly described. Section V consists of the conclusion and future scope of the proposed multipliers in this paper.

II. DADDA MULTIPLIER

Dadda algorithm was invented in the year 1965 by Computer scientist Luigi Dadda.[4] Dadda Algorithm strives

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:32:03 UTC from IEEE Xplore. Restrictions apply.

to reduce the propagation delay and numbers of basic logic gates utilized, unlike Wallace algorithm which only tries to reduces delay as much as possible on each layer. Thus Dadda Multiplier is more efficient.

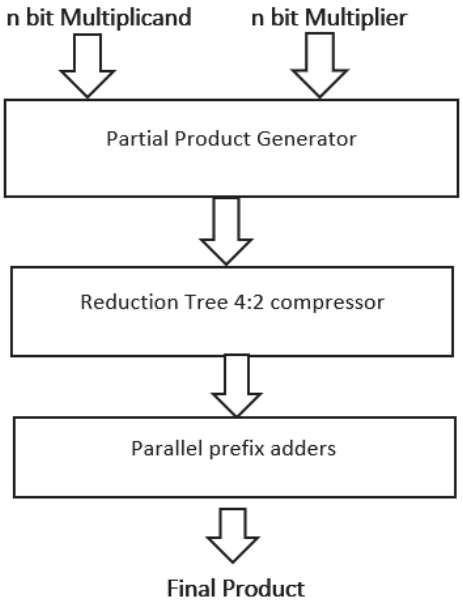


Fig. 1. Block Diagram for Dadda Algorithm

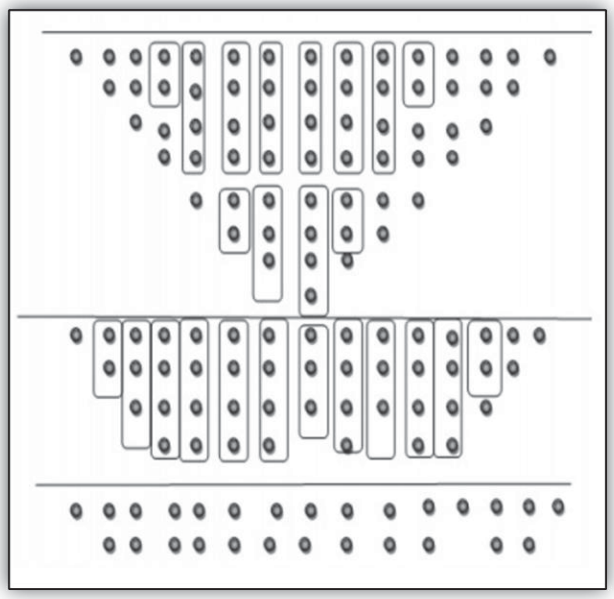


Fig. 2. Dadda Algorithm Dot Diagram[7]

Steps to multiply 2 numbers using Dadda algorithm:

1. Form the partial product by multiplying(and-ing) the multiplicand and multiplier bits by bits, resulting in a matrix known as partial product matrix.
2. The multiplication products of the first stage carry different weights which depends on the position of the multiplied bits.
3. Employing full adders and half adders and 4:2 compressor reduces the number of stages of partial product reduction until two rows are left.

Group the products into 2 numbers and perform addition using parallel prefix adder.

A. 4:2 Compressor:

In multipliers for adding partial product terms, compressors are widely used so as to lessen the number of operands. There are 5 inputs in 4:2 compressor, among which one input is the carry generated in the previous stage and it has 3 outputs which are sum, cout and carry bits.[6] In multiplier when we add many partial products then to reduce the number of operations it is convenient to use 4:2 compressors instead of 3:2 compressors.[8]

Cout = (S1andS2)or(S2andS3)or(S3andS1)

= ((S1xorS2)andS3)or(S1andS2)

Sum = S1xorS2xorS3xorS4xorCin

Carry = (S4xorCin)and(S1xorS2xorS3)or(S4andCin)

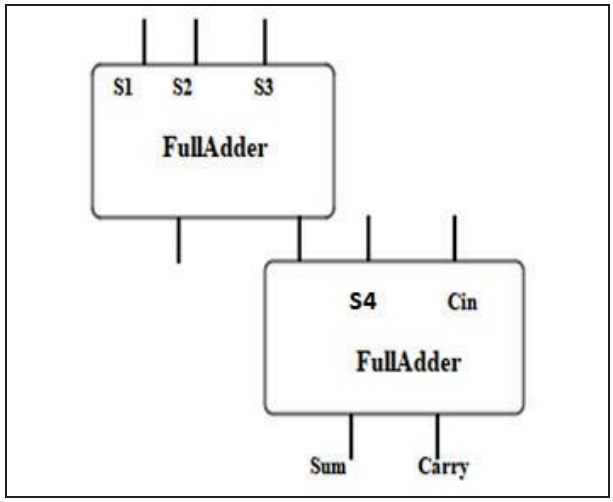


Fig. 3. Traditional Structure of 4:2 Compressor[7]

III. PARALLEL ADDERS

A. Parallel Prefix adder:

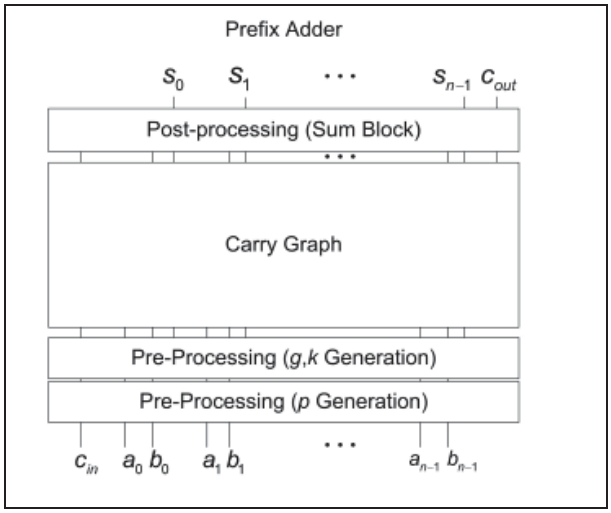


Fig. 4. Prefix Adder[9]

In this paper the proposed Dadda multiplier configuration is improved by employing parallel prefix adders to generate

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:32:03 UTC from IEEE Xplore. Restrictions apply.

the sum of the last two partial products in the final stage addition process to obtain the final product. The main variance between the parallel adders and prefix adders is that the sum and carry generation is performed in the corresponding one-bit block in parallel adder while on the other hand in prefix adder, sum and carry generation is considered as a whole in the carry graph.[9] The carry graph uses the prefix circuit and this lead to the creation of the name "Prefix Adder".

In the pre-processing stage of the parallel prefix adder, the carry generation and carry propagation signals are calculated for the carry graph by each pair of input bits. The generate and propagate equations are mentioned below:

Generate: g = (x and y)

Propagate: p = (x xor y)

After the computation, carries are divided into small packages. These are known as processing components and there following equations are mentioned below:

Pi:k = (Pi:j and Pj-1:k)

Gi:k = (Gi:j or Gj-1:k and Pi:j)

In the post-processing block, Sum and final output carry are calculated:

Si = (Pi xor Ci-1)

Carry graph is important in parallel prefix adders, and it has 3 different components which are mentioned below:

1. Black Cell

2. Gray Cell

3. Buffer

Black cells calculate both carry generate and propagate bits.[15] Grey cells calculate carry generate a signal only. Buffers are used to maintain the loading effect.[15]

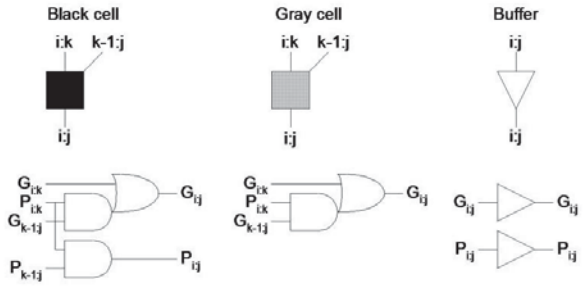


Fig. 5. Carry generation and propagation structures in Parallel Prefix Adders

1) Kogge-Stone Adder:

Kogge stone adder is originated from the logic of carry look ahead adder. It performs a very fast logical addition operation and can be used for arithmetic circuits with top performance. The prime benefit of Kogge Stone adder is it has minimum fanouts. the limitations of this adder are that the operational power is more and it occupies a larger area. The structure of Kogge Stone adder for 16-bits inputs consists of 8 buffers, 15 grey cells and 34 black cells.

2) Sklansky Adder:

The Sklansky adder is based on divide and conquer tree concept. It has a simple prefix structure. In the structure of carry prefix network there is a trade-off between fan out and logic depth, it has high fan out and minimum logic depth. The structure of Sklansky adder for 16-bits inputs consists of 20 buffers, 15 grey cells and 17 black cells.

3) Brent-Kung Adder:

Brent-Kung adder has logic depth in maximum and the number of nodes in minimum. This will also reduce the delay when implemented higher bits adders without compromising the power performance of the adder. It has a higher degree of regularity and has very less wiring crowding which leads to improved performance and requires less chip area to design as compared to Kogge Stone adder. The structure of Brent-Kung adder for 16-bits inputs consists of 20 buffers, 15 grey cells and 12 black cells.

4) Ladner-Fischer Adder:

By hybrid of Sklansky and Brent Kung PPAs, Ladner-Fischer adder is made. This adder design has a large fan-out and has minimum logic depth. The structure of Ladner-Fischer adder for 16-bits inputs consists of 19 buffers, 15 grey cells and 12 black cells.

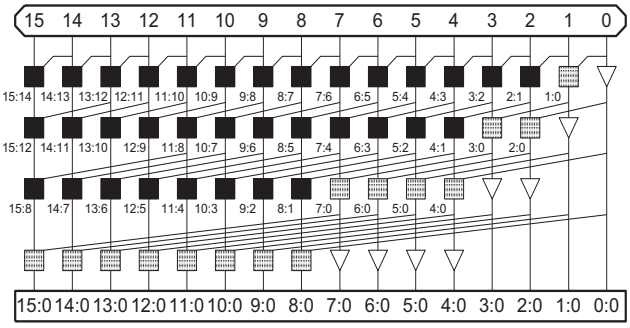


Fig. 6. Kogge-Stone Adder (KSA)

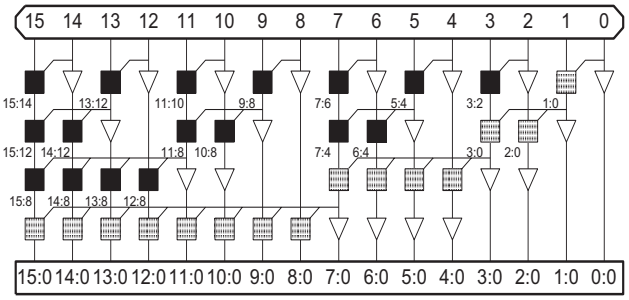


Fig. 7. Sklansky Adder (SA)

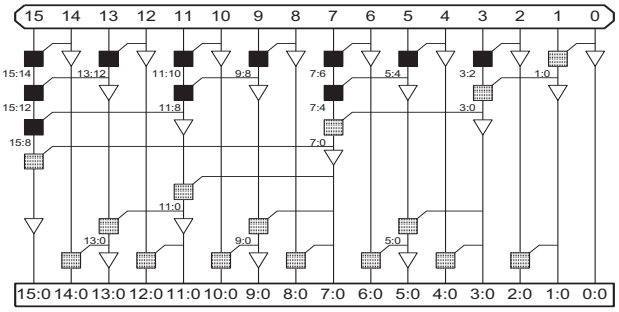


Fig. 8. Brent-Kung Adder (BKA)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:32:03 UTC from IEEE Xplore. Restrictions apply.

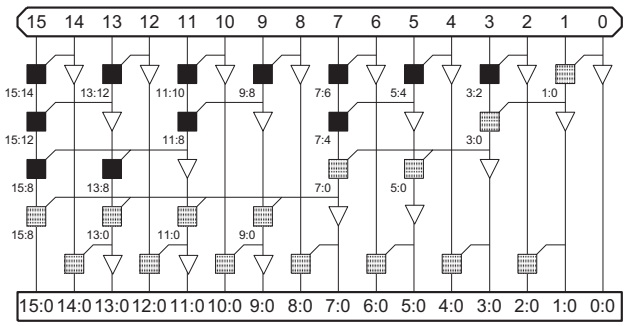


Fig. 9. Ladner-Fischer Adder (LFA)

IV. SIMULATED AND SYNTHESIZED RESULT ANALYSIS

Four Dadda multiplier employing 4:2 compressor structures utilizing four different PPA's (Sklansky Adder, Kogge-Stone adder, Brent-Kung adder and Ladner Fischer adder) are designed for the size of input of 16-bits each in this paper using Verilog HDL. All the multipliers simulation waveforms are shown in Fig10-Fig14. The four proposed multipliers synthesis report results are compared with the synthesis report of Conventional Dadda multiplier using Carry Look Ahead Adder (CLA) in terms of parameter delay in nanoseconds and area (number of LUTs). All the multipliers designed are synthesized and simulated using Xilinx Vivado 2018.2 and selecting Kintex Ultrascale+ xcku15p-ffve1760-1LV-i.

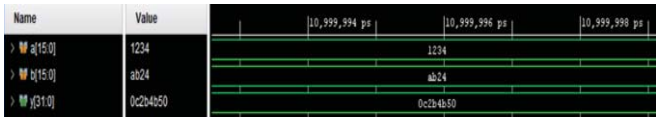


Fig. 10. Waveform simulation of 16-bit Conventional Dadda multiplier employing 4:2 compressor.

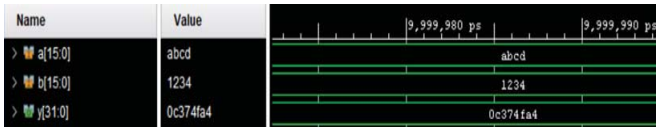


Fig. 11. Waveform simulation of 16-bit Dadda multiplier employing 4:2 compressor and Kogge-Stone Adder.

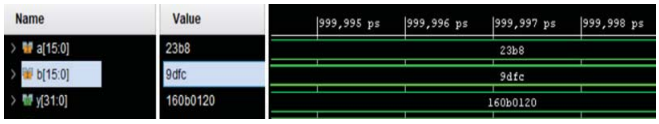


Fig. 12. Waveform simulation of 16-bit Dadda multiplier employing 4:2 compressor and Sklansky Adder.

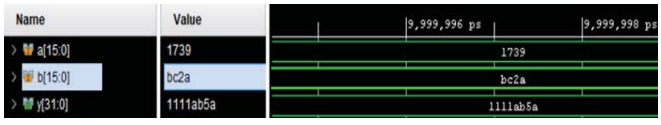


Fig. 13. Waveform simulation of 16-bit Dadda multiplier employing 4:2 compressor and Brent-Kung Adder.

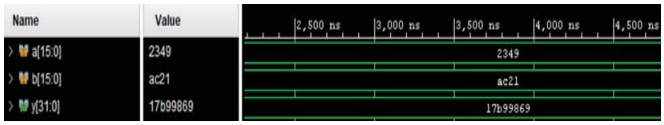


Fig. 14. Waveform simulation of 16-bit Dadda multiplier using 4:2 compressor and Ladner-Fischer Adder.

TABLE I. DELAY(NS) AND AREA (NUMBER OF LUTS) OF DIFFERENT TYPE OF DADDA MULTIPLIER STRUCTURES EMPLOYING 4:2 COMPRESSOR AND DIFFERENT PPAS

Input Size	Type of Multiplier	Delay(ns)	Area (No. of LUT's)
16-bit	Conventional Dadda Multiplier employing 4:2 compressor	9.944	450
	Dadda multiplier employing 4:2 compressor and Kogge Stone adder	8.879	375
	Dadda multiplier employing 4:2 compressor and Sklansky adder	9.207	373
	Dadda multiplier employing 4:2 compressor and Brent Kung adder	9.214	365
	Dadda multiplier employing 4:2 compressor and Ladner Fischer adder	9.201	371

In table 1, a total of five multipliers i.e., four type of proposed multipliers using PPAs and one conventional multiplier utilizing CLA are synthesized and simulated. The synthesis report of each multiplier consists of area details in terms of parameter delay in nanoseconds and the number of LUTs occupied. From table 1, it can be inferred that:

1. Dadda multiplier employing 4:2 compressor and Kogge Stone adder has the least delay, but in terms of area it occupies a greater number of LUTs when compared to other proposed multipliers.
2. The traditional multiplier employing CLA has the maximum delay and number of LUTs occupied.
3. The Dadda multiplier employing 4:2 compressor and Brent Kung adder has the least number of LUTs used but its delay is more than other types of multipliers with PPA's but is less when compared with the traditional structure.
4. The Dadda multiplier employing 4:2 compressor and Ladner Fischer adder lies in between all the other adders compared as its delay and number of LUT's is moderate.

V. CONCLUSION AND FUTURE SCOPE

In this paper, a 16x16 Dadda multiplier employing 4:2 compressor and different PPA's has been realized using Verilog with the help of Xilinx Vivado. The 4:2 compressor improves the performance of the Dadda multiplier during the partial product reduction stage while PPA's help in minimizing the delay in the final product stage. The four-multipliers structure designed in this paper of Dadda multiplier using 4:2 compressor and 4 different PPA's is compared with Conventional Dadda multiplier in terms of parameters like delay and area requirement in term of LUT's. There is a significant improvement in terms of the Delay parameter and number of LUT's used which is observed from table 1. In the multipliers proposed in this paper, Dadda multiplier employing 4:2 compressor and Kogge Stone adder has the least delay i.e., it's the fastest. But the number of LUT's used by it is more as compared to other multiplier structures. While for the area parameter, Dadda multiplier using 4:2 compressor and Brent Kung adder has the least area in terms of the number of LUT's used amongst the other proposed structures.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:32:03 UTC from IEEE Xplore. Restrictions apply.

So, the Dadda multiplier employing 4:2 compressor and Kogge Stone adder circuit can be used for multiplier application which needs to be fast but can cost a little extra area. Whereas for area constricted application Dadda multiplier employing 4:2 compressor and Brent Kung adder can be a good alternative. Finally, if a balanced multiplier application i.e., which needs to be moderately fast but at the same time isn't at the expensive side can use Dadda multiplier using 4:2 compressor and Ladner Fischer adder. In the future scope of this work, compressors like 5:2, 7:2, etc. can be used or the size of input can be increased (i.e., 32-bits, 64-bits).

REFERENCES

[1] N. Weste and D. Harris, CMOS VLSI Design. Reading, MA: Addison Wesley, 2004.

[2] N. Weste and K. Eshragian, Principles of CMOS VLSI Designs: A System Perspective, 2nd ed., Addison-Wesley, 1985-1993.

[3] J. M. Rabaey, Digtal Integrated Circuits—A Design Perspective. Upper Saddle River, NJ Prentice-Hall, 2001.

[4] L. Dadda, "Some schemes for parallel multipliers," Alta Frequenza, Vol. 34, pp. 349-356, May, 1965.

[5] C. S. Wallace, "A suggestion for a fast multiplier," IEEE Trans. Electronic Computers, Vol. 13, pp. 14-17, Feb., 1964.

[6] U. Penchalaiah and S. K. VG, "Design of High-Speed and Energy-Efficient Parallel Prefix Kogge Stone Adder," 2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCA), Pondicherry, 2018, pp. 1-7.

[7] M. V. S. Nandam and S. Alluri, "High Performance 32-bit Dadda Multiplier Using EDA," 2020 IEEE 7th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2020, pp. 1-5, doi: 10.1109/ICSSS49621.2020.9201958.

[8] Rupendra Singh and Neeta Pandey, “A Low Power and Area Efficient Design of Dadda Multiplier by exploring 4:2 Compressors and Brent-Kung Adder,” International Journal of Advanced Science and Technology Vol. 29, No. 3, (2020), pp. 10044 – 10054

[9] Choi, Youngmoon & Jr, Earl. (2005). Parallel Prefix Adder Design with Matrix Representation.90-98. 10.1109/ARITH.2005.35. 17th IEEE Symposium on Computer Arithmetic (ARITH-17 2005), 27-29 June 2005, Cape Cod, MA, USA

[10] Sarnala Butchibabu, Sannikanti Kishore Babu, 2014, Design and Implementation of Efficient Parallel Prefix Adders on FPGA, International Journal of Engineering Research & Technology (IJERT) Volume 03, Issue 07 (July 2014),

[11] P. M. Kogge and H. S. Stone, “A Parallel Algorithm for the Efficient Solution of a General Class of Recurrence Equations,” IEEE Trans. Computers, Vol.C-22, 1973, pp.786 793.

[12] J. Sklansky, “Conditional-sum addition logic,” IRE Transactions on Electronic Computers, vol. 9, pp. 226–231, 1960.

[13] R. P. Brent and H. T. Kung, “A Regular Layout for Parallel Adders,” IEEE Trans. Computers, Vol. C-31, 1982, pp. 260-264.

[14] R.E. Ladner and M.J. Fischer, “Parallel Prefix Computation,” JACM, Vol.27-4, 1980, pp. 831-838.

[15] N Aruna Kumari, M Sai Srinivas, K Aravind, “Implementatio n of 64 Bit Arithmetic Adders,” International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-5S4, February 2019.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:32:03 UTC from IEEE Xplore. Restrictions apply.

2021 2nd International Conference for Emerging Technology (INCET)
Belgaum, India. May 21-23, 2021

Design Of High Bandwidth Circularly Polarised Antipodal Vivaldi Array for 5G Applications

Anirudh Nakra
Dept. of Electronics and Communication
Delhi Technological University
New Delhi, India
anirudhnakra4@gmail.com

Abhijeet Vats
Dept. of Electronics and Communication
Delhi Technological University
New Delhi, India
abhijeetvats01@gmail.com

Asok De
Professor,
Dept. of Electronics and Communication
Delhi Technological University
New Delhi, India
asok.de@gmail.com

2021 2nd International Conference for Emerging Technology (INCET) | 978-1-7281-7029-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/INCET51464.2021.9456438

***Abstract*— In this paper, a circularly polarized antenna array is proposed that has high gain and excellent axial ratio properties. The array is composed of multiple UWB antipodal slotted Vivaldi antennas that have been designed for a resonant frequency of 8.5 GHz and using orthogonal excitations to the array, circular polarization has been successfully achieved. Modified Wilkinson power splitter has been utilized to work over the range of 4-12 GHz and phase shifting has been achieved using aperture coupling techniques. The S parameters for the constructed system are excellent with return loss as low as -25 dB and an axial ratio of almost 0.098 dB at the resonant frequency.**

Keywords—Circular Polarization, Antipodal Vivaldi Antenna, Wilkinson divider, Axial ratio

I. INTRODUCTION

Circularly polarized antennas provide many advantages over a classical arrangement. For one, it allows a greater possibility of a good link between the transmitter and the receiver. It allows good LoS linkages and therefore is an important way forward for the antenna world. Recently with the proliferation of UWB circularly polarized antennas, there has been a rise in research in such areas. Good outputs and axial ratios can be observed over large bandwidths and this has made the demand for an antenna array that much more pivotal.

Some great research has been done on integrating aperture coupling techniques with UWB antennas to create antennas that work better than those using delay line circuits. Narbudowicz et all have employed a fusion system of good isolation power splitters with UWB phase shifters and antipodal Vivaldi antennas.

5G and further generations will benefit from CP antennas due to their extreme resistance to reflection, absorption and other interferences and mismatched polarization. In the paper, we have used an antipodal Vivaldi antenna with slots to maximize the bandwidth while improving isolation between the various

Components using a six stage power divider. We have also compared various different planar antennas such as a rectangular microstrip antenna and Quasi-Yagi antenna before finally landing on the antipodal Vivaldi antenna due to its better bandwidth and antenna efficiency results.

II. DESIGN

We chose the antipodal Vivaldi antenna as our working piece due to its high bandwidth. We achieved a gain of 7.6 dB with an antenna efficiency of 98.7% while also having a practical bandwidth of 7.8 GHz. This all round excellence and relative ease of fabrication made it an obvious choice over the Quasi Yagi antenna which had a slightly better efficiency but compromised on the bandwidth. Furthermore, research on antipodal Vivaldi has come to a point where its characteristics are quite suitable for 5G applications and creating a circularly polarized system would be beneficial for future work.

Two power splitters were tested in the feed network: a 3 stage Wilkinson power splitter and a 6 stage Wilkinson power splitter. The 6 stage Wilkinson power divider designed on a substrate of thickness 0.508 mm and dielectric constant 2.2 and it showed better isolation between the output ports compared to the 3 stage divider.

A Multi Sectional structure with different impedances were used to achieve the operation of the power divider in the required frequency range. The values of the resistances were calculated by the formulas given in [6]. Using the LineCalc tool provided in the schematic interface of Keysight ADS, we were able to define our control parameters to calculate the optimal length and width of the different stages. After calculating the characteristic impedances of the various stages mathematically and using TX line, we calculated, and fine-tuned the dimensions to those that gave us the best results. These widths and lengths were then translated into a 3d model in CST design studio using a delta separation of 1.3mm between the two isolated arms. By analyzing the S parameter curves, we saw that we were able to achieve an isolation greater than 28 dB which was better than the isolation from the splitters in both [3] and [4], two studies we chose as our standard.

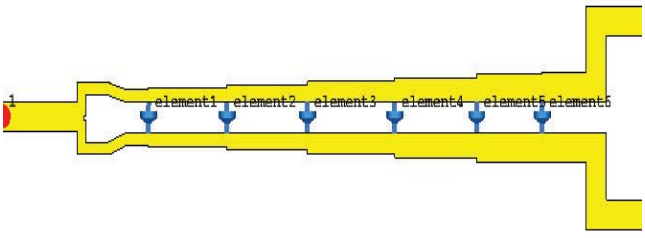


Fig. 1. Wilkinson Power divider

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:32:31 UTC from IEEE Xplore. Restrictions apply.

TABLE I. WILKINSON POWER SPLITTER DIMENSIONAL PARAMETERS AND RESISTOR VALUES

Resistors	Dimensional Parameters of Wilkinson power divider		
	Resistor values	*Width(mm)*	*length*
R1	153	W1=0.42	L1=5.20
R2	166	W2=0.53	L2=5.14
R3	258	W3=0.66	L3=4.99
R4	351	W4=0.75	L4=5.32
R5	299	W5=1.09	L5=4.78
R6	481	W6=1.14	L6=4.27

The phase shifter used in the model is based on broadband aperture coupling technique. The analysis of the Phase shifter is done using the conventional approach done for coupled microstrip line- treating it as a four port device with two of its ports open circuited[2].Previous research has been done on creating microwave phase shifters. However these works have often been non usable on high bandwidth equipment. Edge coupled and Schiffman phase shifters are popular but they suffer from issues such as high fabrication costs and difficulties. The design proposed in [2] has been optimal because of its relative ease to design for own purpose and high bandwidth. Three major steps have been utilized in creating this broadband phase shifter. The first step is to mathematically solve for the even and odd mode characteristic impedances for the proposed ellipses. After solving for them, an elliptical integral of the first kind was calculated for finding out the dimensions of the elliptical slots.

This calculation was complex and has been succinctly described in [2]. The length calculation however is a relatively straightforward process with quarter wave formula being used at the resonant frequency of design. The modelling of these patches is the next major step. Using the calculated dimensions, we cut two nearly circular slots with the size of length 7.9mm and minor axis of 3.2mm in the ground plane. The ground was then sandwiched between two substrates of thickness 0.508mm and dielectric constant 2.2. Above and below the slots, we designed elliptical patches of the same length (7.9 mm) and minor axis 2.41mm.

The two elliptical patches placed below the lower substrate were joined using a microstrip line of characteristic impedance 50 ohms. The topmost patches were joined to the output port and the Wilkinson divider output respectively. The last step we undertook was to optimize the dimensions of both the slots and the elliptical patches to give us a 90 degree phase shift in the lower X band between 8 to 9 GHz. Fine tuning the dimensions was a largely cumbersome process and was solved using trial and error.

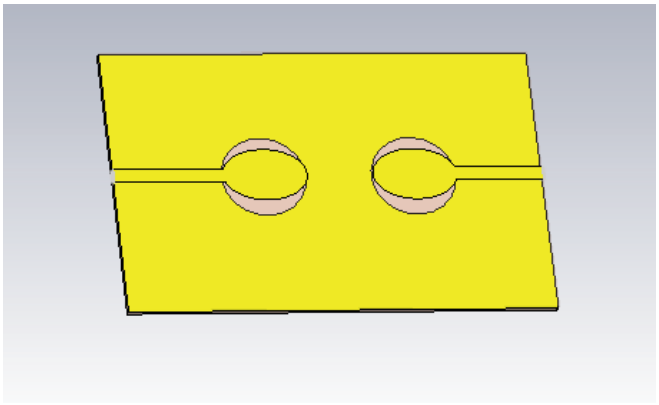


Fig. 2. Phase Shifter

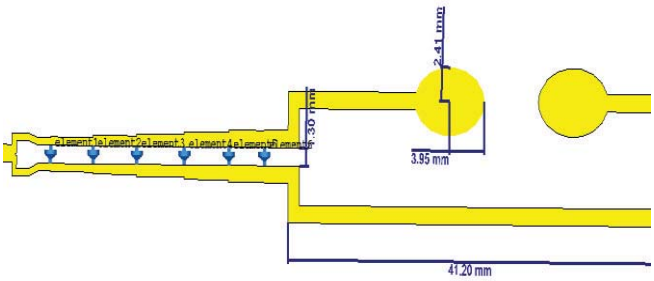


Fig. 3. Wilkinson power splitter and phase shifter combined

The output of the phase shifter and the reference line were fed to the slotted antipodal Vivaldi antenna The Vivaldi antenna was designed using a substrate of dielectric constant 2.98 and thickness 1.5 mm. Using standard design equations in [4], we were able to model the curve as an exponential taper of the type Y=Ae^(Rx). The width and length of the antennas were calculated using the lower limit of the frequency of operation and material parameters like dielectric constant of the substrate, width of the substrate, etc. Our Vivaldi antenna has a width of around 83 mm and a length spanning 91 mm. The microstrip feed line that was tapered was designed for a characteristic impedance of 50 ohms and its width can be calculated easily using software such as Txline or solving equations given in [4].

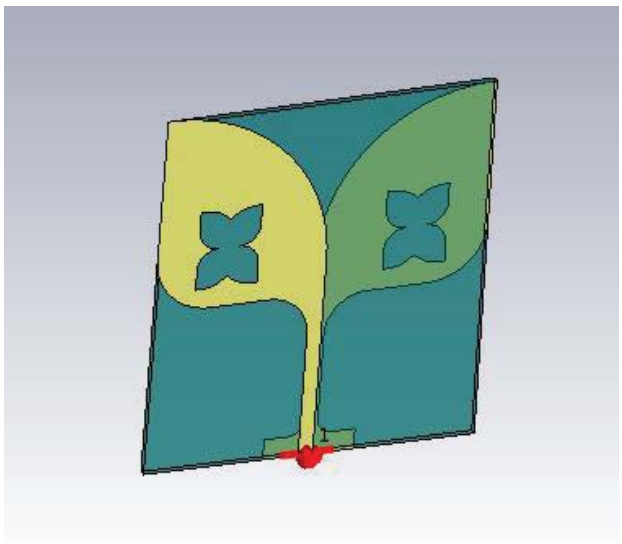


Fig. 4. Antipodal Vivaldi Antenna

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:32:31 UTC from IEEE Xplore. Restrictions apply.

The curves were optimized using a heuristic approach and the changes in the S parameters in CST Studio Suite 2019 were observed to create a broadband and compact antenna design. After the design of the radiators as well as the balun in the Vivaldi antenna, we moved on to the procedure of creating slots to enhance bandwidth. The antenna was already operating with great characteristics at 8.5 GHz but the bandwidth was only 1.2 GHz. Using techniques suggested in [4], we aimed to create floral pattern based slots in the antenna to effectively increase its bandwidth.

The slots introduced in the antenna were constructed by trial and error. The trial and error approach considered the overall impedance of the system. The slots were first designed as rectangular slots by fixing one end and using arbitrary lengths to define the rectangle and then tapering the ends to create a floral structure [4]. To achieve circular polarization, both the antennas must be kept in such a way that their phase centers are close to each other, but factors like size and the space required to prevent coupling counters it. So, for optimization, the antennas are kept at right angles to each other.

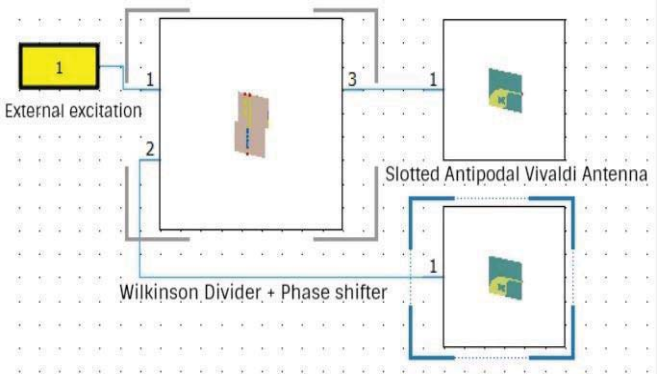


Fig. 5. Schematic of the system in CST design studio

III. RESULTS

At first, we analyzed the results of Wilkinson power divider, the phase shifter and the antipodal Vivaldi antenna individually followed by the analysis of the system. The power divider showed maximally flat response for S21 and S31 while giving great isolation (~30 dB) between the two ports.

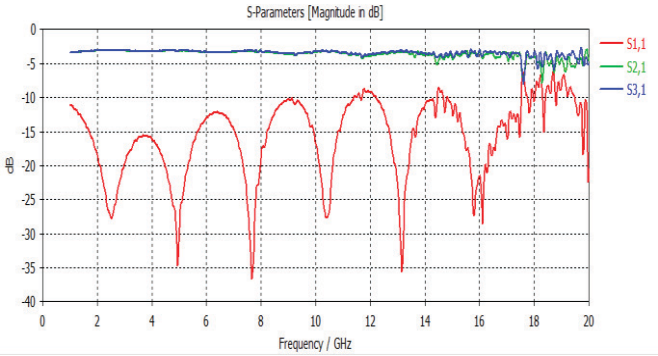


Fig. 6. S11, S21 and S31 of Wilkinson Power splitter

The Phase shifter in the 8-9 GHz band gave 90 degrees phase difference with the reference line. The slotted antipodal Vivaldi antenna showed high bandwidth.

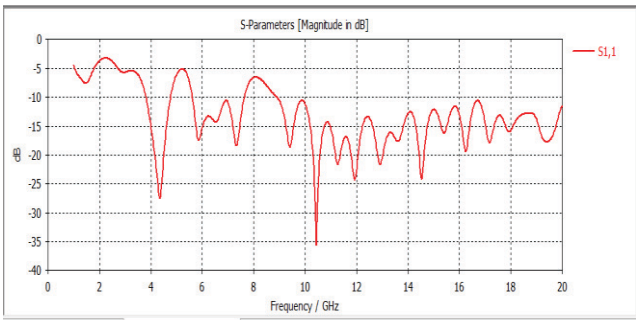


Fig. 7. Return loss of Antipodal Vivaldi Antenna

The combined array's frequency response showed the return loss below 10 db and the axial ratio below 1db for the frequency range of 8-9 GHz. Performance peaked for 8.57 GHz which showed -25 db return loss and an axial ratio of 0.098db.

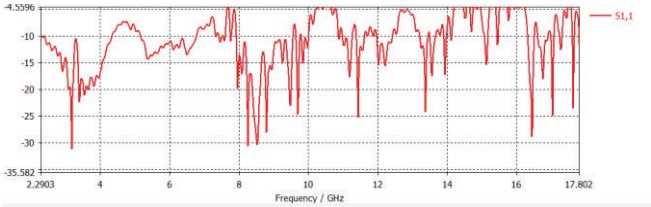


Fig. 8. Return loss of the system

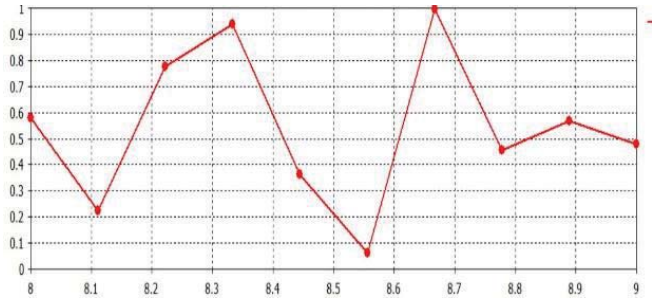


Fig. 9. Axial Ratio v/s Frequency

IV. CONCLUSION

The design of a circularly polarized antipodal Vivaldi antenna array has been presented. This array with CP gives a higher probability of successful link at even higher frequencies than presented before and eliminates the line of sight problem of linear polarization at frequencies above 2.5GHz. Vivaldi antenna in itself has been highly miniaturized and is extremely portable while giving a bandwidth of more than 7 GHz. A high isolation six stage Wilkinson divider has been employed to give maximal isolation of more than 30 dB while not being very loss. An ultra-wideband phase shifter has been designed and optimized to work at low X band frequencies to generate a phase shift of 90 degrees in the feed of the Vivaldi array and the system gives axial ratios (0.098 dB) that are much lower than those in previous literature especially in the target lower X band range.

REFERENCES

[1] Narbudowicz, A., John, M., Bao, X., & Ammann, M. J. (2012). Vivaldi array for generation of UWB circular polarization. Proceedings of the 2012 IEEE International Symposium on Antennas and Propagation. doi:10.1109/aps.2012.6348767

[2] Abbosh, A. M. (2007). Ultra-Wideband Phase Shifters. IEEE Transactions on Microwave Theory and Techniques, 55(9), 1935–1941. doi:10.1109/tmtt.2007.904051

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:32:31 UTC from IEEE Xplore. Restrictions apply.

[3] O. Kizilbey, S. Bozdemir and B. S. Yarman, "2–10 GHz multisection 2-way Wilkinson power divider with enhanced port match and isolation," 2017 IEEE 18th Wireless and Microwave Technology Conference (WAMICON), Cocoa Beach, FL, USA, 2017, pp. 1-4, doi: 10.1109/WAMICON.2017.7930250.

[4] Marek Dvorsky, Harihara S. Ganesh, S. Sadhish Prabhu, "Design and Validation of an Antipodal Vivaldi Antenna with Additional Slots", International Journal of Antennas and Propagation, vol. 2019, Article ID 7472186, 10 pages, 2019. https://doi.org/10.1155/2019/7472186

[5] D. Pozar, Microwave Engineering, 3rd ed. New York: Wiley, 2005.

[6] S. B. Cohn, "A Class of Broadband Three-Port TEM-Mode Hybrids," in IEEE Transactions on Microwave Theory and Techniques, vol. 16, no. 2, pp. 110-116, February 1968, doi: 10.1109/TMTT.1968.1126617.

[7] Guangyou, Fang & Sato, Motoyuki. (2002). Optimization of Vivaldi antenna for deming by GPR.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:32:31 UTC from IEEE Xplore. Restrictions apply.

Detection and Classification of Power Quality Disturbances Using Discrete Wavelet Transform and Neural Network Technique

N.K. Bhagat[1], Shubhechchhu Singh[2]*, Akash Gupta[3], Ankush Tanwar[4]
[1]Assistant Professor, Electrical Engineering, Delhi Technological University, New Delhi, India
[2,3,4]Student, Electrical Engineering, Delhi Technological University, New Delhi, India
Corresponding Author
E-Mail Id: shubhechchhusingh_2k17ee205@dtu.ac.in

ABSTRACT

The nature of electric power and unsettling influences happened in power signal has become a significant issue among the electric power providers and clients. For improving the power quality ceaseless checking of power is required which is being conveyed at customer's locales. Accordingly, recognition of PQ disturbances, and appropriate characterization of PQD is exceptionally attractive. The detection and classification of the PQD in distribution systems are important tasks for protection of power distributed network. The majority of the unsettling influences are non- stationary and transitory in nature thus it requires progressed devices and methods for the investigation of PQ disturbances. In this work a hybrid technique is utilized for portraying PQ disturbances influences utilizing wavelet change and fuzzy logic. A no of PQ events are generated and decomposed using wavelet decomposition algorithm of wavelet transform for accurate detection of disturbances. It is additionally seen that when the PQ disturbances are tainted with noise the detection becomes troublesome and the element vectors to be removed will contain a high level of noise which may debase the characterization accuracy. Consequently, a Wavelet based denoising strategy is proposed in this work before include extraction measure. Two very distinct features common to all PQ disturbances like Energy and Total Harmonic Distortion (THD) are extracted using DWTi are fed as inputs to the fuzzy expert system for accurate detection and classification of various PQ disturbances. The fuzzy expert system orders the PQ disturbances as well as demonstrates whether the disturbance is pure or contains harmonics. A neural network-based Power Quality Disturbance (PQD) location system is likewise demonstrated executing Multilayer Feed forward Neural Network (MFNN).

Keywords: *Power quality, wavelet, DWT classification, ANN*

INTRODUCTION

In this report, we haves used Discrete Wavelet transform for identification/detection of different PQ disturbances as well as pure sinusoidal wave. Discrete Wavelet transform is clearly a better method for detection of these PQ events as wavelet transform can offer time and frequency info of a signal simultaneously. Here db4 mother wavelet and decomposition levels equals 4 and Extracted Feature Vector like skewness, standard deviation,[6] kurtosis are used to reduce large matrix into manageable group for detection and later classification of PQ events. Multi-resolution signal decomposition (MSD) technique was used to decompose the signal into lower resolution.[5]

Power Quality

The term power quality aims at maintaining Unity power factor always in a system. It aims at maintaining voltage or current as pure sinusoidal wave with 1 pu magnitude at 1 pu frequency and phase shift of 120 degree in a three-phase system of voltages. Sometime due to the use of non-linear

 All Rights Reserved

Loads or faults in power system few disturbances such as voltage sag, voltage swell, interruption, harmonics and transients may occur. These disturbances can damage electronic devices connected to the power system.

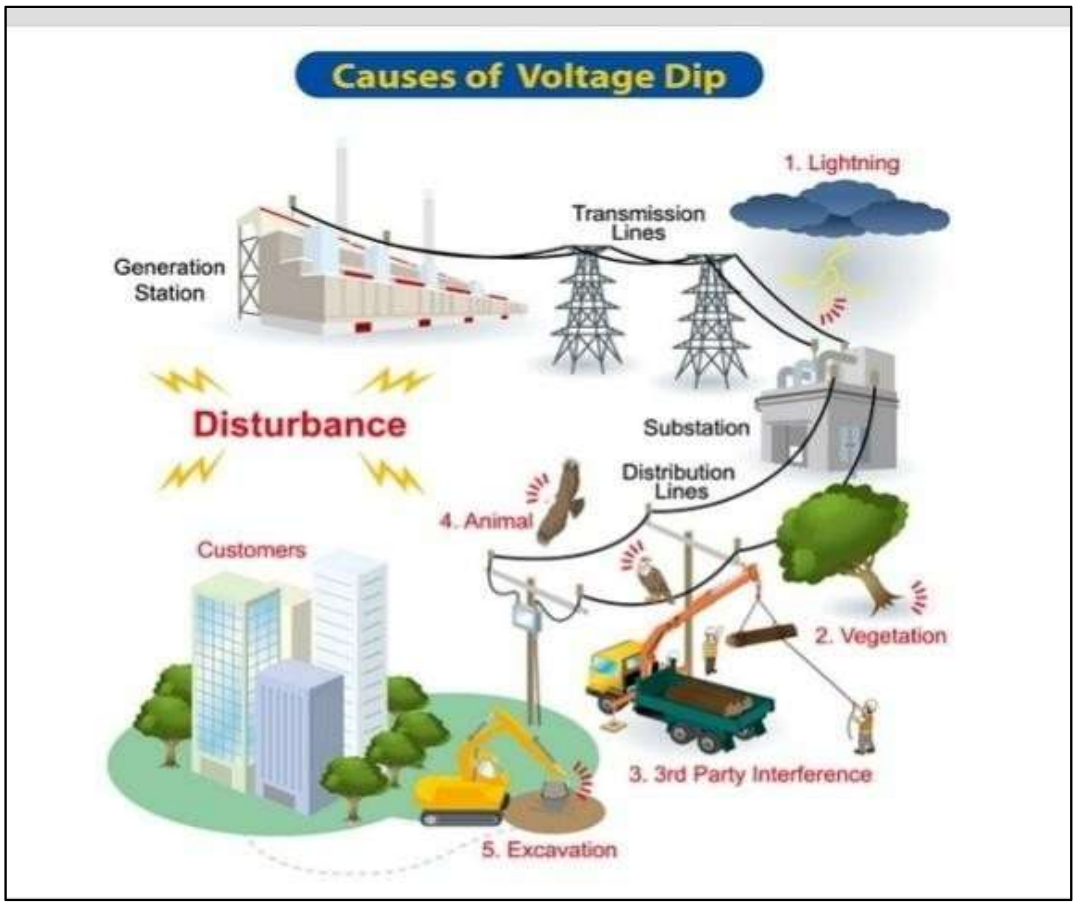


Fig. 1: *Power System Diagram Showing Various Causes of Voltage Dip.*

Types of electric power quality

Transients

- This is a short duration phenomenon.
- The causes are sudden load, sudden switching on of large loads.
- Transients present in the system for a very short duration up to some cycles of the input voltage.
- This will lead to increase or decrease in magnitude of voltage or current.
- There are two sorts of transients: impulsive and oscillatory.

Short and long duration voltage variation[3]

- The power quality disturbances in the voltage that stays for less than 60 seconds fall under the category of short duration. It is classified as sag, swell and voltage fluctuation.
- Power quality disturbances in the voltage that stays for the duration of more than 60 seconds fall under this category of short duration variation. It is of three types under voltage over voltage, voltage fluctuation.
- Voltage sag: It can be defined as a sudden dip in the voltage for a short duration of time.
- Voltage swell: It is the exact opposite of the voltage sag, which is an increase in voltage for a short duration of time.
- Interruptions: If the RMS voltage is under 0.1 per unit for the short duration and short duration here is one minute. It is called momentarily interruption.

- Under-Voltage: It is same as voltage sag but for more than one minute.
- Over-Voltage: Similarly, it is same as the voltage swell but for more than one minute.
- Interruption for more than one minute: It is known as Sustainable interruption.

Voltage imbalance

- It is three phase system phenomena where the phases are not balanced. Its imbalance may be in either phase or magnitude.
- Its imbalance may be in phase or magnitude.
- This results in unbalanced current even when load is balanced. This results in neutral current or increased magnitude of current.
- Any imbalance in phase or magnitude results in unbalanced current and overloading in the system.

Waveform distortion

- In waveform distortion, voltage or current waveforms are not purely sinusoidal.
- The non-sinusoidal voltage or current will increase the loss of the system and loads will get heated.
- Fluctuations, flickering etc can be caused by it.

Voltage fluctuation

Power frequency variation

Causes for disturbances

- Poor load power factor.
- Nonlinear load or unbalanced load.
- Power quality indices.
- Modelling of harmonic distortion—total harmonic distortion and total demand distortion.

Power quality problems

- Poor load power factors.
- Non-linear and unbalanced loads.
- Power Quality Indices.
- Modelling of harmonic distortion. That is Total Harmonic distortion and Total Demand Distortion.

Nonlinear loads

- Loads are classified in to two categories based on the terminal voltage and current relation through linear loads and nonlinear loads.
- Linear loads are defined as the loads in which the load voltage and current are sinusoidal and in phase with each other. The power factor is unity always.
- Nonlinear loads on the other side, in which the voltage and current are not in phase. The waveforms may be sinusoidal or non-sinusoidal. All practical loads come under this category.

Power quality indices

- Total harmonic distortion (THD)
- Power Factor (PF)
- Telephone Influence Factor (TIF)
- C message Index
- IT Product
- VT Product
- K factors
- Crest factors
- Unbalance factor
- Flicker factor

Total harmonic distortion

- It is the measure of closeness of a waveform with respect to its fundamental component.
- It is an important measure of performance in electrical systems.
- Fourier series is the mathematical tool used to extract the fundamental and harmonic components from a distorted waveform.

Total demand distortion

TDD = THD x (Fundamental current / Circuit Rating).

 All Rights Reserved

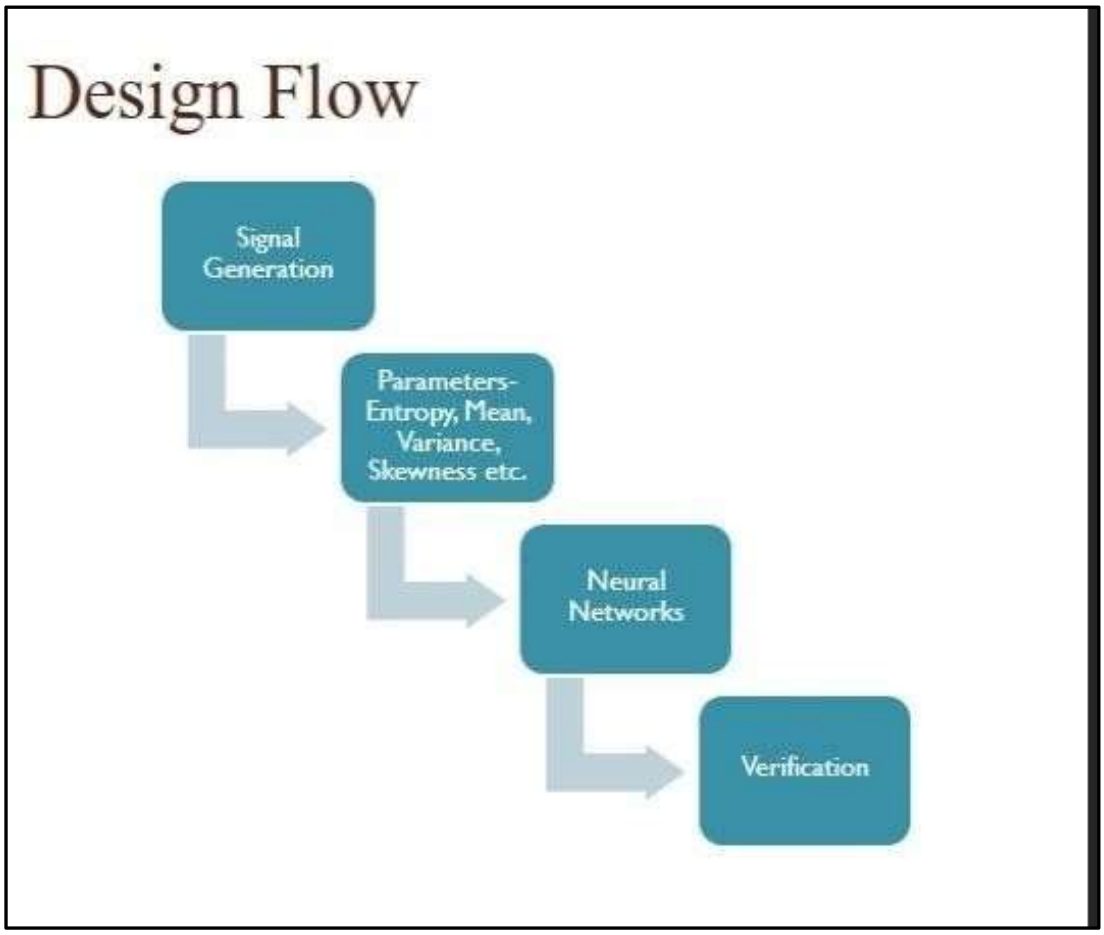


Fig. 2: *Flowchart of Power Quality Analysis.*

Detection and Feature Extraction

Wavelet transform

Real world data or signals often show slowly changing trends on oscillations punctuated with transients. On the other hand, images have smooth regions interrupted by edges on approve changes in contest. These abrupt changes are often the most interesting part of data both perceptually and in terms of the information they provide. The Fourier transform is a powerful tool for data analysis. It does not signify a product change is efficacy. The reason for that is that the Fourier transform represents data as a sum of sine waves which are not localised in time or space.[2] These sine wave oscillates forever. Therefore, to accurately analyse signals and images that have abrupt changes we need to use a new class of functions that are well localised in time and frequency. This introduce us to the topic of wavelet.

A Wavelet is a rapidly decaying wave like oscillation that has zero mean. Unlike sinusoids that extends to infinity of wavelet exists for a finite duration. Wavelet come in different sizes and shapes. This wide range is a key strength of wavelet analysis. To choose the right wavelet we need to consider the application we will use it for. Two important wavelets transform concepts are: Scaling, Shifting

Scaling: It refers to the process of stretching or shrinking the signal in time which Unlike the Sine wave, the wavelet has a Band pass typical in frequency domain.

 All Rights Reserved

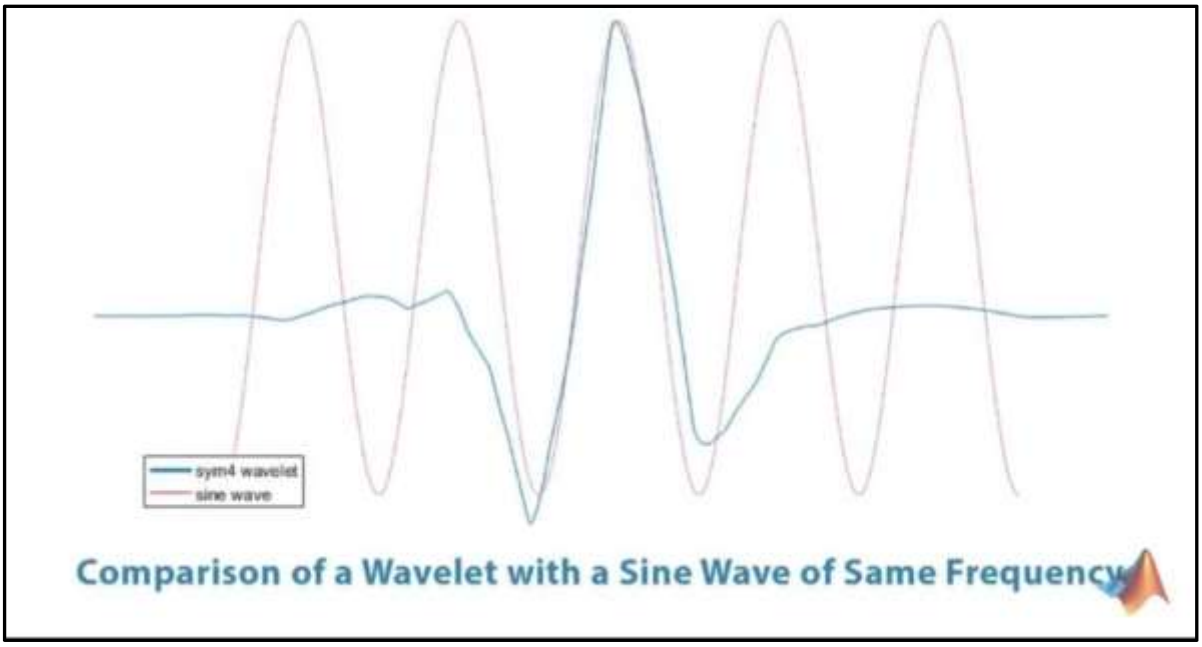


Fig. 3: *Comparison of a Wavelet with a Sine Wave of Same Frequency.*

A stretched wavelet helps in capturing the slowly varying changes in a signal while a compressed wavelet helps in capturing the abrupt changes. Shifting a wavelet simply means delaying advancing the onset of the wavelet along the length of the signal.

A shifted wavelet showed using this notation means that the wavelength is shifted and cantered at K. We required to shift the wavelet to align with the feature we are observing for in a signal.

Two major transforms in wavelet analysis are Continuous and Discrete Wavelet transform.

Discrete wavelet transform

The discrete wavelet transforms, or DWT is ideal for denoising and compressing signals and images as it helps represent many naturally occurring signals and images with fewer coefficients. This enables a sparser representation.

We can obtain different scales by increasing the base scale integer values.

Scaling *2^j (j= 1,2,3,4, .)*

The translation occurs at integer multiples represented in this equation Translation *(2^j)m (m= 1,2,3,4, .)*

This procedure is frequently referred to as Dyadic scaling and shifting. This type of Sampling eradicate redundancy in coefficient. The output of the transform yields the similar number of coefficients as the length of the input signal. Therefore, it requires less memory. The DWT process is equivalent to comparing a signal with discrete multi-rate filter bands.

Given a signal 'S'. The signal is first filtered with special low pass and high pass filters to yield low pass and high pass sub bands. Let us refer these sub bands as A1 and D1. Half of the samples are waste after filtering as per the Nyquist criterion. The filters typically have a small number of coefficients and result in good computational performances.

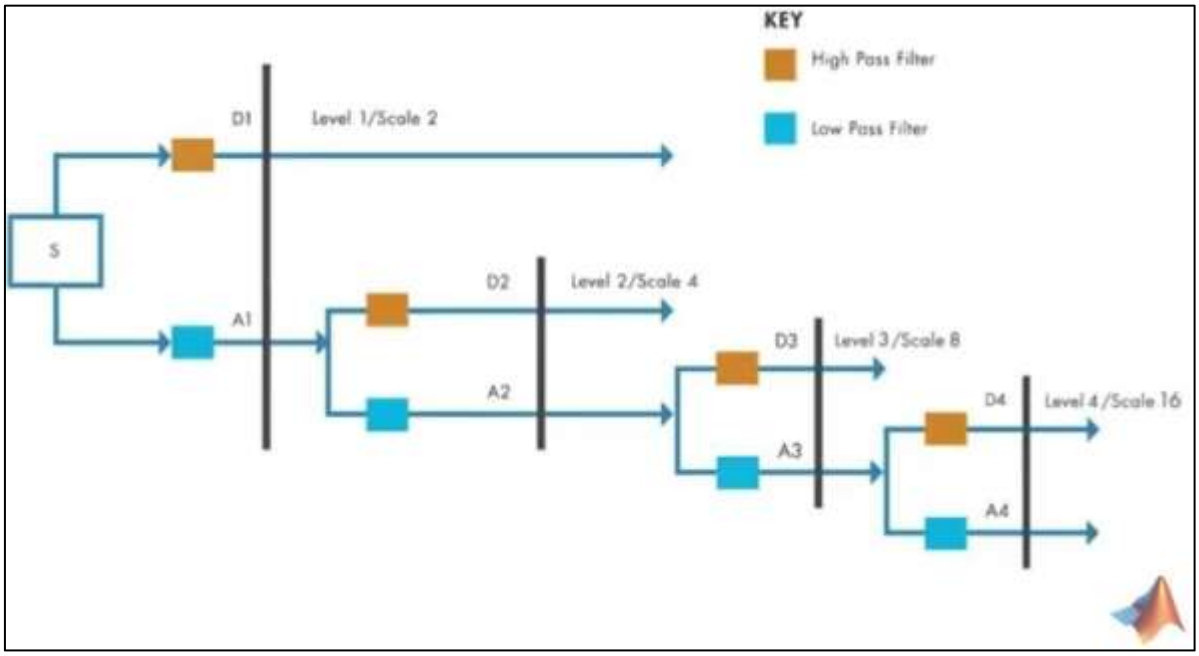


Fig. 4: *Schematic Diagram of Obtaining Coefficients using DWT of a Signal.*

This filters also have the ability to reconstruct the sub band while canceling any aliasing that occurs due to down sampling. For the next level of decomposition iteratively protected by the same technique to yield narrower sub bands A2 and D2 and so on. The length of coefficients in each sub band is half of the number of Coefficient in the preceding stage. With this technique we can capture the signal of interest with a few large magnitude DWT coefficient. While the noise in the signal outcomes in reduced DWT coefficients. This way DWT helps analyse signal at progressively narrower sub bands at different resolutions. It also helps denoise and compressor signals.

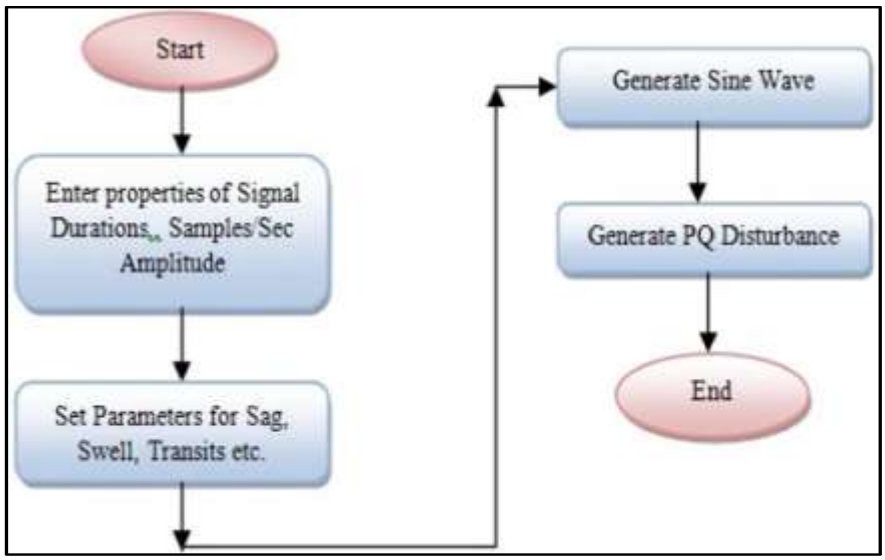


Fig. 5: *Flow Chart for PQ Signal Disturbances.*

METHODOLOGY

An artificial neural network (ANN) (or, neural network) provides a practical general method for, discrete-valued, learning real-valued and vector-valued functions from examples.[7] The type of neural network learning is supervised learning, *i.e.*, the network is provided with example inputs and correct answer for the input. Generally, for regression problem and classification problem, an input belongs to which class (out of several possibilities) that is main concern. For example, we can say, if we want to have network learning to differentiate pictures of cats from pictures dog. Then in that particular case, we would provide the network with a series of pictures and whether the picture is a cat or dog that we will tell for each picture. In case, of regression problem, the main task is learning of real-valued target function. How we will structure our Artificial Neural Network mainly depend on the problem (classification problem or regression problem) we are trying to solve and that has become largely interested subject for the suitability of handling complex and non-linear problems and they are valid for wide range. Artificial Neural Network consists of simple element which is parallel interconnected and is intended to interact with real world as the biological nervous system interacts. An unusual scheme-based programming standpoint is offered by them which exhibit more speed computing compared to other conventional methods. Artificial Neural Network are characterized by the number of interconnections and the node characteristics that are classified by the kind of learning rules has been employed and the type of non-linear elements has been used. An organized topology of Processing Elements (PEs) which are called neurons composes Artificial Neural Network (ANN). In case of MFNN Processing Elements are arranged in layers and PEs in a adjacent layers are connected. The MFNN structure usually comprises of three layers and those are the input layer, the hidden layer and the output layer.

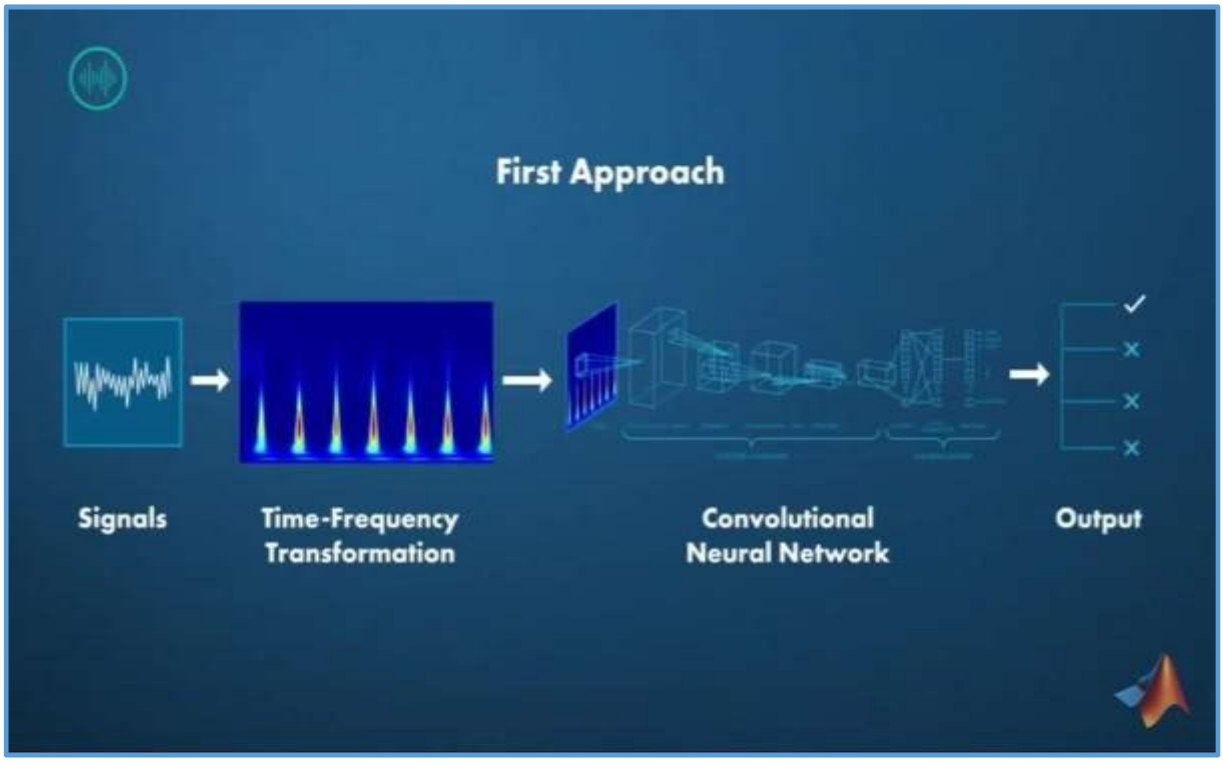


Fig. 6: *Process of Detection and Classification of Signals.*

Power Quality Signal Generation

Electrical signals having voltage sag, voltages well, and interruption is classified by this. But before classification of these signals, detection of the various types of power quality events is necessary. [4]The power quality signals used as input to Discrete Wavelet Transform are synthetic signals. Synthetic power quality signals are produced using MATLAB. [1]For the synthetic signals to signify real-time signal precisely, mathematical parametric equations are used to generate the PQ signals.

The synthetic signals are generated from signal wave equations that are given below,

Disturbance	Model	Parameter
Normal	$x\,t = \sin \omega t$	
Sag	$x\,t = A\,1 - \alpha\,u\,t - t1 - u\,t - t2 \sin \omega t$	$0.1 \leq \alpha \leq 0.9\,,\ T \leq t2 - t1 \leq 9T$
Swell	$x\,t = A\,1 + \alpha\,u\,t - t1 - u\,t - t2 \sin \omega t$	$0.1 \leq \alpha \leq 0.8\,,\ T \leq t2 - t1 \leq 9T$
Interruption	$x\,t = A\,1 - \alpha\,u\,t - t1 - u\,t - t2 \sin \omega t$	$0.9 \leq \alpha \leq 1\,,\ T \leq t2 - t1 \leq 9T$

List of power quality events.

PQ Disturbance	Model	Parameter
Sine	$f(t) = A\sin(\omega t)$,	Frequency = 50 Hz A = 1
Sag	$f(t) = A(1-a(u(t-t_1)-u(t-t_2)))\sin(\omega t)$,	$0.1 \leqslant a \leqslant 0.9$; $T \leqslant t_2 - t_1 \leqslant 9T$
Swell	$f(t) = A(1 + a(u(t-t_1)-u(t-t_2)))\sin(\omega t)$,	$0.1 \leqslant a \leqslant 0.8$; $T \leqslant t_2 - t_1 \leqslant 9T$
Interruption	$f(t) = A(1-a(t\sin(\omega t)$,	$0.1 \leqslant a \leqslant 0.9$; $T \leqslant t_2 - t_1 \leqslant 9T$
Flicker	$f(t) = A(1 + \alpha_f \sin(\beta_f \omega t))\sin(\omega t)$,	$\alpha_f = 0.1-0.2$ $\beta_f = 5-10$ Hz
Harmonic	$f(t) = A(\alpha_1 . \sin(\omega t) + \alpha_3 \sin(3\omega t) + \alpha_5 . \sin(5\omega t) + \alpha_7 . \sin(7\omega t))$	$0.05 \leqslant \alpha_3 \leqslant 0.15$ $0.05 \leqslant \alpha_5 \leqslant 0.15$ $0.05 \leqslant \alpha_7 \leqslant 0.15$ $\sum \alpha_i^2 = 1$
Transient	$f(t) = A\sin(\omega t) + \alpha_{OSC} . \exp\left(-\left(\frac{t-t_1}{\tau_{OSC}}\right)\right) . \sin\omega_{nosc}(t-t_1))$,	$T_{osc} = 0.008-0.04s$ $\omega_{nosc} = 100-400$ Hz
Swell with Harmonic	$f(t) = A(1 + \alpha(u(t-t_1)-u(t-t_2))) . \alpha_1 \sin(\omega t) + \alpha_3 . \sin(3\omega t) + \alpha_5 . \sin(5\omega t) + \alpha_7 . \sin(7\omega t))$,	$0.1 \leqslant \alpha \leqslant 0.8$ $T \leqslant t_2 - t_1 \leqslant 9T$ $0.05 \leqslant \alpha_i \leqslant 0.15$ $\sum \alpha_i^2 = 1$
Sag with harmonics	$f(t) = A(1-a(u(t-t_1)-u(t-t_2)))*A(\alpha_1 . \sin(\omega t) + \alpha_3 \sin(3\omega t) + \alpha_5 . \sin(5\omega t) + \alpha_7 . \sin(7\omega t))$,	$0.1 \leqslant a \leqslant 0.9$; $T \leqslant t_2 - t_1 \leqslant 9T$ $0.05 \leqslant \alpha_3 \leqslant 0.15$ $0.05 \leqslant \alpha_5 \leqslant 0.15$ $0.05 \leqslant \alpha_7 \leqslant 0.15$ $\sum \alpha_i^2 = 1$
Flicker with harmonics	$f(t) = A(1 + \alpha_f \sin(\beta_f \omega t))\sin(\omega t)*A(\alpha_1 . \sin(\omega t) + \alpha_3 \sin(3\omega t) + \alpha_5 . \sin(5\omega t) + \alpha_7 . \sin(7\omega t))$,	$\alpha_f = 0.1-0.2$ $\beta_f = 5-10$ Hz $0.05 \leqslant \alpha_3 \leqslant 0.15$ $0.05 \leqslant \alpha_5 \leqslant 0.15$ $0.05 \leqslant \alpha_7 \leqslant 0.15$ $\sum \alpha_i^2 = 1$
Interruption with harmonics	$f(t) = A(1-a(t\sin(\omega t)*A(\alpha_1 . \sin(\omega t) + \alpha_3 \sin(3\omega t) + \alpha_5 . \sin(5\omega t) + \alpha_7 . \sin(7\omega t))$,	$0.1 \leqslant a \leqslant 0.9$; $T \leqslant t_2 - t_1 \leqslant 9T$ $0.05 \leqslant \alpha_3 \leqslant 0.15$ $0.05 \leqslant \alpha_5 \leqslant 0.15$ $0.05 \leqslant \alpha_7 \leqslant 0.15$ $\sum \alpha_i^2 = 1$
Spikes	$f(t) = \sin(\omega t) + sign(\sin(\omega t)) * \sum k*(u(t-(t_1 + 0.02n))-(u(t-(t_2-0.02n)))$,	$0.1 \leqslant k \leqslant 0.4, 0 \leqslant t_1$ $t_2 \leqslant 0.5T$, $0.01T \leqslant t_2 - t_1 \leqslant 0.05T$

But our concern is only sag, swell, interrupt and noise.

Signal specification

Ts (time period) = 0.2 sec, fs(sampling frequency) =2.5 KHz, f=50Hz,No of cycles=10,No of samples/cycle=5,Total Sampling points=50.Duration of disturbance=81 second. The interval of disturbance from 14 to 95 second of time which is between 75 to 475 sampling points.

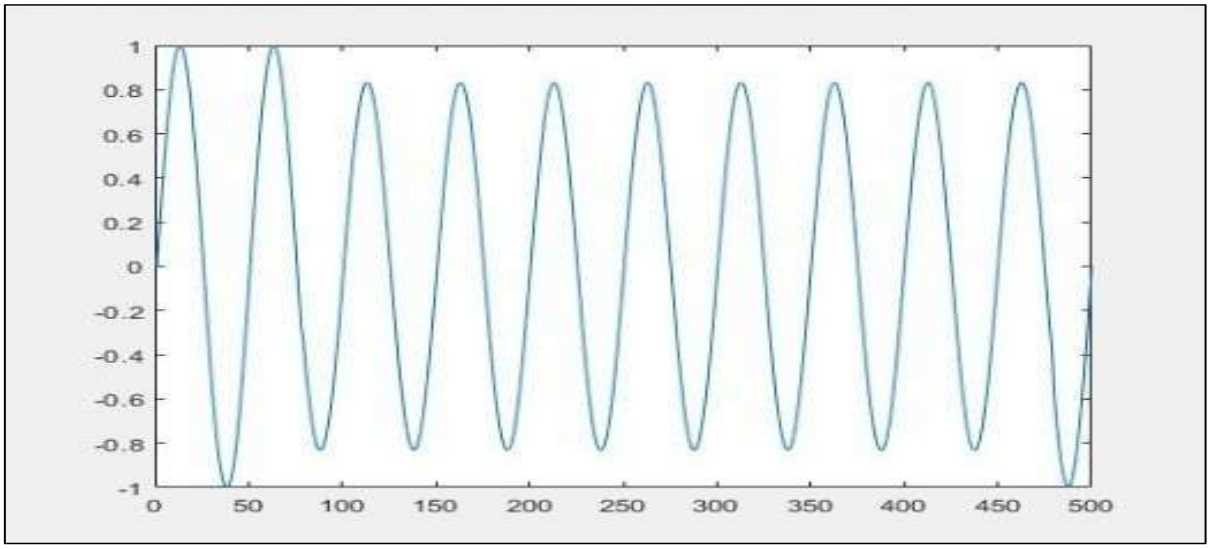


Fig. 7: *Voltage Sag.*

- Voltage sag demonstrated can be characterized as the decline in the nominal rms voltage between 0.1 p.u. and 0.9 p.u. The time length of voltage sag can fluctuate from 0.5 to 1 min.
- Sag happens for the most part because of presence of Single Line to Ground (SLG) faults, starting of heavy motors and heavy current drawn activities.
- Voltage sag can be separated into three wide classes based on the duration. These classes are instantaneous, momentary and temporary sag.

$$f(t) = A(1-a(u(t-t_1)-u(t-t_2)))\sin(\omega t)$$

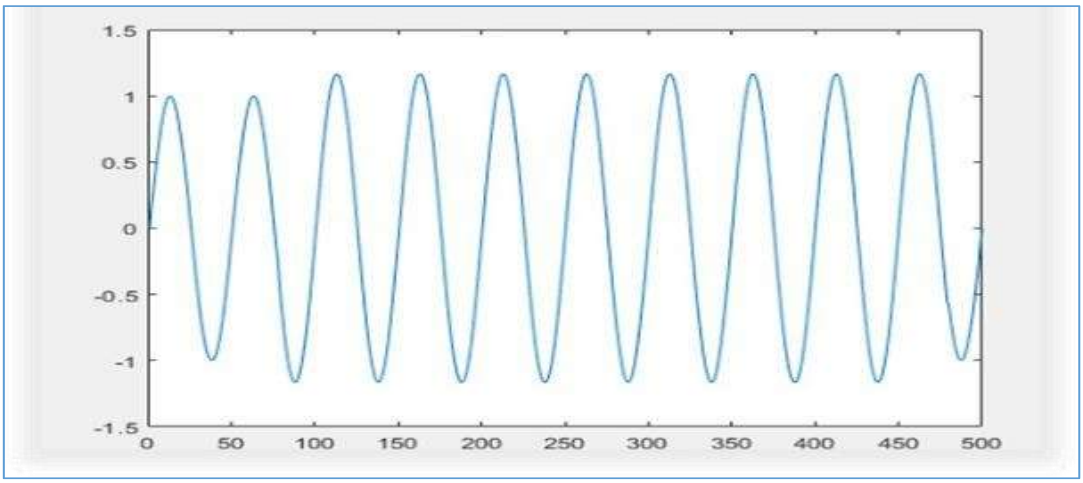


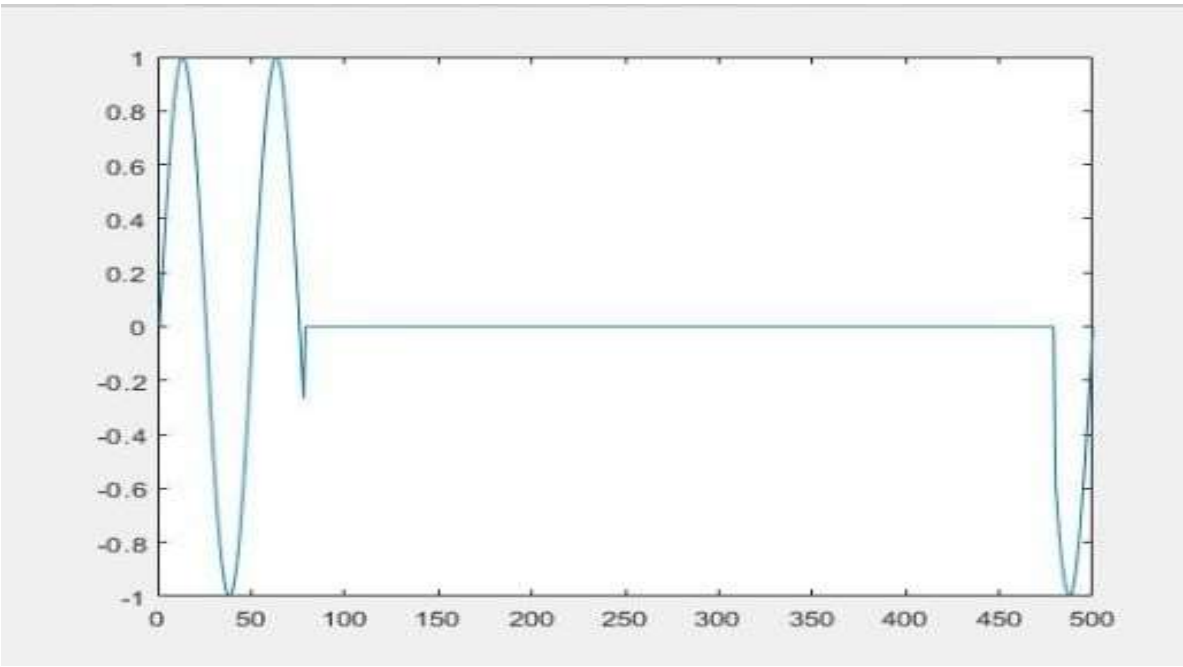
Fig. 8: *Voltage Swell.*

- Voltage swell as indicated can be characterized as the increment in the nominal rms voltage between **1.1 p.u. and 1.8 p.u.**
- The event recurrence of voltage swell is little when compared with event of voltage sag.
- The time length of voltage swell can differ from 0.5 cycles to 1 min, which is same if there should be an occurrence of voltage sag.
- The component that may come about voltage swell incorporates light system loading, incorrect tap setting of the transformer and so on.
- Cause:
- Caused by de-energization of very large loads and energization of a large capacitor bank.

Effects

The severe effects of voltage swell event include overheating of devices and high iron loss in maximum machines applications.

$$f(t) = A(1 + a(u(t-t_1)\text{-}u(t\text{-}t_2))) \sin(\omega t)$$



- Voltage Interruption is can be characterized as the decline in rms voltage such that resultant is not greater than 0.1 p.u.
- It occurs as an aftereffect of loose connection, main faults and reclosing of electrical switch.
- Reclosing of electrical switch activity results in transient phenomenon took after by interruption can likewise be watched.
- The voltage interruption results in inconvenience trip and false operation of the overall system. voltage interruption shows that there are almost zero voltages and no more power supply is available for the system.

$$f(t) = A(1-a(t\sin(\omega t)$$

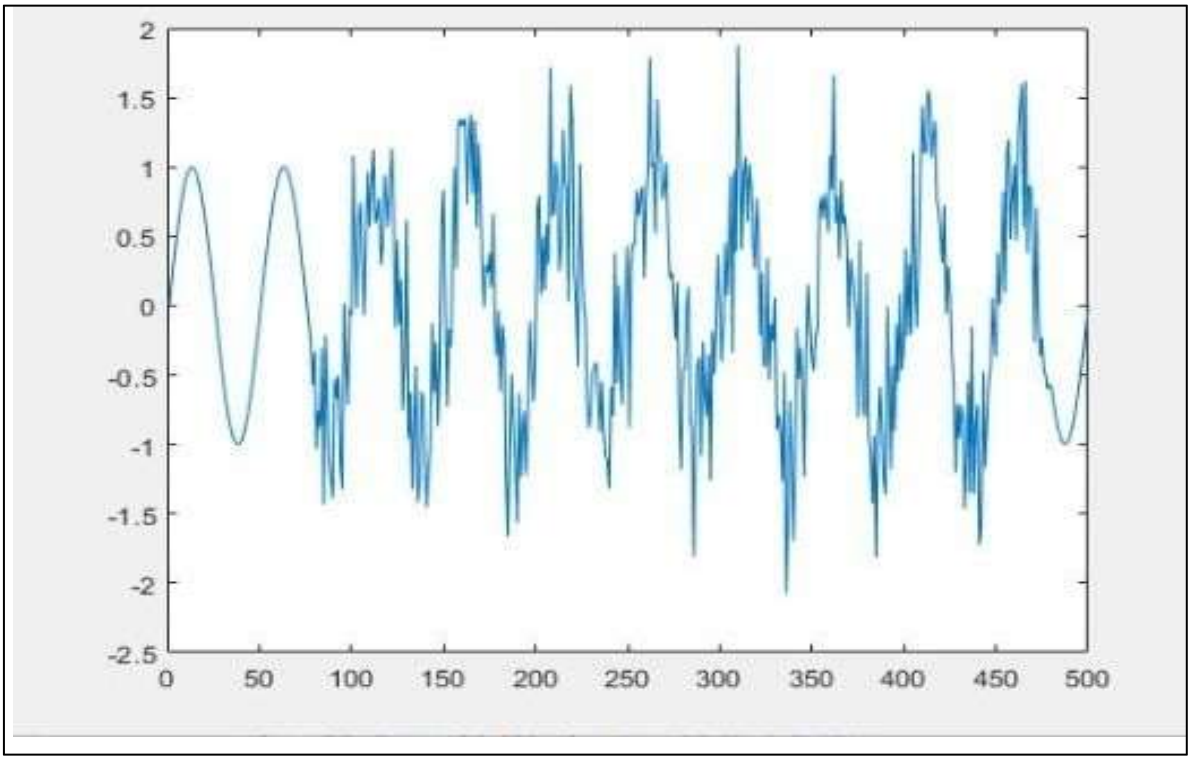


Fig. 9: *Noise Signal.*

- Unwanted electrical signals which produce undesirable noise in the circuit of the control system in which they occur.
- Here for the purpose of detectn we are using Gaussian White noise.
- Gaussian white noise (GWN) is a stationary and ergodic random procedure with zero mean that is well-defined by the following important property: any two values of GWN are statistically independent no matter how close they are in time.[8]
- For all $\tau = 0$, where φw denotes the autocorrelation function, and $E\{\cdot\}$is the "expected value statistical operator.

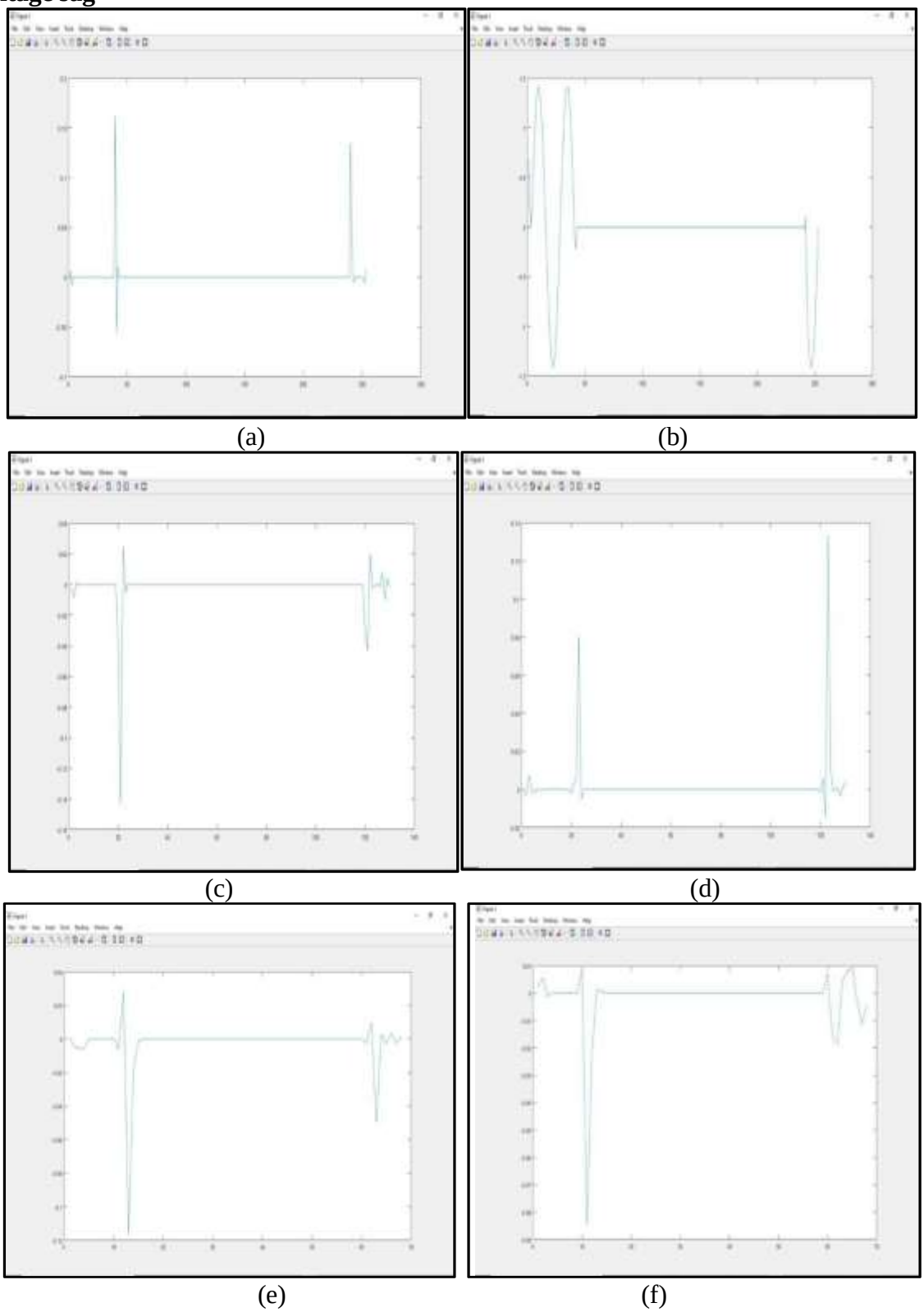
$$\varphi w\{\tau\} = E\{w(t)w\{t - r)\} = E\{w(t)\}.E\{w(t - \tau)\} = 0$$

Detection using WT

The above Disturbances are decomposed into different levels through wavelet decomposition algorithm. The signal is looked at different scales or resolution which is also known as multiresolution analysis(MRA) or sub band coding. With upsurge in every level time resolution declines while frequency resolution increases. The unique deviation of each power quality disturbances from the original sinusoidal waveform is identified both in the approximate and detail coefficients. The different disturbances are studied with different levels.

Usually, one or two scale signal decay is acceptable to discriminate disturbances from their contextual since the decomposed signals at lower scales have high time localization. In extra words, the high scale signal decomposition is not essential since it provides poor time localization. In this case the different power quality disturbances are decomposed up to 4th level for detection purpose.

Interruption in Voltage
Voltage sag



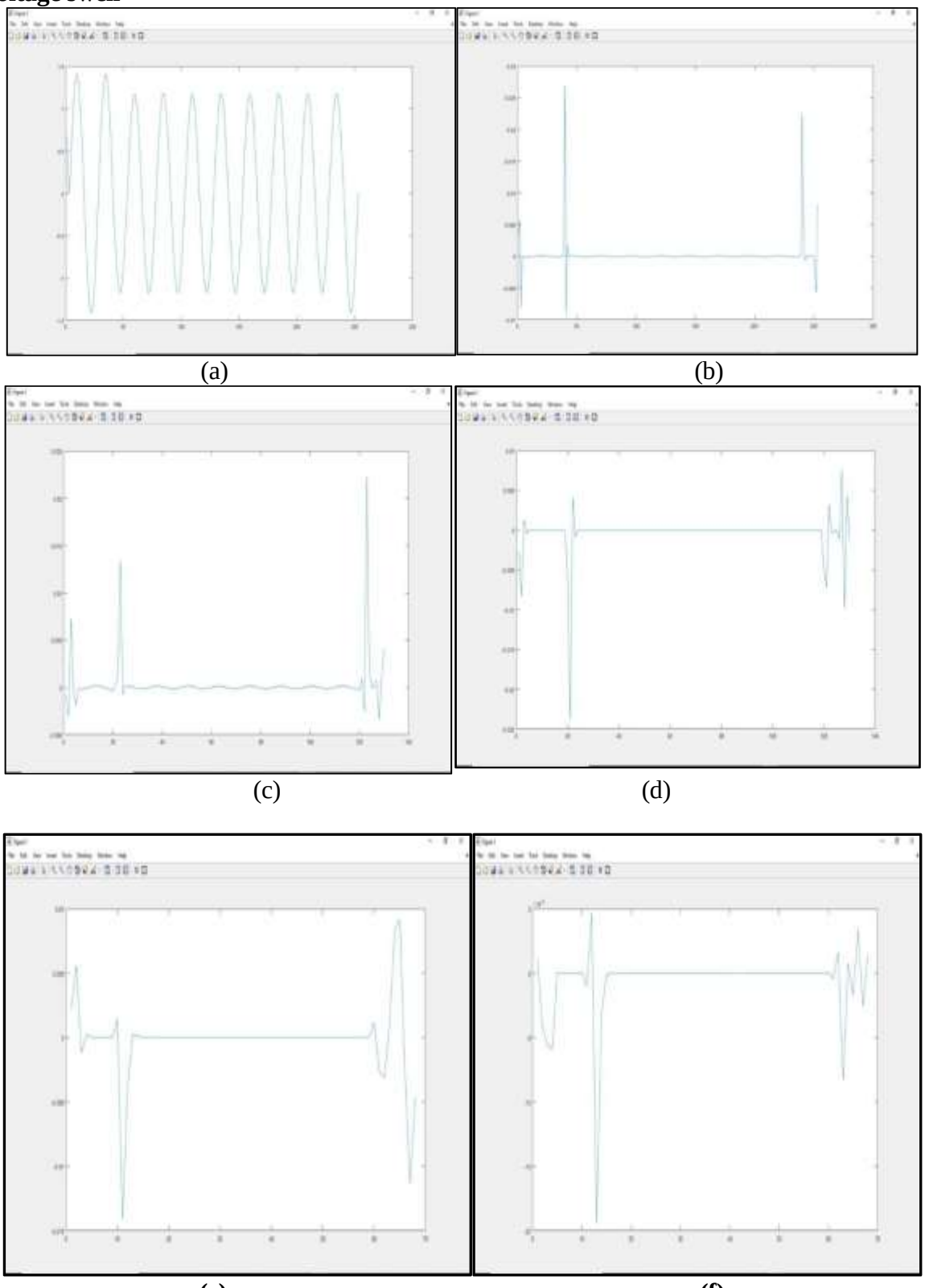
(a) (b)

(c) (d)

(e) (f)

Fig. 10: *(a) Approximate Signal Level 1(a1), (b) Detailed Signal level 1(d1), (c) Approximate Signal Level2 (a2), (d) Detailed Signal Level 2(d2), (e) Approximate Signal Level 3(a3), (f) Detailed signal level 3(d3)*

Voltage swell



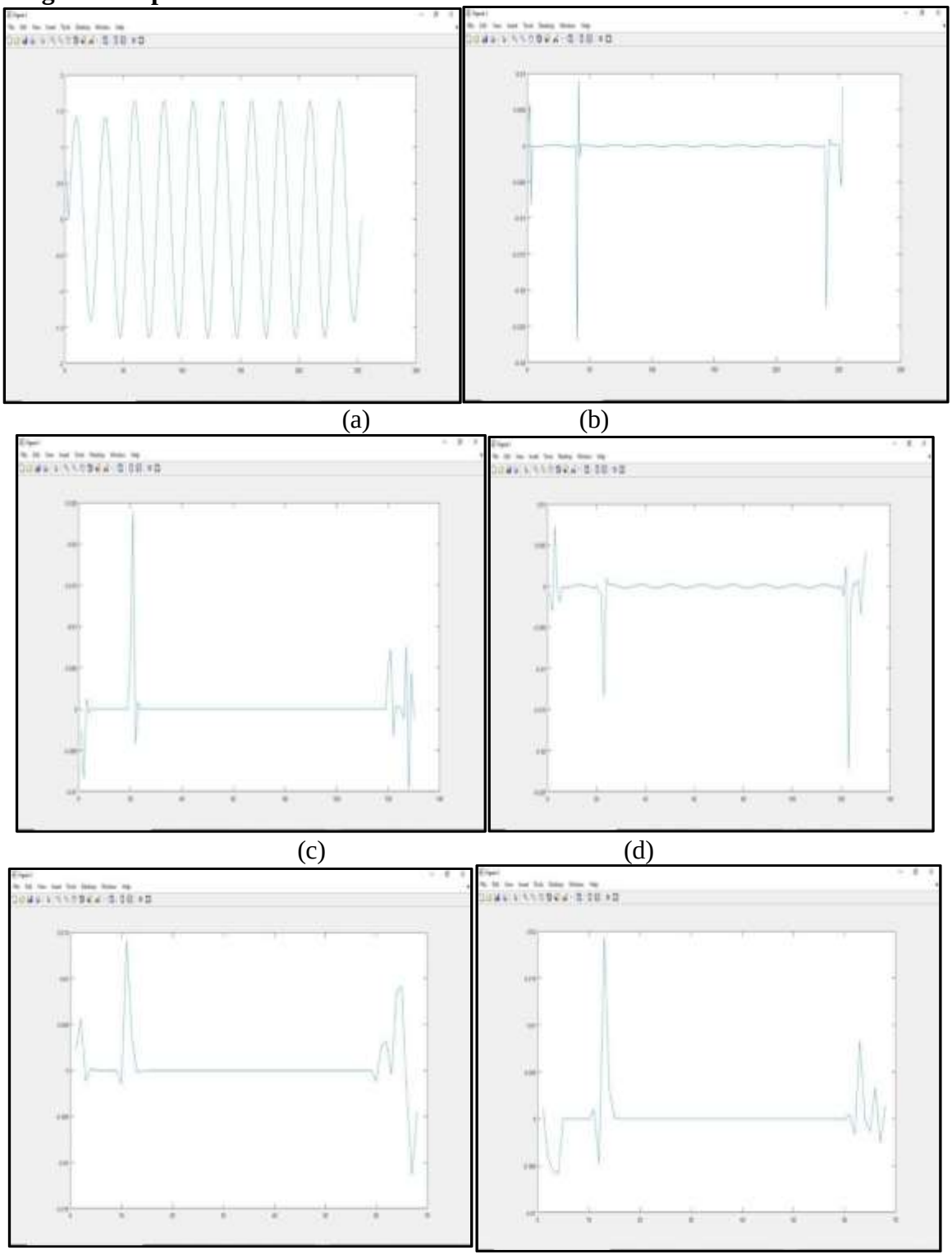
(a) (b)

(c) (d)

(e) **(f)**

Fig. 11: *(a) Approximate Signal Level (a1), (b) Detailed Signal Level 1(d1), (c) Approximate Signal Level 2(a2), (d) Detailed Signal Level 2(d2), (e) Approximate Signal Level 3(a3), (f) Detailed Signal Level 3(d3).*

Voltage interrupt



(a) (b)

(c) (d)

(e) (f)

Fig. 12: *(a) Approximate Signal Level 1(a1), (b) Detailed Signal Level 1(d1), (c) Approximate signal level 2(a2), (d) Detailed signal level 2(d2), (e) Approximate Signal Level 3(a3), (f) Detailed Signal Level 3(d3).*

Compared to the power spectrum of normal signal and the power spectrums of harmonics and its hybrids with sag and swell, reveals that even after low frequency region, repeated peaks contributing non zero power exist in the latter two.

The efficacy of skewness and kurtosis was investigated with the belief that the symmetry and peakness of probability distribution function may change as PQ disturbances leak into the normal signal. Even though sufficient variation in the symmetry of distributions of different classes of disturbances exhibited cannot be observed visually, both were numerically estimated.

CLASSIFICATION

The WNE feature vector constructed for eighth level of decomposition consists of nine dimensions for training and testing the PNN classifier. The total size of the training and testing data-set is 6 × 1000, where 6 is the size of feature set and 1000 is the number of PQ disturbances with 200 samples of each class. From 200 features of each disturbance, 70% used for training, 15% features are used for training and 15% for testing the neural network classifier. The feature vector of 6×150 was applied for training the neural network structure using neural network toolbox 8.2 in MATLAB (R2018a).[9] The PQ disturbances precisely classified are revealed in diagonal elements and the misclassified are shown in off-diagonal elements.

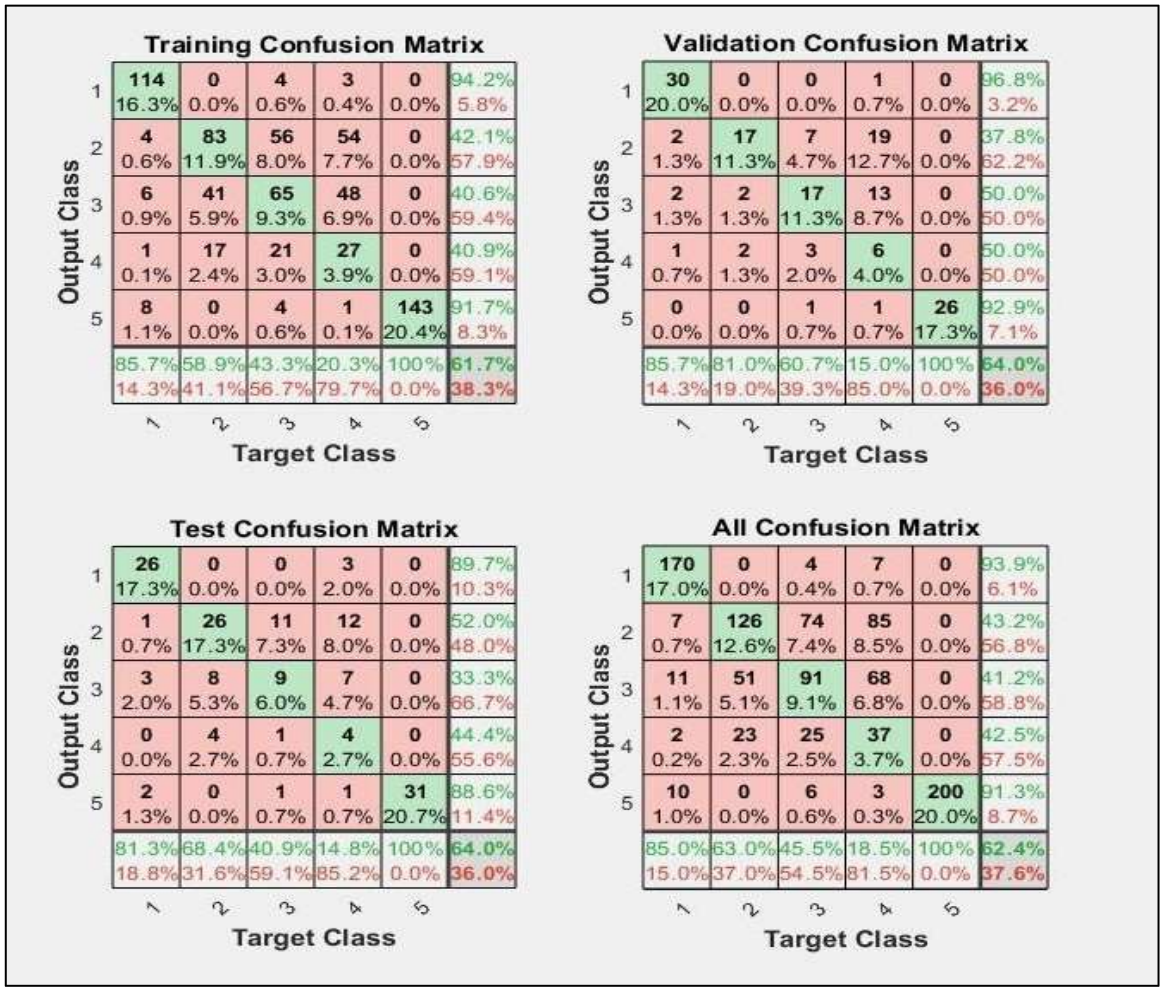


Fig. 13: *Confusion Matrix.*

The various architectures of ANN were also compared for the classification of the PQ disturbances. Every type of ANN is skilled with 100 input data of each class and 100 disturbances of each type are well-thought-out for testing. The classification results of training and testing. The recognition rate of the correctly classified disturbances is the ratio of the correctly classified disturbances to the total number of those disturbances.

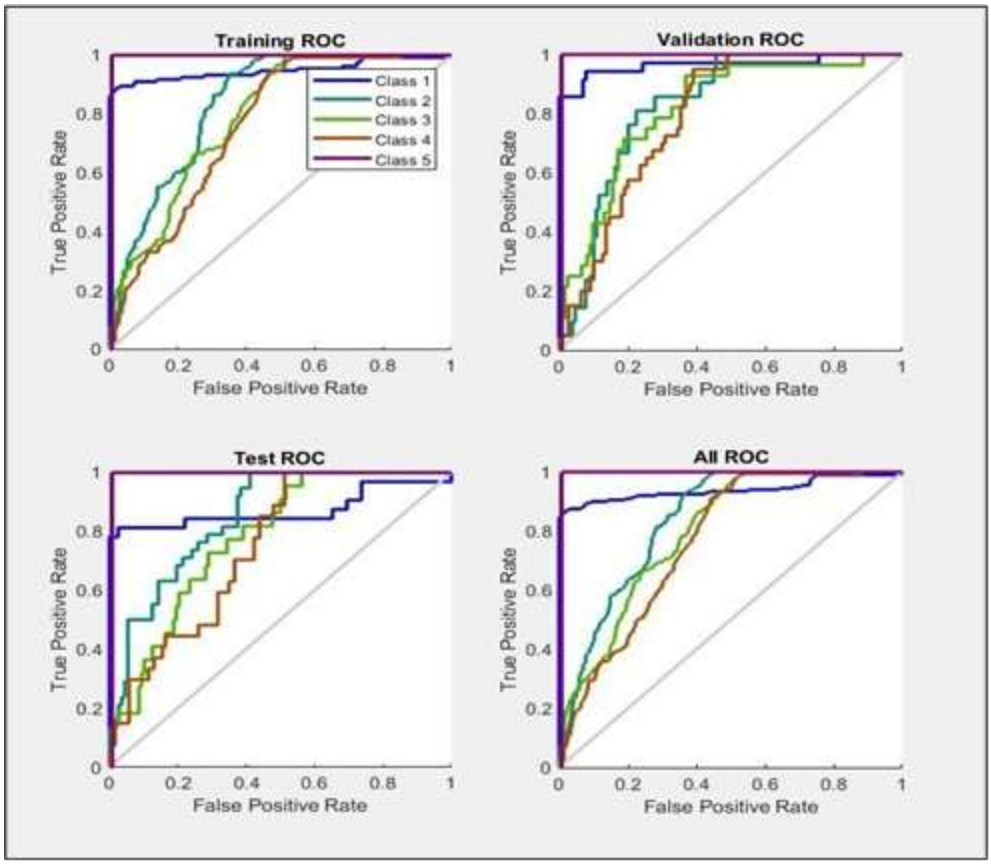


Fig. 14: *True V/s False Positive Rates of Classification.*

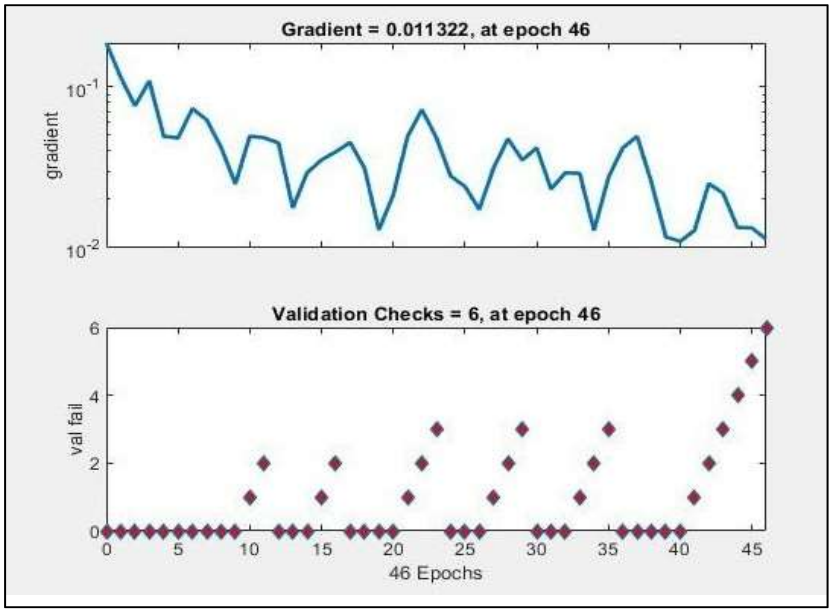


Fig. 15: *Result of Validation Performance.*

RESULT

- Various PQ disturbances are identified and can be seen in above figures like sag, swell and interrupt.
- Feature vectors are obtained that will be used in ANN for classification of PQ disturbances.
- Voltage swell is expansion in the ostensible rms voltage between 1.1 p.u. and 1.8 p.u.
- Time of voltage swell is from 0.5 cycles to 1 min.
- Voltage intrusion happens because of free association, real blames and re closing of electrical switch.
- Gaussian White noise is also obtain .
- Placing quick acting tap changers can moderate voltage swell.

FUTURE SCOPE OF WORK

Classification of PQ Disturbances

- The gigantic number of PQ events happens in distribution frameworks. There is a necessity for a technique that can be utilized for quick and precise recognition of PQ events. The fundamental preferred standpoint of the proposed strategy is low execution time and noteworthy precision. When the disturbance is noticed, it becomes important to categorize which type of events is present.
- We are working on classification through ANN method.
- In this work only two features like THD and Energy are considered for modeling the PQD detection system. More features associated with PQ signals like Entropy, standard deviation etc. can be included for modeling purpose.
- Classification of decomposed power signal using fuzzy logic and comparison the classification accuracy with MFNN and PNN.
- In this report, the effect of noise has not been discussed. In the further research work, the effect of noise part will be included. Instead of simulated signal, the real distorted signal which is measured by digital recorder can also be included in further research work and through a PIC based hardware, it can be implemented by real time.

CONCLUSION

In this report, we haves used Discrete Wavelet transform for identification/detection of different PQ disturbances as well as pure sinusoidal wave. Discrete Wavelet transform is clearly a better method for detection of these PQ events as wavelet transform can offer time and frequency info of a signal simultaneously. Here db4 mother wavelet and decomposition levels equals 4 and Extracted Feature Vector like skewness, standard deviation, kurtosis are used to reduce large matrix into manageable group for detection and later classification of PQ events. Multi-resolution signal decomposition (MSD) technique was used to decompose the signal into lower resolution to obtain Detailed and Approximate coefficient that are further used to extract unique features .Matrix are plotted by using plot tool and hence PQ events are detected .Further these feature vectors can be used to classify the events using ANN which is to be worked on.

After obtaining feature vector, powerful classifier MFNN is used in classification stage 97.905 % average performance is obtained. This thesis projected two types of classification algorithms for the disturbance types considered and could effectively use one of those algorithms to attain a very high error organization rate.

The use of wavelet multi-resolution analysis standard deviation curves along with Fourier domain parameters with the help of a 3-layer feed forward neural network proved to achieve almost a zero error rate for the considered datasets.

REFERENCES

1. Quinquis, A., Radoi, E., Ioana, C., & Mansour, A. (2008). *Digital signal processing using MATLAB* (p. 424). John Wiley & Sons.
2. Wright, P. S. (1999). Short-time Fourier transforms and Wigner-Ville distributions applied to the calibration of power frequency harmonic analyzers. *IEEE transactions on instrumentation and measurement*, 48(2), 475-478.
3. Gu, Y. H., & Bollen, M. H. (2000). Time-frequency and time-scale domain analysis of voltage disturbances. *IEEE Transactions on Power Delivery*, 15(4), 1279-1284.
4. Heydt, G. T., Fjeld, P. S., Liu, C. C., Pierce, D., Tu, L., & Hensley, G. (1999). Applications of the windowed FFT to electric power quality assessment. *IEEE Transactions on Power Delivery*, 14(4), 1411-1416.
5. Polikar, R. (1996). The engineer's ultimate guide to wavelet analysis-the wavelet tutorial. *available at http://www. public. iastate. edu/~ rpolikar/WAVELETS/WTtutorial. html.*
6. Kapoor, R., Gupta, R., Jha, S., & Kumar, R. (2018). Boosting performance of power quality event identification with KL Divergence measure and standard deviation. *Measurement*, 126, 134-142.
7. Kapoor, R., Gupta, R., Jha, S., & Kumar, R. (2018). Detection of power quality event using histogram of oriented gradients and support vector machine. *Measurement*, 120, 52-75.
8. Gaussian White Noise by Wiley Online Library
9. You tube videos on MATLAB tutorials from channel ‘MATLAB’.

Geocarto International

Taylor & Francis
Taylor & Francis Group

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgei20

Detection of engineered surfaces using deep learning approach in AVIRIS-NG hyperspectral data

Shalini Gakhar & Kailash Chandra Tiwari

To cite this article: Shalini Gakhar & Kailash Chandra Tiwari (2021): Detection of engineered surfaces using deep learning approach in AVIRIS-NG hyperspectral data, Geocarto International, DOI: 10.1080/10106049.2021.1953616

To link to this article: https://doi.org/10.1080/10106049.2021.1953616

Accepted author version posted online: 09 Jul 2021.

Submit your article to this journal

Article views: 18

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tgei20

Check for updates

Detection of engineered surfaces using deep learning approach in AVIRIS-NG hyperspectral data

Shalini Gakhar[a*], and Kailash Chandra Tiwari[b]

[a]Information Technology, Delhi Technological University, Delhi, India

[b]Multidisciplinary Centre for Geoinformatics, Delhi Technological University, Delhi, India

*Shalini Gakhar**,** E-mail: shalinigakhar7@gmail.com

Abstract

Hyperspectral remote sensing is opening new avenues for multitude of urban applications. This paper extends target detection method for extraction of engineered surfaces or urban targets, particularly roads and roofs. The study involves application of deep learning using AVIRIS-NG data. In pre-processing, generating ground reference image using Vertex Component Analysis (VCA) is done instead of using it for spectral unmixing of mixed pixels. Principal Component analysis (PCA) is carried out at a scale of 30,40 and 50 components for dimensionality reduction followed by implementation of Convolution Neural Network (CNN) for three window sizes (5,7 and 9). This deep learning measure is effective for high prediction and the results appear significantly higher in comparison to the literature. The time complexity increases with increase in PCA components and window size, making a compromise with accuracy. The study analyses least explored subset of AVIRIS-NG hyperspectral data of Udaipur region (India) to assist urbanisation.

Keywords Target Detection, Hyperspectral Remote Sensing, Dimensionality Reduction, Deep Learning

Accepted Manuscript

1 Introduction

Hyperspectral data have taken a leap over multispectral data, paving way for multiple applications such as soil mapping (Hively, et al. 2011), wildlife and forest mitigation (Hycza, et al. 2018), water body management (Pu, et al. 2017), disaster prediction (Bhangale, et al. 2016), mineral exploration (Arindam 2020), urban land-use classification (Fen, et al. 2018) and many more. The data has advantage of hundreds and thousands of bands giving treasure of spectral and spatial information. This information implies a three-dimensional structure, also known as *hypercube*. The first two dimensions correspond to an arrangement of pixels or location and the third dimension depicts the bands. Radiance or reflectance values at every pixel are used to plot spectral signatures as a function of wavelength in continuous bands. A spectral signature may be considered as a fingerprint of a material, in order to uniquely identify it. Besides, the aforementioned applications, Hyperspectral remote sensing extends its scope over target detection applications (Manolakis, Marden and Shaw 2003) also, where a target may comprise of a single pixel (full-pixel targets) or a part of pixel (sub-pixel targets).

Target detection problem is also regarded as a problem of binary classification in which targets that lie within the circumference of region of interest are extracted from the background (non-target class). Data and the detection algorithms are the two important aspects of target detection. As data apprehends the variations, while the algorithms provide suitable method for extraction of appropriate information. These algorithms can be broadly categorized into spectral-matching and anomaly detection, depending upon the availability of labelled data (Burgers, et al. 2009). Techniques for which, the spectral signature along with additional details are already known from standard spectral libraries, ground truth, in-situ analysis or from the image itself, comes under

the group of spectral matching. These include algorithms like Spectral Angle Mapper (SAM) (James and Richard 2005), Spectral Correlation Measure (SCM) (Robila 2005) etc. In anomaly detection, no *a priori* knowledge about the target is available, therefore incongruities are detected assuming their exceptional behaviour with respect to the background. Orthogonal Subspace Projection (OSP), Reed-Xiaoli (RX) are some algorithms belonging to this category (Tiwari, et al. 2013).

Recently, detection of urban targets (blend of numerous material compositions such as roads, roofs, railways tracks and other built-up surfaces etc) has become a subject of exploration. Increase in urbanisation, requirement of new infrastructures, city planning, maintenance of ancient monuments, population growth analysis and sustainable development in cities are major factors influencing the same (Pham, Yamaguchi and Bui 2011). Urban engineered surfaces/targets have hitherto fore been extracted using standard classification approaches with varying accuracies. Potential use of hyperspectral data for urban land cover classification and its comparison with multispectral data is characterised using SAM and Support Vector Machine (SVM) (Hegde, et al. 2014). Far along, spectral features are fused with the spatial aspect of hyperspectral data to get enhanced results. Morphological profiles (Palmason, et al. 2005), extended morphological profiles (Benediktsson, Palmason and Sveinsson 2005), textural features (Reza and Ghassemian 2017), marker based minimum spanning forest (Akbari, et al. 2016) and many more are further explored as spatial features fused with spectral features. Conventional techniques for collection of associated information are often based on field surveys or manual analysis of aerial images, which are unable to keep pace with the growing urbanisation (Uta, et al. 2007). Wherein, hyperspectral remote sensing provides a time and cost-efficient solution for extraction of urban surfaces, which is a preliminary objective for conducting any sort of urban studies (Weber, et al. 2018).

Apart from aforesaid techniques, machine learning and deep learning-based methods are also evolving for target detection due to their exceptional capabilities to learn from voluminous data, extract useful information and predict for the unknown dataset. Fernando et al implemented Gaussian Synapse artificial neural network (ANN) for detection and unmixing of endmembers within hyperspectral images (López, Crespo and Duro 2009). Kernel based target detection techniques have been very popular in extraction of sparse targets, as they can address the issues like curse of dimensionality, less labelled information, atmospheric noise etc. Wang and Duan have discussed these research issues by developing an SVM based approach jointly employing spatial, spectral and hierarchical structure information and integrated into the classifier with multiple kernels, on standard datasets (Yi and Duan 2018). With advent of high-end equipment and advanced techniques, a lot of data may be captured, with supporting storage capabilities and processing features (Kruse, et al. 2000). Multitude of methods use machine learning as foundation, which is continuously evolving field in itself, instigating deep learning approach which works on automated feature extraction, making use of in-depth architectures. Autoencoders (Windrim, et al. 2019), Recurrent Neural Networks (RNN) (Hang, et al. 2019) (Mou, Ghamisi and Zhu 2017), Convolution Neural Networks (CNN) (Hu, et al. 2015), Generative Adversarial Networks (GAN) (Shi, Liu and Li. 2017), as part of deep learning approaches have also been exploited in near past for hyperspectral remote sensing. However, urban target detection using deep learning approaches in hyperspectral remote sensing, suffers from multiple challenges [refer Table 1].

Most of the research work has been carried out for detection of natural targets such as vegetation, minerals, glacial lakes etc. (Keith, et al. 2014), (Deyu, et al. 2016), which opens scope of expansion in the field of urban target detection. The prevailing limitation of small size of targets such as roofs, pavements arising due to the sensor characteristics, pose an added difficulty in the extraction process. The composition of urban surfaces in hyperspectral data is characterised by use of similar materials, like roads and roof are often constructed from concrete, sand, asphalt etc, leading to similar spectral signatures (Herold, Roberts, et al. 2004). Also, most of the image pixels are mixed with other components, representing spectral information of more than one material (Heiden, et al. 2001). Due to limited access to spectral content, high amount of user interaction is needed to

Accepted Manuscript

delineate impervious surfaces (Van der Linden 2009). The urban surfaces are in a continuous influence of environmental conditions, such as, aging, weathering, humidity which have a direct impact on their appearance (Lemp 2005). Particularly roof surfaces have different geometrical structures, producing a partial signal for the sensor, when light falls on them (Szabó, Burai, et al., Testing algorithms for the identification of asbestos roofing based on hyperspectral data 2014).

Heiden et al have mentioned a hierarchical categorisation of urban surface materials into four levels. Level 1, gives a broader view and is divided into impervious and pervious surfaces. In Level 2, impervious surfaces are further categorised into open and overbuilt areas. Overbuilt constitutes buildings/roofs and open areas have fully and partially impervious surfaces under them. Level 4 is decomposed into urban surface materials of which they are composed of, such as asphalt, metal, clay, concrete etc. Thatch, wooden, slate roofs are mostly replaced by cutting-edge roofing materials like solar tiles, asphalt shingles, stone-coated steel etc. in urban areas. In order to keep pace with frequently altering urban features, it is important to address aforesaid research gaps in an automated way.

Hyperspectral data is rich in providing spectral insights but, high dimensionality of the data invites challenges like requirement of high computational time and space (Kallepalli, Kumar and Khoshelham 2014). Many researchers have used standard datasets for their proposed work such as, University of Pavia (ROSIS Sensor) and Urban (AVIRIS Sensor) dataset, but they may lack significant features of regional materials used for construction of urban surfaces, varying with location (Paoletti, et al. 2019). The spectral signature of a particular pixel is under influence of certain parameters occurring due to sensor's focal length, resolution, altitude; scene's illumination, complexity; atmospheric hindrances such as presence of clouds, fog, haze, temperature and moisture content (Kerekes and Goldberg 2013). Apart from these, spatial (sub-surface or surface, single pixel or sub-pixel target), morphological (point, linear, area), and processing factors play an important role in characterising the spectra (Tiwari, et al. 2013). Various studies have unravelled the high amount of spectral variability of urban surfaces due to the chemistry or morphology inherent to the material and highly depends on particle size, particle shape, packing density (Theiler, et al. 2019). This necessitates the need for automated identification of urban surfaces, as a prerequisite for hyperspectral image analysis. Benchmark datasets are important to evaluate methods using real-time situations, as it becomes difficult to compare approaches adopted by different researchers, under different experimental set-ups (Gewali, Sildomar T and Saber 2018).

Deep learning works on automated extraction of features needed for mining useful information from raw data. High dimensionality of hyperspectral data affects supervised methods in terms of accurate derivation of statistical parameters leading to overfitting of network, in case of supervised approach (Paoletti, et al. 2019). Large number of calculations resulting from deep architectures involve high computational load in terms of time and storage for parallel computing (Cheng, et al. 2017). Learning rate, number of epochs, activation function are some of the additional parameters to be tuned for an optimized network and better detection rate, thereby adding dispensable trial runs for training the deep architecture. The "black-box" conduct of deep learning models may hinder to visualize the parameters, optimise the model and make it hard to interpret the training procedure (Shwartz-Ziv and Tishby 2017).

The objective of this paper is to study an alternative approach for extraction of engineered surfaces by using deep learning approach for their extraction. The novelty of the work lies in the idea of generating the ground reference for an AVIRIS-NG image dataset in first stage, dimensionality reduction as second stage and later on extraction of urban surfaces particularly roads and roofs as last stage. Further, VCA is used for generating the ground truth, PCA reduces the dimensionality of the hyperspectral data and CNN is the deep learning measure to extract urban targets. The approach used for the study appears to be well suited for urban target detection using hyperspectral data. Prior to this work, authors have tried to implement multiple target detection algorithms along with target

Accepted Manuscript

detection on the basis of distance, angle, information and machine learning level (Comparative assessment of target detection algorithms for urban targets using hyperspectral data, manuscript 'in press' Photogrammetry Engineering and Remote Sensing). The presented approach is an extended effort to detect engineered surfaces using deep learning method.

The very next section gives a detailed theoretical background, Section 3 describes the dataset used, Section 4 discusses the methodology and Section 5 demonstrates the corresponding results followed by performance analysis, major Conclusions of the work done.

2 Theoretical Background

2.1. Dimensionality Reduction

The process deals with the computationally intensive image datasets to reduce its complexity, thereby capturing only the required significant details. Enormous amount of data is stored by the sensors, demanding dimensionality reduction to reduce the ambiguity in data. Principal Component Analysis (PCA) is the most common technique used for dimensionality reduction and feature extraction for hyperspectral data (Huang, et al. 2020), (Farrell and Mersereau 2005). A probable group of correlated features are projected to a new space based upon the eigen values and the new generated vectors are known as Principal Components (PCs), arranged in descending order of variance. If x is vector of d random variables, C_x is the covariance matrix then PCA constructs an orthogonal transformation $y = Ax$, where $A = [u_1^T; \dots; u_d^T]$, where $u_k's$ with $k\epsilon[1, d]$ are the unit orthogonal eigenvectors of matrix C_x satisfying equation (1)

$$C_x u_k = \lambda_k u_k \quad (1)$$

Where $\lambda_k's$ are the eigen values of matrix C_x, satisfying $\lambda_1 \geq, \dots, \geq \lambda_d$. It can be shown that $u_k's$ can maximally explain the variance $Var(u_k^T x) = u_k^T C_x u_k$ with maxima λ_k (Su, et al. 2017).

2.2. Ground Reference Generation

Acquisition of the ground reference associated with the area under investigation is a tedious, expensive, time and effort consuming task, also getting the data of nearly same time span as of airborne or space borne imagery is a challenge in itself. In case of supervised classification, image-based endmember came into existence to automate the classification process (Kale, et al. 2017). The spectral signatures in the standard libraries belong to different hyperspectral sensors, therefore number of spectral bands and electromagnetic range of the spectra may largely vary. The procurement conditions (ground based, airborne, spaceborne) along with atmospheric hindrances also influence the spectra in many ways. Consequently, to ensure spectral coherence, spectral signatures are derived from the image for endmember realization (Zhu 2017). Recently general method to label hyperspectral imagery is introduced to mark endmembers and their abundances. Endmember being characterised as pure signature of class represents single material within the scene. Numerous algorithms have been proposed for the extraction process in literature, such as, Pixel Purity Index (PPI) (Boardman 1995), N-FINDER (Winter 1999), Vertex Component Analysis (VCA) (J. M. Nascimento 2005) and many more. These techniques are based on assumption that, at least one pure pixel of each endmember is present in the data, which may not be true for every dataset. VCA is used to derive ground reference data for the proposed study, as it exploits two facts: the endmembers are the vertices of a simplex (can be broken down into constituent n-dimensional elements) and the affine transformation of simplex is also simplex. (J. M. Nascimento 2005). The mathematical interpretation of VCA involves observed spectral vector, given by equation (2)

Accepted Manuscript

$$r = x + n = M\,\gamma\alpha + n \quad (2)$$

Where r is an L-vector (L is the number of bands), $M \equiv [m_1, m_2, \ldots\ldots\ldots m_p]$ is the mixing matrix (m_i denotes the i^{th} endmember signature and p is the number of endmembers present in the covered area), $S \equiv \gamma\alpha$ (γ is the scale factor modelling illumination variability due to surface topography), $\alpha = [\alpha_1, \alpha_2, \ldots\ldots\ldots \alpha_p]^T$ is the abundance vector containing the fractions of each endmember. Here $(.)^T$ stands for transpose and n is the additive noise. The abundance fraction always satisfies the physical constraints of non-negativity ($\alpha \geq 0$)and the positivity constraint $1^T\alpha = 1$, where 1 is a $p \times 1$ vector of ones.

The spectral signatures of the endmembers may be chosen from the standard digital spectral libraries, (such as USGS), captured from the field, processed in laboratory or derived from the imagery. The abundance fractions are generated using Dirichlet distribution (Nascimento and Bioucas-Dias 2011), given by Equation (3):

$$p(\alpha_1, \alpha_2, \ldots, \alpha_p = \frac{\tau(\mu_1+\mu_2+\cdots+\mu_3)}{\tau(\mu_1)\tau(\mu_2)\ldots\tau(\mu_p)} \times \alpha_1^{\mu_1-1}\alpha_2^{\mu_2-1} \ldots \alpha_p^{\mu_1-1} \quad (3)$$

Where $0 \leq \alpha_i \leq 1$, $\sum_{i=1}^{p}\alpha_i = 1$, $E[\alpha_i] = \mu_i/\sum_{k=1}^{p}\mu_k$ is the expected value of the i^{th} endmember fraction, and $\tau(.)$ denotes the gamma function. Parameter γ is Beta (β_1, β_2) distributed, constructing equation (4):

$$p(\gamma) = \left(\frac{\tau(\beta_1+\beta_2)}{\tau(\beta_1)\tau(\beta_2)}\right)\gamma^{\beta_1-1}(\gamma-1)^{\beta_2} - 1 \quad (4)$$

2.3. Convolution Neural Network

In recent years, various deep learning architectures have been proposed by researchers and CNN being employed as feature extractor and classifier for spectral, spatial and spectral-spatial configurations. In spectral models, the data is exploited pixel wise by extracting spectra of individual pixels or a particular class in the scene, therefore a priori information is necessary to identify the pixels belonging to target of interest (W. Y. Hu 2015). Spatial features may also be derived from the complete spectral data by applying dimensionality reduction or cropping spatial patches of a particular window size (Zhao 2015), (Haut 2019). In the third approach, the complete three-dimensional data is utilized for extraction of sub-volumes in terms of spectral and spatial knowledge because of flexible architecture of CNN (Mei 2017).

CNN is a multi-layer architecture to process different features where each layer has decisive impact on the model. Apart from input and output layer, the most common layers comprised in CNN are convolution layer, activation layer, down sampling layer and a fully connected layer.

2.3.1. Convolution layer

This layer contains a set of neurons acting as linear kernels to analyse the statistical characteristics of a hypercube ($lines \times samples \times bands$) in which data features are equally distributed with respect to spatial positions. Therefore, the features learned for a particular target can be applied to other regions to locate the same target. Fix sized filters are convolved by sliding-window mechanism overlapping the input data at certain interval called stride. For computation dot product of weights and input layer is performed with an added bias (equation 5)

$$X^{(l)} = (W^{(l)} \times X^{(l-1)} + b^{(l)})_{K^{(l)}\times k^{(l)}\times k^{(l)}\times q^{(l)}} \quad (5)$$

Accepted Manuscript

Output volume $X^{(l)}$ is generated when l^{th} convolution layer applies $K^{(l)}$ linear 3 D kernels over the input layer $X^{(l-1)}$ which performs a dot product between weights $W^{(l)}$ and biases $b^{(l)}$ respectively, and small fragments of input data. The general feature calculation is done using equation 6 (Paoletti, et al. 2019).

$$x_{i,j,t}^{l_{(z)}} = \sum_{i=0}^{k^{(l)}-1} \sum_{j=0}^{k^{(l)}-1} \sum_{t=0}^{q^{(l)}-1} (w_{i,j,t}^{(l)} \cdot x_{(i\cdot s^l+i),(j\cdot s^l+j),(t\cdot s^l+t)}^{(l-1)}) + b^{(l)} \quad (6)$$

2.3.2. Activation layer

Activation functions are mathematical equations which are the decisive to fire an input neuron relevant for model prediction. Computational efficiency of an activation function is equally important because it is calculated for a lot of neurons iteratively. They are linear (behaves like simple linear regression model) or non-linear (for complex data and mapping between input and output layer with stack of hidden layers) in nature. The result of this layer constitutes of output volume containing the activation values from the previous layer, as shown in equation 7.

$$X^{(l)} = f(X^{(l-1)}) \quad (7)$$

$f(\cdot)$ can be varied as linear or non-linear manner. The common activation functions are linear, sigmoid, tanh and ReLU. Depending upon the application, the activation functions may be chosen wisely.

2.3.3. Down sampling layer

Time and space complexity for implementation of deep neural networks is quite high, therefore to reduce the size of feature maps generated by the previous layers down-sampling is used. Down-sampling can be attained by varying the stride of convolution or by pooling layer. In all cases, pooling helps to make the representation become approximately invariant to small translations of the input. Invariance to translation means that if we translate the input by a small amount, the values of most of the pooled outputs do not change (Goodfellow 2016). The objective of the pooling is to, (i) reduce the spatial dimensions and volume, (ii) contribute the data with certain invariance to small transformations, (iii) reduce computation time and space (Paoletti, et al. 2019).

2.3.4. Fully connected layer

In this layer, each node of a layer is connected to all the nodes in the following layer which is required for end-to-end training. The architecture of CNN is shown with respect to hyperspectral image processing in Figure 1.

3 Dataset Description

3.1. Aviris-NG Data Collection

The Hyperspectral data analysis is acquired as part of Airborne Visible and Infrared Imaging Spectrometer – Next Generation (AVIRIS-NG) data collection campaign held in February 2016. The area being considered for investigation is Udaipur, Rajasthan scene having 1295 × 388 pixels with 425 bands in the range of 376 nm to 2500 nm, with 5 nm spectral resolution and 8.1 m spatial resolution. Level 2 (derived geospatial variables at same resolution and location as Level 1 source data) spatial subset of 400 lines and 400 samples, which constitutes maximum urban area has been selected manually for conducting the set of experiments.

Accepted Manuscript

Engineered surfaces/targets may be regarded as a surface which is created artificially and has unique characteristic with some specific thickness to get a unique spectral signature. It may consist of either individual or multiple components (end-members) and can be considered stationary over a large span of time. For example: Roads, Roofs, railway tracks, pavements, parking lots etc. Engineered surfaces may be characterized as *pervious* (allowing water to pass through) and *impervious* (not allowing fluid to pass through) as Level 1 categorization. They are further divided on the basis of constituent materials, such as, concrete road, bitumen roof, metal bridge etc. as Level 2 classification. The current work is restricted to Level 1 detection, that too for subset of engineered surfaces/targets as roads and roofs.

3.2. Ground Data Collection

The spectroradiometer used for ground spectra collection for the very same region spans its range over 2151 channels with 1 nm spectral sampling. Fig. 2, shows the RGB (R = 51, G = 33, B = 20) representation of a subset of size 400 × 400 of the complete Udaipur image. The rectangle in the top middle represents the target roof, which is easily distinguishable from the background. The second rectangle in the bottom left is an enlargement of highway road whereas the third rectangle at the bottom right depicts another type of roof. Table 2 provides a brief introduction about the image and ground data collection parameters considered for the study.

4 Methodology

4.1. Pre-processing

4.1.1. Raw data conversion and dissemination

Three kinds of product are available from airborne data collection campaign. Level 0: Raw data captured by sensor, Level 1: Calibrated Radiance Image Cube and Level 2: Atmospherically corrected surface reflectance. Onboard and vicarious calibrations are done to monitor radiometric response or gain coefficients of sensor spectral bands over time. 6Sv2 Radiative Transfer Model is used to predict the sensor's Top of Atmosphere (TOA) radiance using ground reflectance measurement with other external parameters like total ozone, vapor content etc. Other factors like sensor geometry are taken form header file and solar geometry is derived from image acquisition time. When airborne instrument was flying overhead, ISRO investigators provided coincident data related to atmospheric conditions and ground reflectance, enabling end-to-end AVIRIS-NG radiometric and spectral calibration. For radiometric correction, method from (Thompson, W., et al. 2018) is used whereas, atmospheric correction is done by (Thompson, Gao, et al. 2015). The inhouse software 'Specsoft' developed by ISRO is used for pre-processing the data. Therefore, the authors received atmospherically corrected Level 2 data from ISRO, on which the further analysis is performed. Rest of the details about the campaign can be found at https://vedas.sac.gov.in/aviris_web/pdf/AVIRIS_NG_FIRST_PHASE_REPORT.pdf.

4.1.2. Removal of bad bands

As part of pre-processing, Level 2 (derived geospatial variables at same resolution and location as Level 1 source data) AVIRIS-NG data having 425 bands is processed and bands severely affected by atmospheric gases and water vapors are removed. A total of 53 bands ranging from 1348-1433 nm in near infrared (NIR) region and 1778-1954 nm in short wave infrared (SWIR) region are not considered for further analysis of hyperspectral imagery. After removal of bad bands, the implementation is carried out on 372 bands.

Accepted Manuscript

4.1.3. Selection of Urban targets

According to National Geographic Society, urban areas are very developed, meaning there is a density of human structures such as houses, commercial buildings, roads, bridges, and railways. Therefore, a subset of existing infrastructure, that is roads and roofs are chosen as engineered surface targets to be detected. The pure endmembers are selected, considering the V-I-S (Vegetation- impervious surface- soil) model, vegetation and soil are taken as background while road and roofs as targets, therefore other details are masked out leaving only built-up area (Ridd 1995), (Phinn, et al. 2002).

4.1.4. Creation of Training File

A training file of size 1600 samples derived from ground data and image is employed for making the network to learn. Wherein, 400 spectra each constitute the reflectance values of roof, road, vegetation and soil respectively. The Training and Testing sets are categorised as 75% and 25% of the total samples individually, where Training set constitutes 1200 and testing set contains 400 known samples. While implementing the algorithm for a particular target (road or roof), all pixels belonging to other category are treated as background or non-target.

4.1.5. Ground Reference Generation

The mean spectral signature is generated for each feature and assumed as pure endmember for generation of ground reference image of the scene under investigation, as shown in Figure 3. The five-class ground reference displays roads (red), roofs (yellow), vegetation (green), soil (brown) and water (masked in black). The abundance values of the considered endmembers are generated using Dirichlet distribution followed by VCA (mentioned in Subsection 2.2.) for assigning the label to particular classes. As engineered surfaces, particularly roads and roofs are to be extracted, therefore ground reference with only built-up area is generated for implementation of CNN [refer Figure 3]. Also, ground reference image is georeferenced and overlayed on Google Earth imagery, validated against 59 and 75 locations belonging to roads and roofs respectively (collected during field data acquisition).

4.2. Target detection using CNN

4.2.1. Dimensionality reduction through PCA

Higher dimensionality of hyperspectral satellite imagery may lead to computation burden in terms of time and space complexity. PCA transforms the data and reduces it to components contributing maximum for the detection accuracy. Also, neighbouring bands of hyperspectral imagery are highly correlated and often redundant, so optimum combination of bands accounting for variation in terms of target detection are identified. For the proposed work, the number of components is confined to 30,40 and 50 as the information content is indirectly proportional to the number of principal components because most of the information is comprised in initial bands (Rodarmel and Shan 2002).

4.2.2. Architecture of CNN

For extraction of roads and roofs, supervised CNN is implemented that utilizes spectral data generated from AVIRIS-NG hyperspectral imagery. The complete stage constitutes a convolution layer, non-linear activation operation and a drop-out layer, which is replicated again to formulate a deep network, followed by fully connected layer. The user defined filter is convolved over a

Accepted Manuscript

fragment of input data, for identifying features from the previous layers to yield the detected targets. The network is made to learn these features with the supporting non-linear operation, here Rectified Linear Unit (ReLU) trailed by drop-out layer which summarises the output of multiple neurons. Drop-out layer is active during training process while inactive during evaluation acting as a regularization technique. It reduces, (i) overfitting of the network, (ii) superfluous feature dependencies and (iii) spatial size of the output generated. The dropout value is varied with respect to the regularization required, for the initial stage it is set as 0.25 and 0.5 for the later stage.

In order to ensure parallel processing, the image data is alienated into three-dimensional patches of size *'n'* (where n = 5,7 and 9) which are clustered together in form of small batches, acting as initial input to CNN. The layer contains filters of size *'d'* (where d = 3), stride *'s'* set as 1 and padding *'p'* of size 2, with ReLu function to be performed. After another replication, the output generated is flattened and sent to the fully connected layer. For the current work, two fully connected layers are taken generating the output using Equation 5, with a dropout layer of rate 0.25 and 0.5 respectively.

The labelled spectral data is split in ratio 75:25, where 75% of the complete data constitutes the training file and 25% is the testing ratio. Other hyperparameters encompassing, learning rate is set as 0.0001, momentum as 0.9, epoch as 15. Training parameters like performance error, epochs, momentum, learning rate, etc. have been tuned by trial and error in order to improve accuracy of the network. The architecture is explained in Figure 4.

4.3. Performance Evaluation

The performance evaluation metrics used to evaluate the approach are precision, recall and F1 along with binary output images, where bright pixels represent the presence of target (road or roof) and dark pixels depict the non-target or background region. True positive (TP) and true negatives (TN) are the observations that are correctly predicted and we want to minimize false positives (FP) and false negatives (FN). Precision is the ratio of correctly predicted positive targets to the total predicted positive observations. Recall is the ratio of correctly predicted positive observations to the all observations in actual class. F1 Score is the weighted average of Precision and Recall as it helps in balancing the both parameters. The equations supporting them as given from Equation numbers 8 to 10.

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

$$2 * (Recall \times Precision) \,/\, (Recall + Precision) \quad (10)$$

4.4. Implementation Background

The software used to carry out VCA for generation of ground reference image is MATLAB 2018 b (the code is from (J. M. Nascimento 2005)- is reproduced) and Python is used for implementation of CNN using keras library (the code from (Makantasis 2015) is amended for the area of interest). The architecture is implemented using Intel® Core ™ i7-8550U CPU with 16 GB RAM, NVIDIA GeForce MX130. The block diagram of the implementation is shown in Figure 5.

5 Results and Discussion

Accepted Manuscript

In intermediate past, researchers have proposed and developed multiple techniques to detect roads using hyperspectral data. The approaches are a combination of spectral, spatial, textural features, along with some machine-learning and deep learning methods. Classification of road surfaces on spectral basis often generates spectral similarities between specific roof types and asphalt roads. Also shadows of vegetation fencing the roads may cover the road surfaces. Highlighting the limitation of AVIRIS data to extract the centerline of roads, Herold et al pointed to associate contextual information for better accuracy (Herold, Gardner, et al. 2003). Object based recognition of roads and roofs is also carried out as an alternative approach using Binary Partition Tree (BPT) to delineate the roads and buildings; dealing with the problem of spectral variability (Valero, et al. 2011). Man, et al have used fusion based approach by combining hyperspectral and LiDAR data and extract multiple urban features such as roads, railway lines, residential areas etc. using SVM (Man, Dong and Guo 2015). The common inferences that may be drawn from the literature is that, challenges like spectral similarity, spectral variability still pertain in airborne hyperspectral data. In reference to road extraction, deep learning approaches are yet to be traversed, due to advantages like automatic learning from the data.

Modern roofs in urban areas can be made of metal sheets, asbestos, asphalt, cement, clay etc., depending upon the climatic and architectural factors. Techniques like discriminant analysis (Chisense, Hahn and Engels 2012), SAM, SVM (Szabó, Burai, et al. 2014), (Taherzadeh and Shafri 2011) are already mentioned in literature to identify the roofs and their condition. Since durable materials are required for urban construction, there is a scope to eliminate the need of handcrafted spectral and spatial features to automate the complete extraction procedure.

This paper presents an alternative approach for the extraction of engineered surfaces. In the process of detection of these, a series of experiments have been implemented and a comparative assessment has also been carried out. The dataset used for this study is already discussed in Section 3. SAC, ISRO provided Level 2 which is atmospherically corrected reflectance data. Besides, field data was also collected using field spectroradiometer for different types of engineered surfaces including road and roof. Collection of field data was synchronized with AVIRIS-NG data collection campaign. The approach for detection of roads and roofs along with their accuracy table and resultant imagery are shown in this section. The white pixels correspond to the presence of targets whereas the black pixels represent non-target or background.

5.1. Detection of Roads

The images in Figure 6 show the results obtained in different window size (5,7 and 9) with varying PCA components (30,40 and 50) and the prediction of each pixel was performed by assigning values as 1 for target and 0 for non-target class. Window size 11 has also been considered but due to minimal performance, the results for the same are not shown.

5.1.1. Window Size 5 × 5

The overall accuracy for considering the patch as 5, results into a balance between precision and recall [Table 3], the street roads along with major roads are detected in the area considered, whereas some traces of bare soil are also detected. Still, CNN is widely used for image recognition because it is invariant to small rotation and shifts (Hinton 2012). Fluctuating the PCA components from 30,40 and then 50 is having a hardly noticeable change in the illustrated results in Figure 6 (a, b, c) and also in accuracy assessment.

5.1.2. Window Size 7 × 7

Accepted Manuscript

For detection of roads, window size of 7 × 7 with 50 components seems to give the maximum value for performance evaluation parameters [Refer Table 3 and Figure 6 [d, e, f]]. The reason articulated for this is that roads belong to linear features and due to spatial resolution of the AVIRIS-NG sensor the roads are best detected in 7 × 7 patch size, whereas there is very less perceptible decrease in accuracy corresponding to 5 × 5 and 9 × 9 window. The complex variations in urban materials make pixel by pixel analysis difficult, therefore such robust and invariant features are required. Unlike Zhang et al who took a window size of 3 × 3 with lowest root mean square error (RMSE) for urban dataset (https://rslab.ut.ac.ir/data) using their approach to apply pixel based and cube-based CNN (Zhang, et al. 2018).

5.1.3. Window Size 9 × 9

On visually examining the resultant images, shows detection of additional features similar to roads on spectral grounds [Figure 6 [g, h, i]] and the decrease in overall extraction rate. Another observation derived from the results is, increase in PCA derived components may lead to decrease in accuracy of the model after some extent, as only initial components are involved maximum in the detection process. PCA is introduced to abridge the whole image which reduces the data dimension to an acceptable scale (here, 30, 40 and 50 components).

Less variation is captured in the resultant images while changing the window size, on visual examination. The proposed approach has consistent performance but have confused few pixels of roads and soil at certain positions, it detects patches of soil along with extraction of roads on top right side of the images. Since the material used for construction of roads may include soil, therefore it is detected at some parts. One of the major challenges in target detection for hyperspectral imagery include presence of targets in different sizes, different orientations and at very close positions (Khan, et al. 2017). Also, CNN exploits the advantage of large data samples, which requires more computational resources to extract features for better data fitting and higher detection performance (Zhang, et al. 2020). Therefore, at the time of implementation, though the training time is compromised but the method effectively improves the detection rate.

5.2. Detection of Roofs

The visual analysis of the images illustrated in Figure 7 represents detection of roofs as clusters in urban areas, using CNN and Table 4 shows the corresponding performance evaluation metrics.

5.2.1. Window Size 5 × 5

Patch of 5 with 30 components appears to be the most suitable for extraction of roofs for the current dataset used for the study [Figure 7 [a, b, c] and Table 4]. Roofs appear in the form of clusters; therefore, smaller size window is able to detect most of them. Also, the computation time for smaller window and a smaller number of components is comparatively low in contrast with other experiments considered. Sun et al have proposed to extract impervious surfaces with 3 D – CNN using WorldView-2 and airborne LiDAR data with a Producer's Accuracy (PA) maximum of 91% for combined of impervious surfaces (Sun, et al. 2019). In contrast, the maximum F1 value generated for roof detection for this approach and dataset is estimated as 0.93.

5.2.2. Window Size 7 × 7

Consequently, the patch size of 7 also shows remarkable results for detecting the roof surfaces on pictorial grounds, as given in Figure 7 (d, e, f). Roof being high albedo reflects

Accepted Manuscript

maximum energy falling on them, creating a confusion with other features such as barren lands, which might get detected additionally in big window sizes.

5.2.3. Window Size 9 × 9

The window of size 9 appears to be large for detecting the small roofs when number of components are increased to 50 as demonstrated in Figure 7 (i). PCA components 30 and 40 exemplifies marginal difference in terms of detection [Table 4].

CNN has been recently explored for hyperspectral data with ease of implementation, preferred for large datasets, minimal human intervention generates high accuracy. But with increase in window size, computation burden in terms of time complexity also intensifies.

6 Conclusion

This paper presents an alternative approach for extraction of engineered surfaces using hyperspectral data. The problem of extraction of roads and roofs has been successfully dealt using deep learning approach. VCA algorithm which is used for spectral unmixing as reported in literature, is used for generating the ground reference which is further supplied to CNN as input. In this work, CNN has been implemented using recently acquired hyperspectral AVIRIS- NG data of Udaipur region in India.

The algorithm has been compared under three categories of patch sizes, (i) 5 × 5 (ii) 7 × 7 (iii) 9 × 9 where roads and roofs have been treated as targets. It is observed that increase in the window size (5,7,9) and the number of components (30,40,50) often degrades the performance of the target detector. By visual examination of the results, the highway roads, street roads made up of asphalt, bitumen are mostly detected, whereas unmetalled or dirt roads are confused with bare soil. Similarly for roofs, flat surfaces appear to be detected completely, while there is a pixel loss in case of high or low slope roofs found in industrial areas. Besides, the spatial resolution of data used in these experiments is 8.1 meters and this might be insufficient for detection of engineered surfaces as their scale is much smaller. A sharp delineating boundary between roads and roofs is therefore not seen/detected in the resultant mages.

Deep learning methods appear to be more attractive and show their tremendous potential in detecting targets with less execution time and high accuracy rate. Their learning capability from known data and prediction for unknown data make them outperform among all the algorithms. Overall, experimentation appears to perform uniformly, for detection of roads and roofs. Apart from this, material used for construction of roads and roofs is also similar, such as concrete, asphalt, bitumen which could be regarded as another reason for lesser detection rate in some cases. Removal of redundant bands is needed to realize a better target detection system. Dimensionality reduction methods, other than PCA can also be explored for decreasing the computation time for implementation.

As observed, consideration of only spectral aspect leads to confusion in classes made up of similar materials, therefore spatial attributes which exhibit strong dependencies in hyperspectral data shall be considered for enhanced detection in future work. The study is also bounded till detection of roads and roofs only, further their condition with respect to aging can also be determined. The proposed approach assumes targets as full pixel; however, mixed pixels are also encountered during aerial data collection due to sensor configuration. In recent times, many feature selection and extraction techniques have been developed which can be used in association of deep learning measures for real-time processing of data.

Accepted Manuscript

Disclosures

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Acknowledgements

The work is in collaboration with Defence Research and Development Organization (DRDO) and is sponsored by the Department of Science and Technology (DST). The authors extend their gratitude to ISRO (Indian Space and Research Organization) for supporting the project. This research was funded by Ministry of Science & Technology, Department of Science and Technology (DST), grant number BDID/01/23/2014-HSRS.

Data Availability Statement

Data subject to third party restrictions: The data that support the findings of this study are available from ISRO. Restrictions apply to the availability of these data, which were used under license for this study. Data are available at https://avirisng.jpl.nasa.gov/aviris-ng.html with the permission of Jet Propulsion Laboratory (JPL), National Aeronautics and Space Administration (NASA).

Disclosure Statement

I hereby declare that, there are no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Accepted Manuscript

References

Akbari, Davood, Saeid Homayouni, Abdolreza Safari, and Naser Mehrshad. 2016. "Mapping urban land cover based on spatial-spectral classification of hyperspectral remote-sensing data." *International Journal of Remote Sensing* 37 (2): 440-454.

Arindam, Guha. 2020. "Mineral exploration using hyperspectral data." In *Mineral exploration using hyperspectral data*, 293-318. Elsevier.

Benediktsson, Jón Atli, Jón Aevar Palmason, and Johannes R. Sveinsson. 2005. "Classification of hyperspectral data from urban areas based on extended morphological profiles." *IEEE Transactions on Geoscience and Remote Sensing* 43 (3): 480-491.

Bhangale, Ujwala M., Kuldeep R. Kurte, Surya S. Durbha, Roger L. King, and Nicolas H. Younan. 2016. "Big data processing using hpc for remote sensing disaster data." *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Beijing, China: IEEE. 5894-5897.

Boardman, Joseph W., Fred A. Kruse, and Robert O. Green. 1995. *Mapping target signatures via partial unmixing of AVIRIS data*. https://trs.jpl.nasa.gov/bitstream/handle/2014/33635/94-1613.pdf?sequence=1.

Burgers, Kate, Yohannes Fessehatsion, Sheida Rahmani, and and T. Wittman. J. Seo. 2009. "A comparative analysis of dimension reduction algorithms on hyperspectral data." *LAMDA Research Group* 1-23.

Cheng, Yu, Duo Wang, Pan Zhou, and Tao Zhang. 2017. *A survey of model compression and acceleration for deep neural networks.* arXiv preprint arXiv:1710.09282.

Chisense, C., M. Hahn, and J. Engels. 2012. "Classification of roof materials using hyperspectral data." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 39 (B7): 103.

Deyu, An, Gengxing Zhao, Chunyan Chang, Zhuoran Wang, Ping Li, Tongrui Zhang, and Jichao Jia. 2016. "Hyperspectral field estimation and remote-sensing inversion of salt content in coastal saline soils of the Yellow River Delta." *International Journal of Remote Sensing* 37 (2): 455-470.

Farrell, Michael D, and Russell M. Mersereau. 2005. "On the impact of PCA dimension reduction for hyperspectral detection of difficult targets." *IEEE Geoscience and Remote Sensing Letters* 2 (2): 192-195.

Fen, Chen, Huajun Jiang, Tim Van de Voorde, Sijia Lu, Wenbo Xu, and Yan Zhou. 2018. "Land cover mapping in urban environments using hyperspectral APEX data: A study case in Baden, Switzerland." *International journal of applied earth observation and geoinformation* 70-82.

Gewali, Utsav B., Monteiro Sildomar T, and Eli Saber. 2018. "Machine learning based hyperspectral image analysis: a survey."

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning (Adaptive Computation and Machine Learning series).*

Hang, Renlong, Qingshan Liu, Danfeng Hong, and Pedram Ghamisi. 2019. "Cascaded recurrent neural networks for hyperspectral image classification." *IEEE Transactions on Geoscience and Remote Sensing* 57 (8): 5384-5394.

Haut, Juan Mario, Mercedes E. Paoletti, Javier Plaza, Antonio Plaza, and Jun Li. 2019. "Visual attention-driven hyperspectral image classification." *IEEE Transactions on Geoscience and Remote Sensing* 57 (10): 8065-8080.

Hegde, Gaurav, J. Mohammed Ahamed, R. Hebbar, and Uday Raj. 2014. "Urban land cover classification using hyperspectral data." *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 8 (8): 751-754.

Heiden, Uta, Sigrid Roessner, Karl Segl, and Hermann Kaufmann. 2001. "Analysis of spectral signatures of urban surfaces for their identification using hyperspectral HyMap data." *IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas.* 173-177.

Herold, Martin, Dar A. Roberts, Margaret E. Gardner, and Philip E. Dennison. 2004. "Spectrometry for urban area remote sensing—Development and analysis of a spectral library from 350 to 2400 nm." *Remote Sensing of Environment* 91 (3-4): 304-319.

Herold, Martin, M. Gardner, Val Noronha, and D. Roberts. 2003. "Spectrometry and hyperspectral remote sensing of urban road infrastructure." *Online Journal of Space Communications* 3: 1-29.

Hinton, Geoffrey E. 2012. "A practical guide to training restricted Boltzmann machines." *Neural networks: Tricks of the trade* 599-619.

Hively, W. Dean, Gregory W. McCarty, James B. Reeves, Megan W. Lang, Robert A. Oesterling, and Stephen R. 2011. "Use of airborne hyperspectral imagery to map soil properties in tilled agricultural fields." *Applied and Environmental Soil Science.*

Hu, Wei, Yangyu Huang, Li Wei, Fan Zhang, and Hengchao Li. 2015. "Deep convolutional neural networks for hyperspectral image classification." *Journal of Sensors.*

Hu, Wei, Yangyu Huang, Li Wei, Fan Zhang, and Hengchao Li. 2015. "Deep convolutional neural networks for hyperspectral image classification." *Journal of Sensors.*

Huang, Hong, Guangyao Shi, Haibo He, Yule Duan, and Fulin Luo. 2019. "Dimensionality reduction of hyperspectral imagery based on spatial–spectral manifold learning." *IEEE Transactions on Cybernetics* 50 (6): 2604-2616.

Hycza, Tomasz, Krzysztof Stereńczak, and Radomir Bałazy. 2018. "Potential use of hyperspectral data to classify forest tree species." *New Zealand Journal of Forestry Science* 48 (1): 18.

James, Falcone A., and Gomez Richard. 2005. "Mapping impervious surface type and sub-pixel abundance using hyperion hyperspectral imagery." *Geocarto International* 20 (4): 3-10.

Kale, Karbhari V., Mahesh M. Solankar, Dhananjay B. Nalawade, Rajesh K. Dhumal, and and Hanumant R. Gite. 2017. "A research review on hyperspectral data processing and analysis algorithms." *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences.* 541-555.

Accepted Manuscript

Kallepalli, A., A. Kumar, and K. Khoshelham. 2014. “Entropy based determination of optimal principal components of Airborne Prism EXperiment (APEX) imaging spectrometer data for improved land cover classification.” 781-786.

Keith, Darryl J., Blake A. Schaeffer, Ross S. Lunetta, Richard W. Gould Jr, Kenneth Rocha, and Donald J. Cobb. 2014. “Remote sensing of selected water-quality indicators with the hyperspectral imager for the coastal ocean (HICO) sensor.” *International Journal of Remote Sensing* 35 (9): 2927-2962.

Kerekes, John P., and Daniel S. Goldberg. 2013. “Hyperspectral image quality for unmixing and subpixel detection applications.” *Image Quality and System Performance* (International Society for Optics and Photonics) 8653: 865304.

Khan, Muhammad Jaleed, Adeel Yousaf, Nizwa Javed, Shifa Nadeem, and Khurram Khurshid. 2017. “Automatic target detection in satellite images using deep learning.” *Journal of Space Technology* 7 (1): 44-49.

Kruse, F. A., J. W. Boardman, A. B. Lefkoff, J. M. Young, K. S. Kierein-Young, T. D. Cocks, R. Jensen, and P. A. Cocks. 2000. “HyMap: an Australian hyperspectral sensor solving global problems-results from USA HyMap data acquisitions.” *InProc. of the 10th Australasian Remote Sensing and Photogrammetry Conference.* 18-23.

Lemp, Dirk and Uwe Weidner. 2005. “Improvements of roof surface classification using hyperspectral and laser scanning data.” *Int. Symp. Remote Sens. Data Fusion Over Urban Areas (URBAN).* ISPRS. 14-16.

López, Peña Fernando, Juan Luís Crespo, and Richard J. Duro. 2009. “Unmixing low-ratio endmembers in hyperspectral images through Gaussian synapse ANNs.” *IEEE Transactions on Instrumentation and Measurement* 59 (7): 1834-1840.

Makantasis, Konstantinos, Konstantinos Karantzalos, Anastasios Doulamis, and Nikolaos Doulamis. 2015. “Deep supervised learning for hyperspectral data classification through convolutional neural networks.” IEEE. 4959-4962.

Man, Qixia, Pinliang Dong, and Huadong Guo. 2015. “International Journal of Remote Sensing.” *Pixel-and feature-level fusion of hyperspectral and lidar data for urban land-use classification* 36 (6): 1618-1644.

Manolakis, Dimitris, David Marden, and Gary A. Shaw. 2003. “Hyperspectral image processing for automatic target detection applications.” *Lincoln laboratory journal* 14 (1): 79-116.

Mei, Shaohui, Jingyu Ji, Junhui Hou, Xu Li, and Qian Du. 2017. “Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks.” *IEEE Transactions on Geoscience and Remote Sensing* 55 (8): 4520-4533.

Mou, Lichao, Pedram Ghamisi, and Xiao Xiang Zhu. 2017. “Deep recurrent neural networks for hyperspectral image classification.” *IEEE Transactions on Geoscience and Remote Sensing* 55 (7): 3639-3655.

Nascimento, José MP, and José M. Bioucas-Dias. 2011. “Hyperspectral unmixing based on mixtures of Dirichlet components.” *IEEE Transactions on Geoscience and Remote Sensing* 50 (3): 863-878.

Nascimento, José MP, and José MB Dias. 2005. “Vertex component analysis: A fast algorithm to unmix hyperspectral data.” *IEEE transactions on Geoscience and Remote Sensing* 43 (4): 898-910.

Palmason, Jon Aevar, Jon Atli Benediktsson, Johannes R. Sveinsson, and Jocelyn Chanussot. 2005. “"Classification of hyperspectral data from urban areas using morphological preprocessing and independent component analysis.” *IEEE International Geoscience and Remote Sensing Symposium, IGARSS'05.* IEEE. 4.

Paoletti, M. E., J. M. Haut, J. Plaza, and A. Plaza. 2019. “Deep learning classifiers for hyperspectral imaging: A review.” *ISPRS Journal of Photogrammetry and Remote Sensing* 158: 279-317.

Pham, Hai Minh, Yasushi Yamaguchi, and Thanh Quang Bui. 2011. “A case study on the relation between city planning and urban growth using remote sensing and spatial metrics.” *Landscape and Urban Planning* 100 (3): 223-230.

Phinn, S., M., P. Scarth Stanford, A. T. Murray, and P. T. Shyy. 2002. “Monitoring the composition of urban environments based on the vegetation-impervious surface-soil (VIS) model by subpixel analysis techniques.” *International Journal of Remote Sensing* 23 (20): 4131-4153.

Pu, Hongbin, Dan Liu, Jia-Huan Qu, and Da-Wen Sun. 2017. “Applications of imaging spectrometry in inland water quality monitoring—a review of recent developments.” *Water, Air, & Soil Pollution* 228 (4): 131.

Accepted Manuscript

Reza, Seifi Majdar, and Hassan Ghassemian. 2017. “A probabilistic SVM approach for hyperspectral image classification using spectral and texture features.” *International Journal of Remote Sensing* 38 (15): 4265-4284.

Ridd, Merrill K. 1995. “Exploring a VIS (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities.” *International journal of remote sensing* 16 (12): 2165-2185.

Robila, Stefan. 2005. “An investigation of spectral metrics in hyperspectral image preprocessing for classification.” *In Geospatial goes global: from your neighborhood to the whole planet, ASPRS Annual Conference.* Baltimore, Maryland. 7-11.

Rodarmel, Craig, and Jie Shan. 2002. “Principal component analysis for hyperspectral image classification.” *Surveying and Land Information Science* 62 (2): 115-122.

Shi, Qian, Xiaoping Liu, and Xia Li. 2017. “Road detection from remote sensing images by generative adversarial networks.” *IEEE access* 25486-25494.

Shwartz-Ziv, Ravid, and Naftali Tishby. 2017. *Opening the black box of deep neural networks via information.* arXiv preprint arXiv:1703.00810 .

Su, Jinya, Dewei Yi, Cunjia Liu, Lei Guo, and Wen-Hua Chen. 2017. “Dimension reduction aided hyperspectral image classification with a small-sized training dataset: experimental comparisons.” *Sensors* 17 (12): 2726.

Sun, Zhongchang, Xiangwei Zhao, Mengfan Wu, and Cuizhen Wang. 2019. “Extracting urban impervious surface from worldView-2 and airborne LiDAR data using 3D convolutional neural networks.” *Journal of the Indian Society of Remote Sensing* 47 (3): 401-412.

Szabó, Szilárd, Péter Burai, Zoltán Kovács, and György Szabó. 2014. “Testing algorithms for the identification of asbestos roofing based on hyperspectral data.” *Environmental Engineering and Management Journal* 143 (11): 2875-2880.

Szabó, Szilárd, Péter Burai, Zoltán Kovács, György Szabó, and Attila Kerényi. 2014. “Testing algorithms for the identification of asbestos roofing based on hyperspectral data.” *Environmental Engineering and Management Journal* 143 (11): 2875-2880.

Szabó, Szilárd, Péter Burai, Zoltán Kovács, György Szabó, Attila Kerényi, István Fazekas, and Tamás Buday, and Gergely Szabó. "." Mónika Paládi. 2014. “Testing algorithms for the identification of asbestos roofing based on hyperspectral data.” *Environmental Engineering and Management Journal* 143 (11): 2875-2880.

Taherzadeh, Ebrahim, and Helmi ZM Shafri. 2011. “Using hyperspectral remote sensing data in urban mapping over Kuala Lumpur.” *Joint Urban Remote Sensing Event,*. IEEE. 405-408.

Theiler, James, Amanda Ziemann, Stefania Matteoli, and Marco Diani. 2019. “Spectral variability of remotely sensed target materials: causes, models, and strategies for mitigation and robust exploitation.” *IEEE Geoscience and Remote Sensing Magazine* 7 (2): 8-30.

Thompson, David R., Boardman Joseph W., Michael L. Eastwood, Robert O. Green, Justin M. Haag, Pantazis Mouroulis, and Byron Van Gorp. 2018. “Imaging spectrometer stray spectral response: In-flight characterization, correction, and validation.” *Remote Sensing of Environment* 204: 850-860.

Thompson, David R., Bo-Cai Gao, Robert O. Green, Dar A. Roberts, Philip E. Dennison, and Sarah R. Lundeen. 2015. “Atmospheric correction for global mapping spectroscopy: ATREM advances for the HyspIRI preparatory campaign.” *Remote Sensing of Environment* 167: 64-77.

Tiwari, Kailash C., Manoj Kumar Arora, Dharmendra P. Singh, and Deepti Soni Yadav. 2013. “Military target detection using spectrally modeled algorithms and independent component analysis.” *Optical Engineering* 52 (2): 026402.

Uta, Heiden, Karl Segl, Sigrid Roessner, and Hermann Kaufmann. 2007. “Determination of robust spectral features for identification of urban surface materials in hyperspectral remote sensing data.” *Remote Sensing of Environment* 111 (4): 537-552.

Valero, Silvia, Philippe Salembier, Jocelyn Chanussot, and Carles M. Cuadras. 2011. “Improved binary partition tree construction for hyperspectral images: Application to object detection.” IEEE. 2515-2518.

Accepted Manuscript

Van der Linden, S., and P. Hostert. 2009. “The influence of urban structures on impervious surface maps from airborne hyperspectral data.” *Remote Sensing of Environment* 113 (11): 2298-2305.

Weber, Cody, Rahim Aguejdad, Xavier Briottet, Sophie Fabre J. Avala, Jean Demuynck, and Emmanuel Zenou et al. 2018. “Hyperspectral imagery for environmental urban planning.” *IGARSS IEEE International Geoscience and Remote Sensing Symposium.* IEEE. 1628-1631.

Windrim, Lloyd, Rishi Ramakrishnan, Arman Melkumyan, Richard J. Murphy, and Anna Chlingaryan. 2019. “Unsupervised feature-learning for hyperspectral data with autoencoders.” *Remote Sensing* 11 (7): 864.

Winter, Michael E. 1999. “N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data In Imaging Spectrometry .” *International Society for Optics and Photonics* 5 (3753): 266-275.

Yi, Wang, and and Hexiang Duan. 2018. “Classification of hyperspectral images by SVM using a composite kernel by employing spectral, spatial and hierarchical structure information.” *Remote Sensing* 10 (3): 441.

Zhang, Gaigai, Shizhi Zhao, Wei Li, Qian Du, Qiong Ran, and Ran Tao. 2020. “HTD-net: A deep convolutional neural network for target detection in hyperspectral imagery.” *Remote Sensing* 12 (9): 1489.

Zhang, Xiangrong, Yujia Sun, Jingyan Zhang, Peng Wu, and Licheng Jiao. 2018. “Hyperspectral unmixing via deep convolutional neural networks.” *IEEE Geoscience and Remote Sensing Letters* 15 (11): 1755-1759.

Zhao, Wenzhi, Zhou Guo, Jun Yue, Xiuyuan Zhang, and Liqun Luo. 2015. “On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery.” *International Journal of Remote Sensing* 36 (13): 3368-3379.

Zhu, Feiyun. 2017. *Hyperspectral unmixing: ground truth labeling, datasets, benchmark performances and survey.* arXiv preprint arXiv:1708.05125

Accepted Manuscript

Table 1. Challenges and Research Gaps Identified in the field.

S No.	Domain	Challenges and Research Gaps
1.	Urban Target Detection	1.1. Major research is focused on natural targets such as vegetation, soil, minerals. 1.2. Small size of urban targets such as roofs, pavements etc 1.3. Higher within class variability and less inter class variability 1.4. Heterogeneity within urban environment
2.	Hyperspectral remote sensing	2.1. Curse of dimensionality 2.2. Non-availability of corresponding ground truth 2.3. Interference of multiple factors (sensor, scene related and atmospheric factors) 2.4. Spectral variability 2.5. Unavailability of standard datasets
3.	Deep learning approach	3.1. Complex and time-consuming training process 3.2. High computation complexity 3.3. Overfitting of developed model 3.4. Selection of hyper-parameters for high performance 3.5. Blackbox nature of training procedure 3.6. Lack of labelled or known samples

Accepted Manuscript

Table 2: Description of Image and Ground Data Collection.

Parameters	**Image Data Description**	**Ground Data Description**
Location	Udaipur, Rajasthan, India	
Date of Data acquisition	02/02/2016	16/07/2017 − 21/07/2017
Spatial Resolution	8.1 meter	-
Spectral Resolution	5 nm	3nm(700 nm),8nm(1500 nm), 6 nm(2100 nm)
Channels	425	2151
Level of data	Level 1 and Level 2	
Main features in the images	Urban features such as roads, bridges, different types of roofs (Concrete, bitumen, sandstone, CGI sheet), vehicles, railway tracks etc.	

Accepted Manuscript

Table 3: Accuracy Assessment for detection of Roads as engineered surfaces.

S No	Window Size	PCA Components	Road		
			Precision	Recall	F1
1	5 × 5	30	0.93	0.90	0.92
		40	0.94	0.89	0.91
		50	0.92	0.91	0.92
2	7 × 7	30	0.92	0.93	0.92
		40	0.92	0.91	0.92
		50	0.94	0.89	0.92
3	9 × 9	30	0.90	0.88	0.89
		40	0.90	0.90	0.90
		50	0.34	1.00	0.51

Accepted Manuscript

Table 4: Accuracy Assessment for detection of Roofs as engineered surfaces.

S No	Window Size	PCA Components	Roof		
			Precision	Recall	F1
1	5 × 5	30	0.93	0.94	0.93
		40	0.91	0.95	0.93
		50	0.93	0.93	0.93
2	7 × 7	30	0.91	0.92	0.92
		40	0.93	0.93	0.93
		50	0.91	0.96	0.93
3	9 × 9	30	0.91	0.91	0.91
		40	0.92	0.91	0.92
		50	0.89	0.84	0.82

Accepted Manuscript

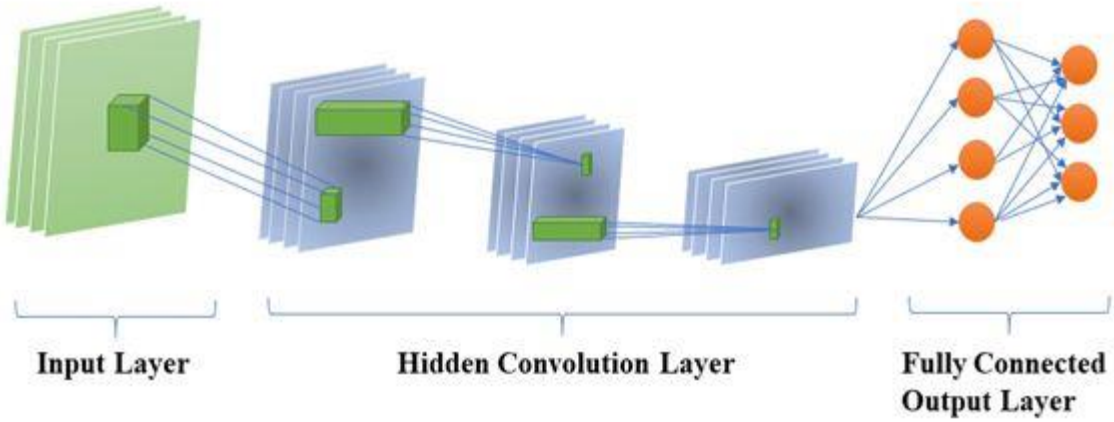


Figure 1: General Architecture of Convolution Neural Network.

Accepted Manuscript

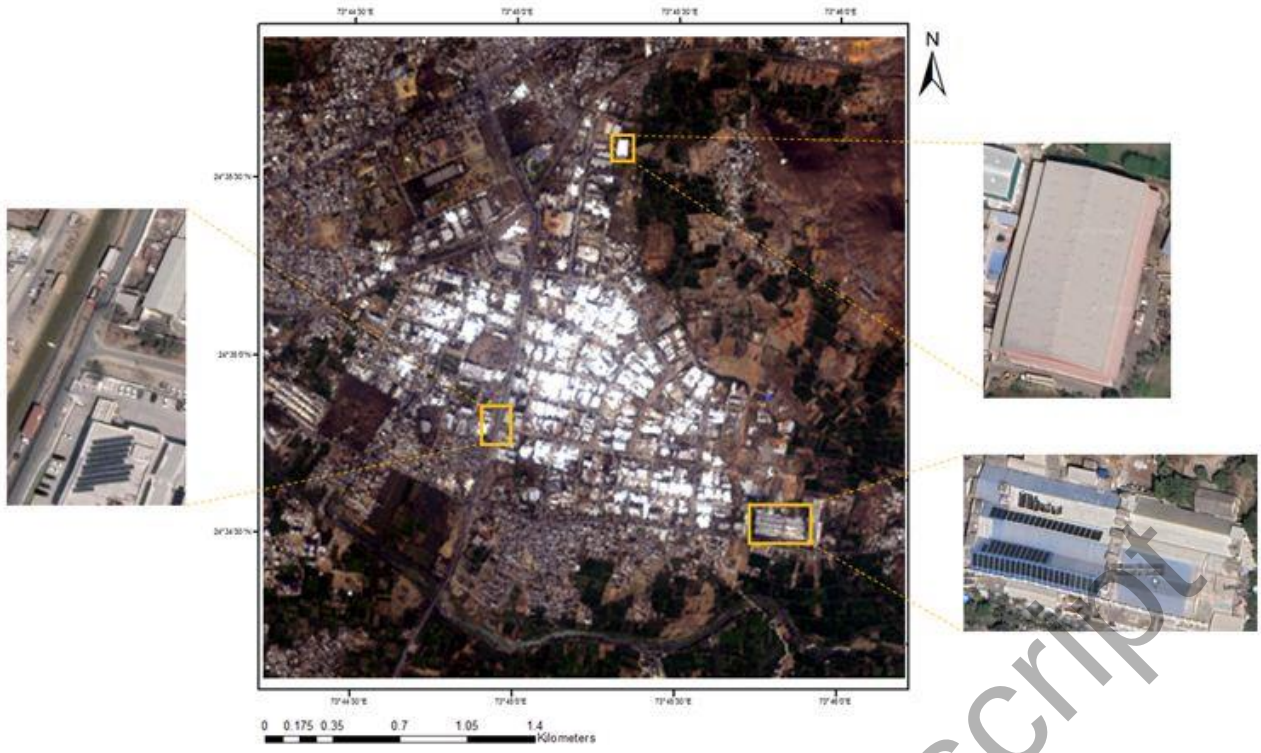


Figure 2. Image Dataset of AVIRIS-NG Hyperspectral data.

Accepted Manuscript

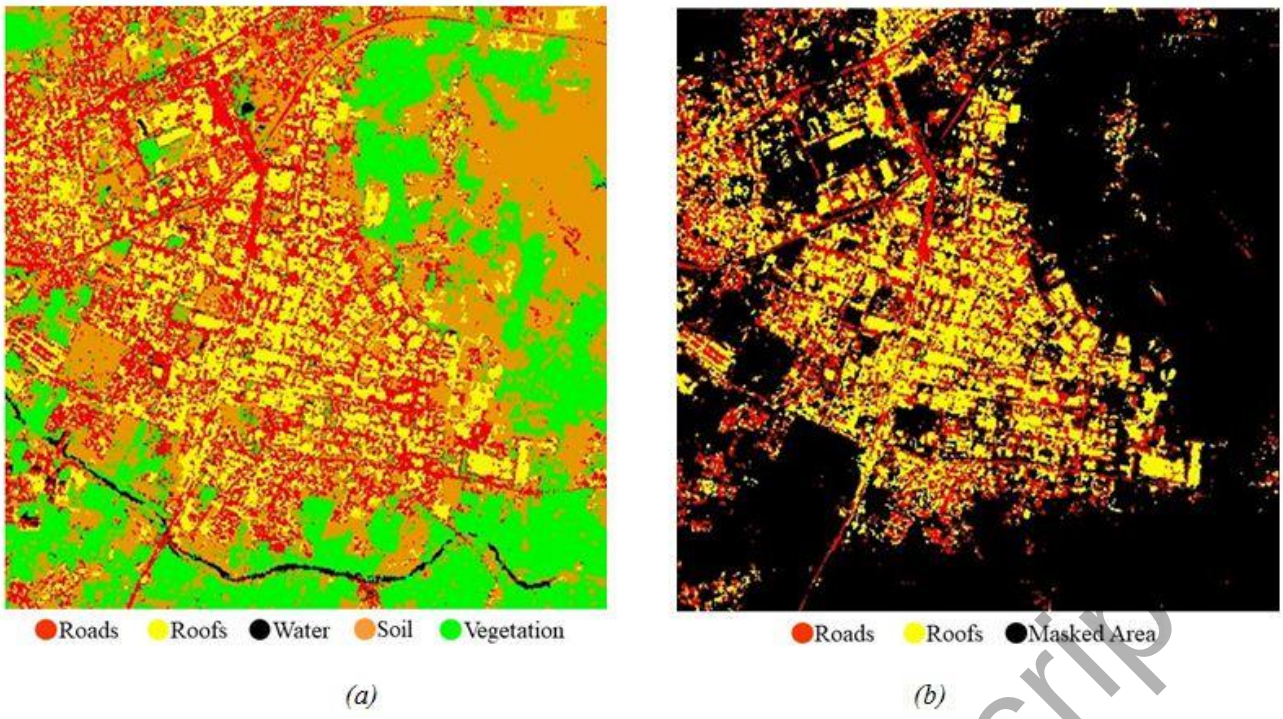


Figure 3. (a) Ground Reference generated using VCA for road, roof, vegetation and soil **(b)** Ground Reference for engineered surfaces (roads and roofs).

Accepted Manuscript

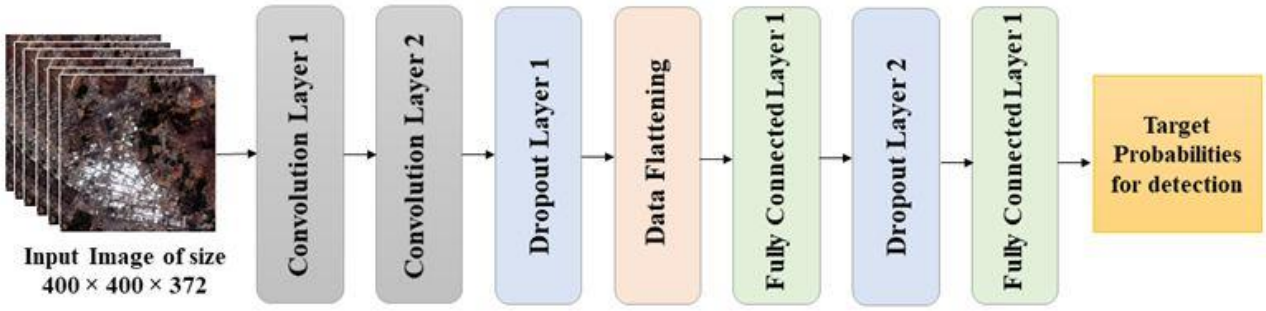


Figure 4. CNN Architecture for the proposed approach.

Accepted Manuscript

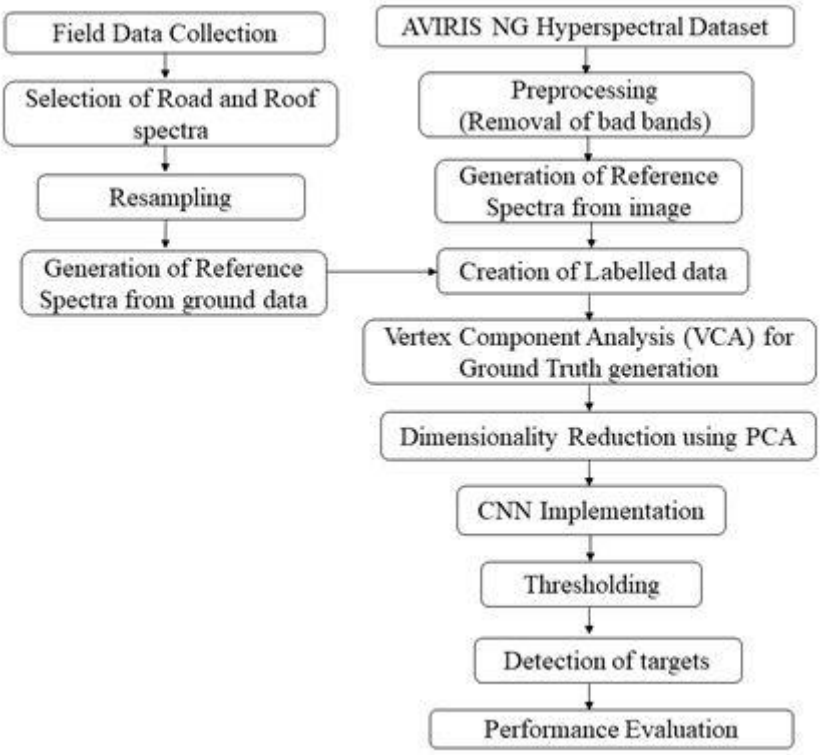


Figure 5. Block Diagram of Methodology.

Accepted Manuscript

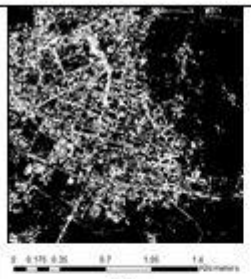
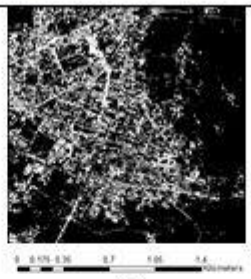
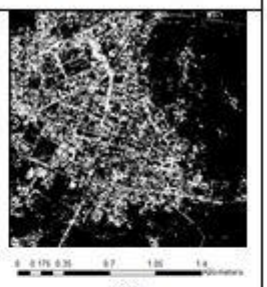
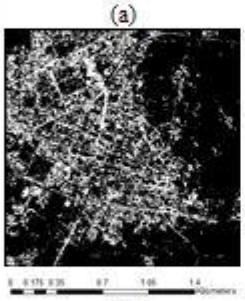
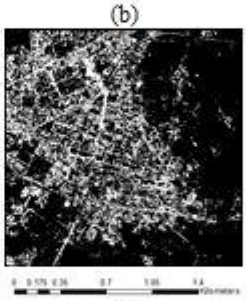
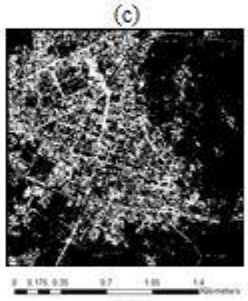
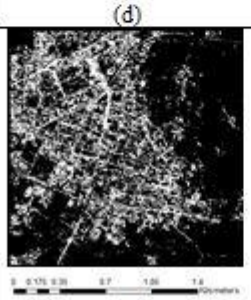
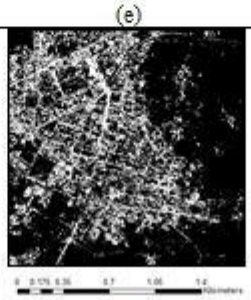
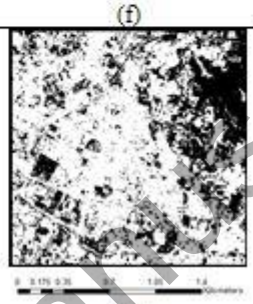
S No	Window Size	PCA Components = 30	PCA Components = 40	PCA Components = 50
1	5×5	 (a)	 (b)	 (c)
2	7×7	 (d)	 (e)	 (f)
3	9×9	 (g)	 (h)	 (i)

Figure 6. Detection of Roads using proposed approach.

Accepted Manuscript

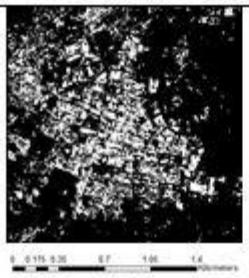
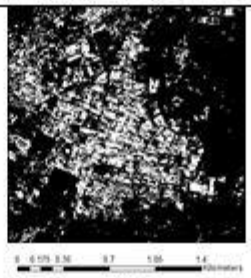
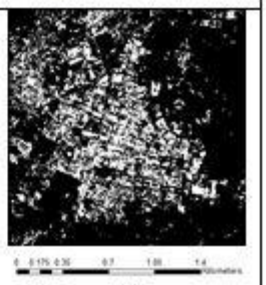
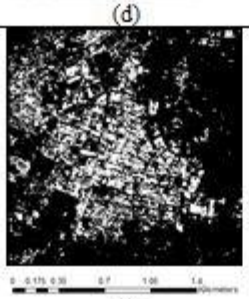
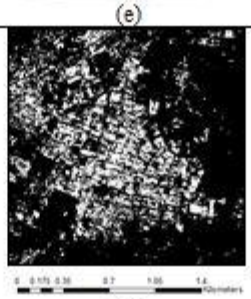
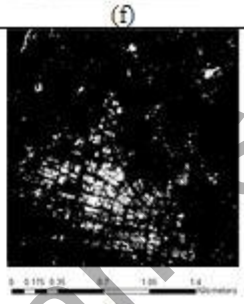
S No	Window Size	PCA Components = 30	PCA Components = 40	PCA Components = 50
1	5×5	 (a)	 (b)	 (c)
2	7×7	(d)	(e)	(f)
3	9×9	 (g)	 (h)	 (i)

Figure 7. Detection of Roofs using proposed approach.

Accepted Manuscript

2021 2nd International Conference for Emerging Technology (INCET)
Belgaum, India. May 21-23, 2021

Detection of Pulsars by Classical Machine Learning Algorithms

Deepanshu Beniwal
Department of Information Technology
Delhi Technological University
Delhi, India
beniwaldeepanshu442@gmail.com

Abhishek Roy
Department of Information Technology
Delhi Technological University
Delhi, India
royabhishek250@gmail.com

Himanshu Yadav
Department of Information Technology
Delhi Technological University
Delhi, India
h4himanshuyadav@gmail.com

Anamika Chauhan
Department of Information Technology
Delhi Technological University
Delhi, India
anamika@dtu.ac.in

2021 2nd International Conference for Emerging Technology (INCET) | 978-1-7281-7029-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/INCET51464.2021.9456250

***Abstract*—This paper compares the classical machine learning algorithms to classify a pulsar (rapidly spinning neutron star) in the HTRU-2 dataset [6]. This dataset was gathered during the survey, called the High Time Resolution Universe (South), and is an imbalanced dataset consisting of more than 17 thousand records. Being highly imbalanced, several methodologies have been implemented (such as sampling, resampling, and feature selection) to overcome the imbalance between target classes (binary in this case) to achieve higher accuracies. The results of the proposed work show that by applying different methodologies, it is possible to overcome the class imbalance of the dataset, which in turn helps the classical machine learning algorithms to obtain promising accuracies.**

Keywords—neutron star, pulsar, general relativity, gravitational waves, radio telescopes, binary classification, supervised machine learning, astronomy, sampling, feature selection

I. INTRODUCTION

Pulsars are quickly spinning neutron stars in a nutshell, having a diameter of about 19-20 km across. If the mass of our Sun is considered to be 1 Solar Masses, then the masses of Pulsar range between 1.18 and 1.97 Solar Masses, but most Pulsars have a mass of about 1.35 Solar Masses. In 1967, Jocelyn Bell was a postgraduate student when she was building a radio telescope. Bell noticed a noise pattern in the data recorded by their telescope, which they find unfamiliar. After many nights of analyzing those noise patterns, they finally get to know that the noise was coming from an astronomical body. She had discovered the first known pulsar. The rapid rotation of pulsars paired with their intense magnetic fields generates twin beams of energy away from their surface. As the star rotates, these beams get detected as a flash of increased brightness or pulse from Earth. This pulse carries the light of different spectrums such as visible, radio, and x-rays.

Pulsars are the leftover cores of massive stars that exploded in a supernova. Some pulsars spin with periods of a few milliseconds (called millisecond pulsar) to a couple of minutes. Some of them form a binary system with a nearby star. In rare cases, two pulsars or neutron stars get close enough such that they start orbiting each other and are known to be of very scientific importance. While orbiting each other, the neutron stars generate continuous ripples in the fabric of space-time, thus causing what we call gravitational waves, which according to cosmologists, has confirmed an important prediction of the general theory of relativity proposed by Albert Einstein in 1916.

Modern telescopes used to detect these bizarre objects produce terabytes of data per observation, pushing supercomputers to their limits and emerging as a problem for astronomers. Analyzing such data requires astronomers to be able to think computationally to solve problems. Thus, automation in detection and classification through machine learning is needed with a requirement of optimization to permit efficient and real-time throughput processing of data. The proposed work investigates the utility of supervised Machine learning in pulsar detection from the data fed into them.

The proposed paper is organized in the following manner: In Section-II, a summary of the studies previously done under pulsar detection, along with what new we're proposing in our work, is presented. Section-III explains the working methodology. In Section-IV, Experiments with setup and dataset variables description are presented. Results of the experiments are presented in Section-V. Based on the results obtained in the previous section, conclusions are made in Section-VI.

II. LITERATURE REVIEW

Researches in the field of astronomy have seen significant growth over the past couple of years. One of the reasons for this is the large amount of data generated by modern-day telescopes for every single observation they make, which astronomers have found quite overwhelming, as it is nearly impossible to deal with such an amount of data manually. Thus, computer scientists and algorithm experts who already have the required experience of handling such data have shown interest in automating astronomy-related problems. Under astronomy, the detection of pulsars is one such big problem. Pulsars are a rare type of stars and since they're rare, detecting them through petabytes of data generated is not an easy task.

Researchers and algorithm experts around the world have contributed some of their work towards this area of astronomy. Several Machine Learning and Artificial

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:25:49 UTC from IEEE Xplore. Restrictions apply.

Intelligence models are putting forward to help astronomers automating this task.

A Novel AI system called PICS or the Pulsar Image-based Classification System is introduced by [1], where pulsar detection is made possible through image pattern recognition with the help of a 9000 neurons-based deep-learning model. A survey pipeline called the PALFA or the Pulsar Arecibo L-Band Feed Array has integrated this system and, to date, has successfully discovered six new pulsars.

In [2], the authors have worked on pulsar classification with very few features, primarily focusing on feature selection procedures. They proposed two feature selection algorithms called GS (Grid Search) and RFE (Recursive Feature Elimination) and found promising recall and false-positive rates.

[3] Developed a novel approach for pulsar classification by extracting deep-features in pulsar candidate's image domain with DCGAN (deep convolutional generative adversarial networks) and by training a classifier developed using Multi-Layer Perceptron with PILAE (pseudo-inverse learning auto-encoder) algorithm.

Sometimes either due to interstellar medium or due to a farther remote location from the Earth, the signals received from a pulsar aren't strong enough to be used by traditional pulsar detection models based on FFT (Fast Fourier Transform) spectrum search. Thus, a different approach is required to overcome this weak strength of the signal. One such work is introduced by [4], where the authors have combined Bi-Spectrum (higher-order spectrum) and a deep convolutional neural network to detect weak pulsar signals. Their method involved a non-uniform sampling strategy and an extra de-noising process giving their model an accuracy greater than 90%.

The work done by [5] involves a machine learning pipeline to detect pulsars on the HTRU-2 dataset. They proposed two different strategies to overcome unbalanced data. By implementing random resampling methods with K-Fold Cross Validation (as the sampling method) and Correlation (as the feature selection method), their model gave a good insight into how resampling techniques can help overcome the unbalanced data. Their findings have inspired our work. Thus, our work can be considered an extension or a branching to their work. In our proposed work, we've implemented various resampling techniques and have introduced Stratified K-Fold Cross-Validation along with the ordinary K-Fold Cross-Validation as a sampling method. We've introduced ANOVA F-Test (as the feature selection method) and have used Accuracy and F1-Score metrics as a measure to find results through Experiments.

III. Methodology

Figure-1 represents the pipeline of the proposed work. Each component of the methodology is discussed below.

A. Data Analysis

Data Analysis is the first step in the proposed pipeline. Under this step, the attributes and properties such as the number of rows & columns, different data types (like int and float), and null values are analyzed in the dataset. Statistical measures such as mean and standard deviation of the dataset are also analyzed.

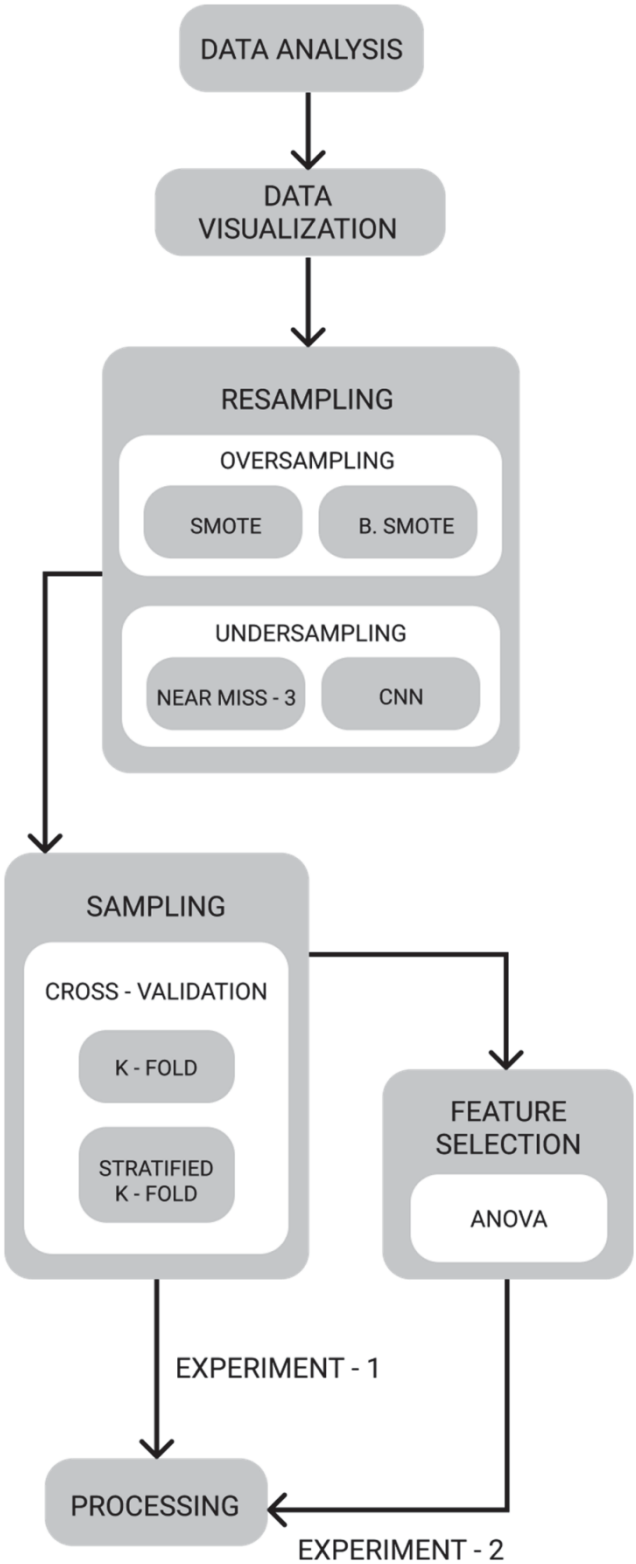


Fig. 1. Pipeline of the proposed work

B. Data Visualization

In machine learning, data visualization is used to understand data or information in a visual format such as a bar chart or a graph. It is an essential step in the exploration of a dataset in terms of class distribution. Data Visualization makes it easier to identify any underlining patterns, outliers, and class imbalances within a dataset.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:25:49 UTC from IEEE Xplore. Restrictions apply.

C. Resampling

Resampling (also known as Monte Carlo Estimation) refers to the techniques, which efficiently use data samples in order to improve the accuracy metric and assists in quantifying the uncertainty of the estimate. The two most common types of resampling techniques, oversampling and under-sampling, are discussed below.

1) Over-sampling: Oversampling is a type of resampling technique that deals with datasets having an imbalance between the target classes by generating synthetic instances for minority classes until a more balanced distribution is attained. The proposed work involves two unique oversampling methods, Smote [8] and Borderline Smote [9].

2) Under-sampling: Under-sampling is another type of resampling technique (opposite of oversampling) that deals with datasets where there is an imbalance between the target classes by discarding pre-existing instances of the majority class until a more balanced distribution is attained. This work involves two under-sampling methods, namely, Near Miss-3 and Condensed Nearest Neighbors [11].

D. Sampling

Sampling is a statistical approach concerned with choosing the individual observation, making it easier to draw statistical inferences about the data. A type of sampling method, called Cross-Validation, used for the validation of models, is described below:

1) Cross-Validation: It is a sampling technique used to train models by the subset of the dataset, which in turn gets evaluated by the complementary subset of the same dataset. In our work, we've implemented two Cross-Validation techniques, namely K-Fold Cross-Validation and Stratified K-Fold Cross-Validation.

E. Feature Selection

It is a means of selecting the most important features of a dataset based on different methods such as Pearson correlation, recursive feature elimination, and more. Restricting the number of features helps in reducing irrelevant or redundant features to be used in the training of the model, which in turn often improves the model's accuracy.

The proposed work uses the ANOVA F-test (analysis of variance f-test) [10] as the basic criteria for feature selection, which is a parametric statistical hypothesis test for determining whether the means from two or more data samples come from the same distribution or not.

F. Processing

Processing is the last step in the pipeline yet the most important. It forms the basis for comparison between results obtained by different classical ML algorithms [12] [13]. The classical algorithms used in the proposed work are listed as follows:

- Logistic Regression
- K – Nearest Neighbors (KNN)
- Decision Trees
- Random Forest
- Naive Bayes

IV. Experiments

A. Dataset

This work uses the HTRU-2 data-set consisting of samples of pulsar candidates gathered through the High Time Resolution Universe Survey and are all verified by human analysts.

This dataset consists of 17,898 instances, of which 16,259 are false instances originated by RFI/noise belonging to the majority negative class with a class label of 0 (negative), and 1,639 are real pulsar instances belonging to the minority positive class with a class label of 1 (positive).

For every pulsar candidate in the dataset, there are a total of eight statistical variables, half of which belong to the DM-SNR curve profile, and the other half belong to the integrated pulse profile. The statistical variables used are mean, excess kurtosis, standard deviation, and skewness for both the profiles.

This section explains the followings terms:

- Integrated pulse profile
- DM-SNR curve
- Skewness
- Kurtosis

1) Integrated Pulse Profile: The individual pulse or blip of increased brightness from a pulsar lacks a constant shape or amplitude in nature, making them not very useful for identifying pulsars uniquely. Since the independent emissions from a pulsar are different in nature, astronomers stack up multiple recorded pulses thus, creating an average integrated pulse profile to characterize a particular pulsar, as shown in figure 2 [7].

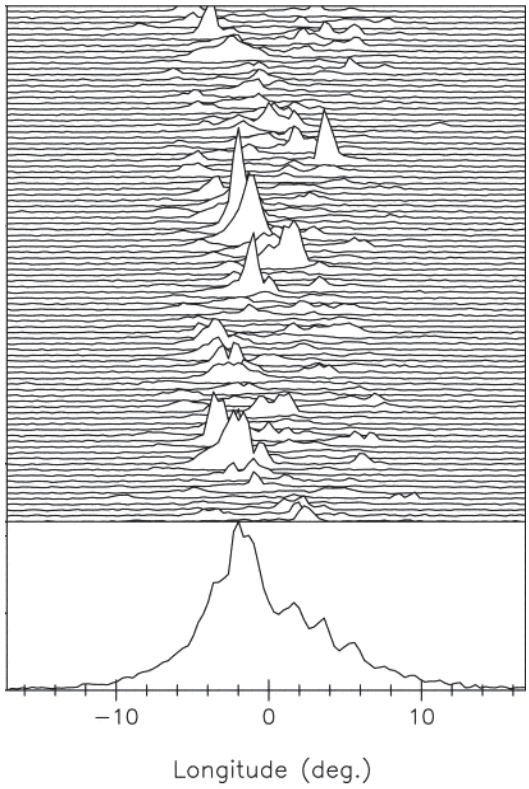


Fig. 2. Average integrated pulse profile

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:25:49 UTC from IEEE Xplore. Restrictions apply.

2) DM-SNR Curve: Another uncertainty that gives rise to a new profile is that the pulses arrive at different times across different radio frequencies. Ionized interstellar medium gives rise to this delay from frequency to frequency (figure 3 [7]) and, this phenomenon is called dispersion. To overcome this effect, Astronomers use several approaches to compensate for the shape of the delay. Even after that, there is always some uncertainty associated with the fit, expressed by the DM-SNR curve, as depicted in figure 4 [7].

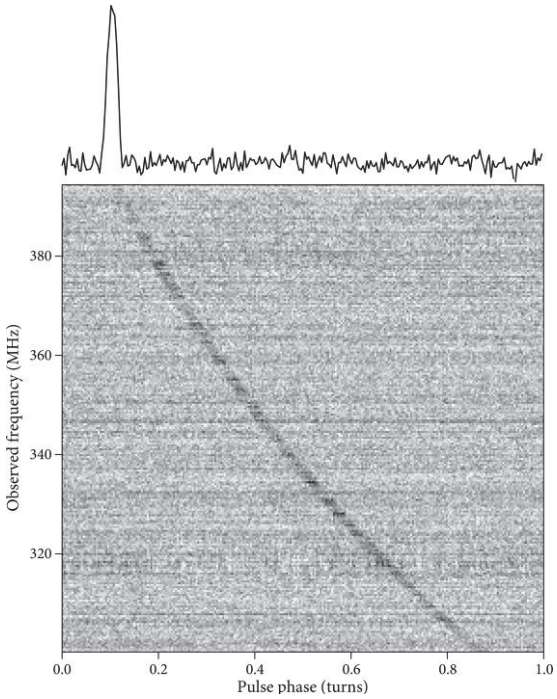


Fig. 3. Delay from frequency to frequency

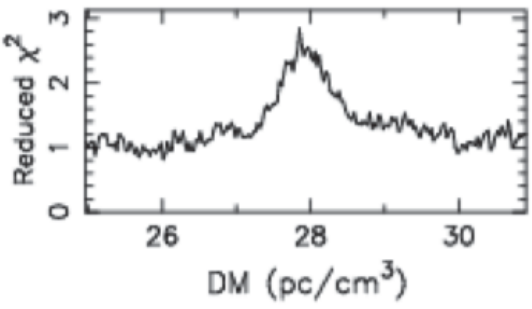


Fig. 4. The DM-SNR curve

3) Skewness: The Skewness of a probability distribution measures the asymmetry of the probability distribution about its mean. Skewness can have values zero, positive, negative, or undefined. The below figure shows the general graphical representation of Skewness:

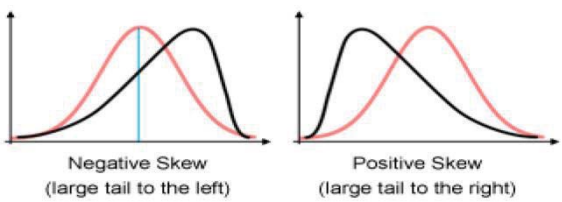


Fig. 5. General graphical representation of Skewness

4) Kurtosis: Kurtosis measures the "peakedness" of the distribution and outlines the shape of the same. The below figure shows the general graphical representation of Kurtosis:

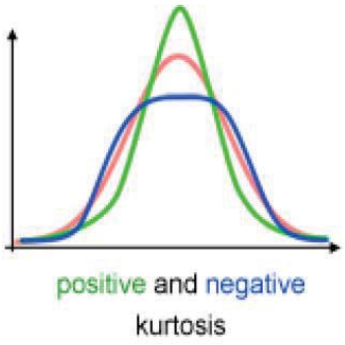


Fig. 6. General graphical representation of Kurtosis

B. Experimental Setup

Both experiments were carried out using five variants of the HTRU-2 dataset in the view of two under-sampling and two over-sampling algorithms in order to repress the imbalance of target classes in the dataset.

1) Dataset naming Convention:

a) ORIGINAL: The original or genuine HTRU-2 dataset without any modification.

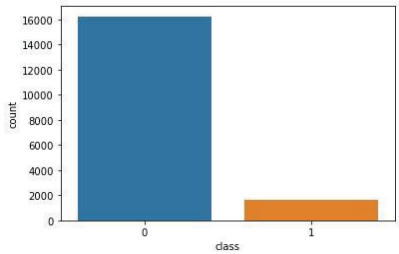
b) SMOTE: Dataset generated using the Smote oversampling algorithm over the genuine dataset.

c) B. SMOTE: Dataset generated using the Borderline Smote oversampling algorithm over the genuine dataset.

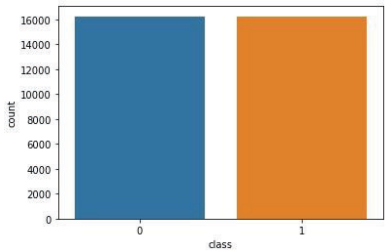
d) NM-3: Dataset generated using the Near Miss-3 undersampling algorithm over the genuine dataset.

e) CNN: Dataset generated using the Condensed Nearest Neighbors undersampling algorithm over the genuine dataset.

2) Visual Class Distribution of all Datasets

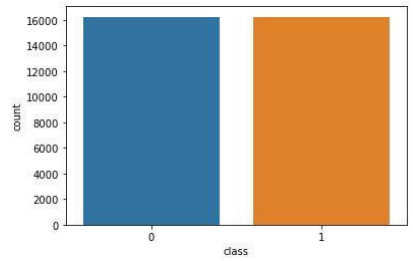


(a)

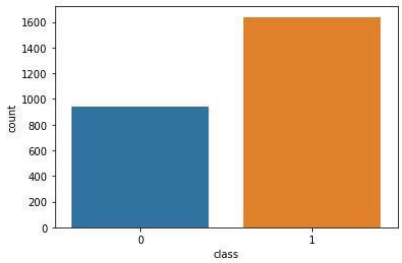


(b)

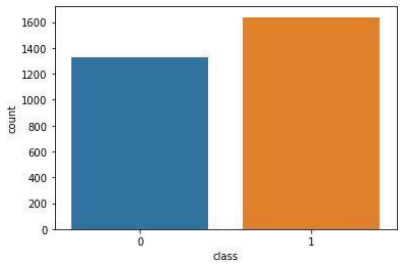
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:25:49 UTC from IEEE Xplore. Restrictions apply.



(c)



(d)



(e)

Fig. 7. Visual Class Distribution of all five datasets: (a) ORIGINAL Dataset, (b) SMOTE Dataset, (c) B. SMOTE Dataset, (d) NM-3 Dataset, and (e) CNN Dataset.

C. Experiment-1

This experiment is performed as per the five datasets mentioned above without feature selection with K-Fold Cross Validation for sampling. The dimensions of datasets used in this experiment are shown in Table-I below:

TABLE I. PROPERTIES OF DATASET USED IN EXPERIMENT-1

Dataset Name	Class Distribution	Dimensions
ORIGINAL	[16259, 1639]	[17898, 9]
SMOTE	[16259, 16259]	[32518, 9]
B. SMOTE	[16259, 16259]	[32518, 9]
NM-3	[1639, 941]	[2580, 9]
CNN	[1639, 1330]	[2969, 9]

D. Experiment-2

This experiment is performed as per the five datasets mentioned above with only six important features, obtained using the ANOVA F-Test Algorithm with Stratified K-Fold Cross Validation for sampling.

The six important features obtained by ANOVA F-Test are:

1) *Mean of the IPP*
2) *Excess kurtosis of the IPP*
3) *Excess kurtosis of the DCP*
4) *The skewness of the IPP*
5) *The standard deviation of the DCP*
6) *Mean of the DCP*

Where,

- IPP stands for *Integrated Pulse Profile*
- DCP stands for *DM-SNR Curve Profile*

The dimensions of datasets used in this experiment are shown in Table II below:

TABLE II. PROPERTIES OF DATASET USED IN EXPERIMENT-2

Dataset Name	Class Distribution	Dimensions
ORIGINAL	[16259, 1639]	[17898, 7]
SMOTE	[16259, 16259]	[32518, 7]
B. SMOTE	[16259, 16259]	[32518, 7]
NM-3	[1639, 941]	[2580, 7]
CNN	[1639, 1330]	[2969, 7]

V. RESULTS

A. Experiment-1

The experiment was performed using the K-Fold Cross Validation sampling method with k (number of folds) = 10. Table-3 below shows the results when accuracy is taken into account as a measure.

TABLE III. EXPERIMENT-1 RESULTS USING ACCURACY AS A MEASURE

Classifiers	Datasets				
	ORIGINAL	*SMOTE*	*B. SMOTE*	*NM-3*	*CNN*
Naive Bayes	0.94	0.90	0.86	0.78	0.84
Logistic Regression	0.97	0.93	0.93	0.82	0.88
KNN	0.97	0.95	0.96	0.61	0.80
Random Forest	0.97	0.97	0.98	0.78	0.87
Decision Trees	0.96	0.95	0.97	0.75	0.81

The above results clearly show that neither oversampling nor under-sampling performed as expected. The original dataset with a high imbalance between target classes performs better than any other dataset. Apart from this, under-sampling has performed poorly in comparison to oversampling. Below is a trend of datasets from best to worst (based on Accuracy):

ORIGINAL > B. SMOTE > SMOTE > CNN > NM-3

TABLE IV. EXPERIMENT-1 RESULTS USING F1-SCORE AS A MEASURE

Classifiers	Datasets				
	ORIGINAL	*SMOTE*	*B. SMOTE*	*NM-3*	*CNN*
Naive Bayes	0.67	0.80	0.75	0.59	0.52
Logistic Regression	0.84	0.87	0.82	0.63	0.54
KNN	0.79	0.84	0.85	0.57	0.52
Random Forest	0.83	0.93	0.93	0.62	0.54
Decision Trees	0.75	0.86	0.89	0.62	0.54

Table-IV shows the results when f1-score acts as a measure.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:25:49 UTC from IEEE Xplore. Restrictions apply.

The results in Table-IV show that oversampling is a better approach for lesser misclassification (i.e., lesser false positives and false negatives) than others as oversampling results in the highest F1-Scores (since it is the weighted harmonic mean of precision and recall). Under-sampling, on the other hand, has a poor F-1 score which makes it clear why even in the results of Table-III, under-sampling had poor accuracies. The original dataset stands between over-sampling and under-sampling. Below is a trend of datasets from best to worst (based on F1-Score):

SMOTE > B. SMOTE > ORIGINAL > CNN > NM-3

Common findings for Experiment-1:

- KNN performs worst, whereas Logistic Regression performs best in under-sampling.
- Naive Bayes performs worst, whereas Random Forest performs best in oversampling.

B. Experiment-2

The experiment was performed using the Stratified K-Fold Cross Validation sampling method with k (number of folds) = 10. Table-5 below shows the results when accuracy is taken into account as a measure.

TABLE V. EXPERIMENT-2 RESULTS USING ACCURACY AS A MEASURE

Classifiers	Datasets				
	ORIGINAL	*SMOTE*	*B. SMOTE*	*NM-3*	*CNN*
Naive Bayes	0.94	0.90	0.84	**0.80**	0.84
Logistic Regression	0.97	0.94	0.93	**0.86**	**0.89**
KNN	0.97	0.95	0.93	**0.76**	**0.83**
Random Forest	0.97	0.97	0.96	**0.84**	**0.88**
Decision Trees	0.96	0.95	0.95	**0.79**	**0.83**

From the above results, one can see that using feature selection didn't improve the classification by any means. The reason why under-sampling performed better (marked in bold) in this Experiment, compared to that in Experiment-1 is, because of the use of Stratified K-Fold Cross Validation, instead of normal K-Fold Cross Validation, which found use in Experiment-1.

Comparing to the results of Experiment-1, the accuracies of only under-sampling are improved. The rest is almost the same as Experiment-1.

Below is a trend of datasets from best to worst (on the basis of Accuracy):

ORIGINAL > SMOTE > B. SMOTE > CNN > NM-3

Table-VI shows the results when f1-score acts as a measure.

TABLE VI. EXPERIMENT-2 RESULTS USING F1-SCORE AS A MEASURE

Classifiers	Datasets				
	ORIGINAL	*SMOTE*	*B. SMOTE*	*NM-3*	*CNN*
Naive Bayes	**0.74**	**0.89**	**0.83**	**0.82**	**0.84**
Logistic Regression	**0.87**	**0.94**	**0.93**	**0.89**	**0.90**
KNN	**0.84**	**0.95**	**0.92**	**0.80**	**0.84**
Random Forest	**0.88**	**0.97**	**0.95**	**0.87**	**0.89**
Decision Trees	**0.81**	**0.95**	**0.94**	**0.84**	**0.85**

We already know that feature selection didn't improve our model in any way. The F1-Scores of Table-6 above implies that Stratified K-Fold Cross Validation alone has reduced the misclassifications previously made in Experiment-1 to a significant amount as the F1-Scores in Table-VI are way better (marked in bold) than that they were in Experiment-1.

Oversampling has performed the best here too, but the interesting part is that under-sampling has performed better than the original dataset, which was opposite in the case of F1-Scores of Experiment-1, where under-sampling performed poorly compared to the original dataset. Below is a trend of datasets from best to worst (on the basis of the F-1 Score):

SMOTE > B. SMOTE > CNN > NM-3 > ORIGINAL

Common findings for Experiment-2:

- KNN performs worst, whereas Logistic Regression performs best in under-sampling.
- Naive Bayes performs worst, whereas Random Forest performs best in oversampling.

VI. CONCLUSION

- Sampling methods such as Stratified K-Fold Cross-Validation can drastically improve the misclassifications made by the model resulting in higher F1- Scores.
- Random Forest & Logistic Regression classifiers can always be used in pulsar classification as they perform the best in over-sampling and under-sampling, respectively.
- Feature Selection doesn't seem to improve the model accuracy as all the features hold some information crucial to classify candidates as pulsars.
- As being highly imbalanced, there is always a need for sampling and resampling methods in order to repress the imbalance of target classes in the HTRU-2 dataset.

REFERENCES

[1] W.W. Zhu, A. Berndsen, E.C. Madsen, M. Tan, I.H. Stairs, A. Brazier, P. Lazarus, R. Lynch, P. Scholz, K. Stovall, S.M. Ransom, S. Banaszak, C.M. Biwer, S. Cohen, L.P. Dartez, J. Flanigan, G. Lunsford, J.G. Martinez, A. Mata, M. Rohr, A. Walker, B. Allen, N.D.R. Bhat, S. Bogdanov, F. Camilo, S. Chatterjee, J.M. Cordes, F. Crawford, J.S. Deneva, G. Desvignes, R.D. Ferdman, P.C.C. Freire, J.W.T. Hessels, F.A. Jenet, D.L. Kaplan, V.M. Kaspi, B. Knispel, K.J. Lee, J. van Leeuwen, A.G. Lyne, M.A. McLaughlin, X.Siemens, L.G. Spitler, and A. Venkataraman: SEARCHING FOR PULSARS USING IMAGE PATTERN RECOGNITION. The Astrophysical Journal, Volume 781, Number 2, https://iopscience.iop.org/article/10.1088/0004-637X/781/2/117/pdf

[2] Haitao Lin, Xiangru Li, Ziying Luo: Pulsars Detection by Machine Learning with Very Few Features. https://arxiv.org/abs/2002.08519v1

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:25:49 UTC from IEEE Xplore. Restrictions apply.

[3] Mohammed A.B. Mahmoud, Ping Guo: Learning from imbalanced pulsar data by combine DCGAN and PILAE algorithm, https://doi.org/10.1016/j.newast.2020.101561

[4] Wang, Longqi & Jin, Jing & Jiang, Yu & Shen, Yi. (2019). A Method for Weak Pulsar Signal Detection Combining the Bispectrum and a Deep Convolutional Neural Network. The Astrophysical Journal. 873. 17. 10.3847/1538-4357/ab0308.

[5] Alexander Ylnner Choquenaira Florez, Braulio Valentin Sanchez Vinces, Diana Carolina Roca Arroyo, and Josimar Edinson Chire Saire Patrıcia Batista Franco {alexanderchf, braulio.sanchez, dianaroca, jecs89, patricia.franco}@usp.br: Machine Learning Pipeline for Pulsar Star Dataset, https://arxiv.org/pdf/2005.01208.pdf

[6] R. J. Lyon, B. W. Stappers, S. Cooper, J. M. Brooke, J. D. Knowles, Fifty Years of Pulsar Candidate Selection: From simple filters to a new principled real-time classification approach, Monthly Notices of the Royal Astronomical Society 459 (1), 1104-1123, DOI: 10.1093/mnras/stw656

[7] James J. Condon and Scott M. Ransom: Essential Radio Astronomy, https://www.cv.nrao.edu/~sransom/web/Ch6.html#S1

[8] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, SMOTE: Synthetic Minority Over-sampling Technique, Journal Of Artificial Intelligence Research, Volume 16, pages 321-357, 2002, DOI: 10.1613/jair.953, https://arxiv.org/abs/1106.1813

[9] Han H., Wang WY., Mao BH. (2005) Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. In: Huang DS., Zhang XP., Huang GB. (eds) Advances in Intelligent Computing. ICIC 2005. Lecture Notes in Computer Science, vol 3644. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11538059_91

[10] Max Kuhn, Kjell Johnson: FEATURE ENGINEERING AND SELECTION: A Practical Approach for Predictive Models, 2019, Page 242

[11] P. Hart, "The condensed nearest neighbor rule (Corresp.)," in IEEE Transactions on Information Theory, vol. 14, no. 3, pp. 515-516, May 1968, doi: 10.1109/TIT.1968.1054155.

[12] M. Wadera, M. Mathur and D. K. Vishwakarma, "Sentiment Analysis of Tweets- A Comparison of Classifiers on Live Stream of Twitter," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 968-972, doi: 10.1109/ICICCS48265.2020.9121166.

[13] P. Meel, M. Singhal, M. Tanwar and N. Saini, "Predicting Flight Delays with Error Calculation using Machine Learned Classifiers," 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 2020, pp. 71-76, doi: 10.1109/SPIN48934.2020.9071159.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:25:49 UTC from IEEE Xplore. Restrictions apply.

Dynamic Analysis of Steel Silo

Uzma Bashir Wani, Nirendra Dev

Abstract: ***The structural response of any structure is the result of various dynamic phenomenon which lead to vibrations or shaking of the structure , depending on the duration of the ground motion, its frequency and time period. In the present work, dynamic analysis of a typical steel silo is done by using linear Time History Analysis and Response Spectrum method for earthquake Zone V as per Indian code. Two analyses are carried out namely, Time History Analysis (THA) and Response Spectrum Analysis (RSA) using STAAD.ProV8i software. The Load combinations are worked out as per IS-1893-2002. The results in terms of Fundamental natural period, Design Base shear, Lateral Displacements, are compared for the two different silo models considered in the present study.***

Keywords: ***Plate elements, STAAD. ProV8i, Response spectrum method, Time history analysis.***

I. INTRODUCTION

The objective of this study is to learn about the different types of forces acting on the structure and its consequences. The typical Silo is modelled and analyzed for the different combinations of loading. Dynamic analysis is performed to obtain the design seismic force and its distribution to different levels of the structure. The dynamic analysis is carried out using Time-History analysis method and Response-Spectrum method. The results of plate elements are discussed in preceding chapters.

II. METHODOLOGY

- Steel silos are considered fixed based.
- Behaviour of the soil is not included in this study.
- Two different models of steel silos are modelled using Finite Element Method of different dimensions.
- Dimensions are as follows;one for 15m diameter and 31.5 m height and other for 18m diameter and 37.65m height.
- For understanding the of steel silo IS-code 9178-Part-1 and Part-2 should be considered.
- Modelling of two different steel silos is done.
- Perform the Dynamic analysis of the silo using Response Spectrum analysis and Time-History analysis for seismic effects.
- Evaluate the analysis results and compare the results.

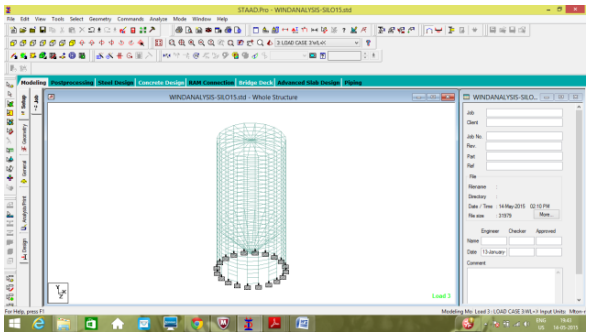


Fig.1.FEM os steel silo (15m dia)

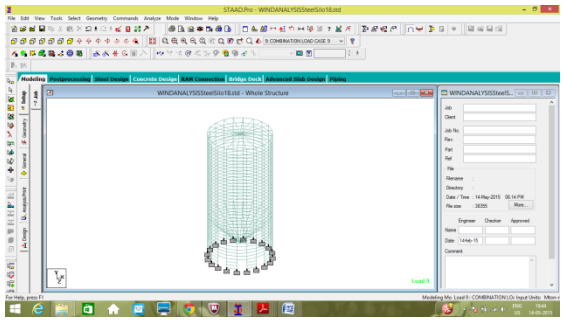


Fig.2.FEM of steel silo (18m dia)

III. RESULTS AND DISCUSSIONS

TIME HISTORY ANALYSIS

In the finite element modelling top roof nodes of the silo are selected to get the maximum deflection at in the structure. The deflection of the two silos are compared, it can be analysed the silo with larger diameter and height will show increased deflection as compared to one with smaller diameter and height. The time-displacement graphs obtained from the STAAD are given below:

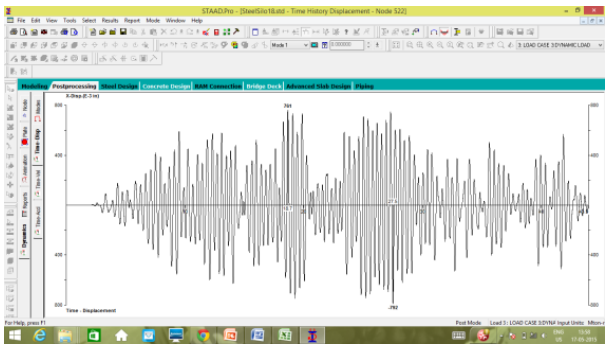


Fig. 3. Time history analysis in X-direction (18 m dia.)

Revised Manuscript Received on April 15, 2020.

Uzma Bashir Wani, Department of Structural Engineering, Delhi Technological University, Delhi, India. Email: uzmawani@gmail.com

Dr. Nirendra Dev, Head, Department of Civil Engineering, Delhi Technological University, Delhi, India. E-mail: nirendradev@dce.ac.in

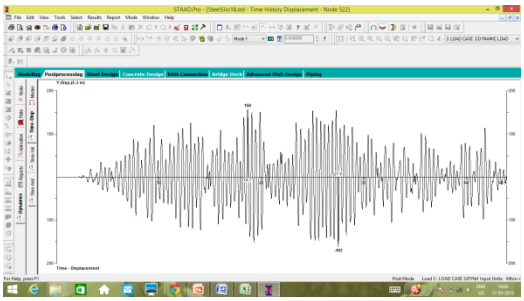


Fig. 4.Time history analysis in Y-direction (18 m dia.)

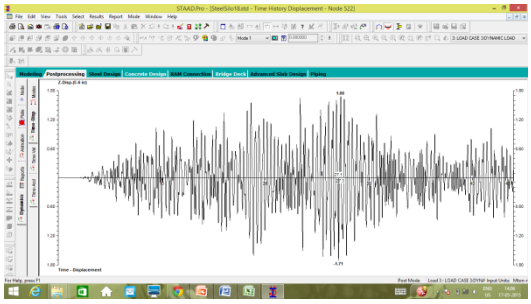


Fig.5.Time history analysis in Z-direction (18 m dia.)

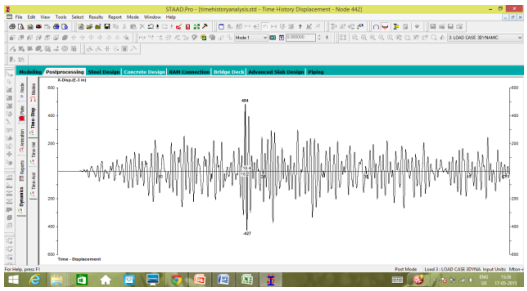


Fig.6.Time history analysis in X-direction (15 m dia.)

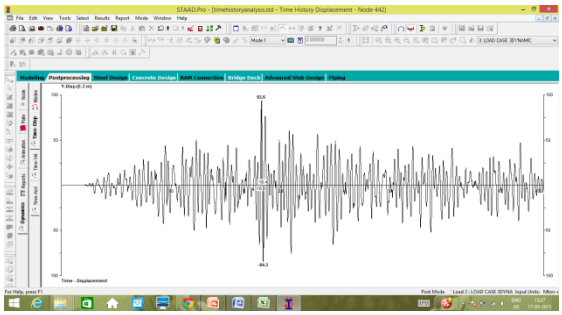


Fig.7. Time history analysis in Y-direction (15 m dia.)

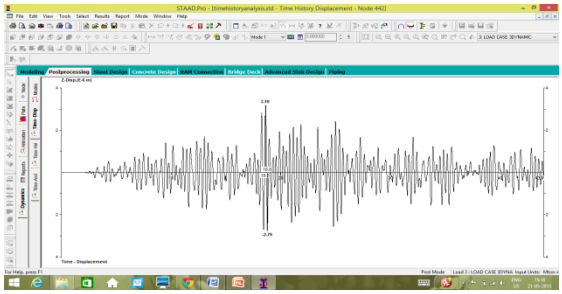


Fig.8. Time history analysis in Z-direction (15 m dia.)

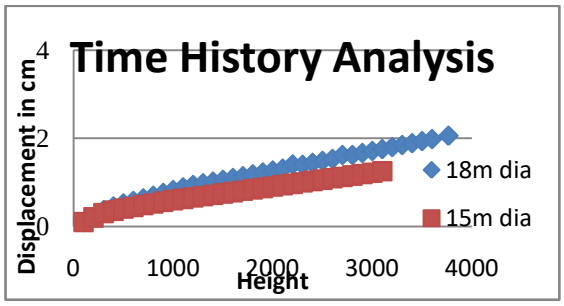


Fig.9.Displacement variation with height.

Table.I Displacement variation along height STAAD output for THA

Height	18m dia	15m dia
100	0.09652	0.1016
200	0.21082	0.20574
300	0.3683	0.30226
400	0.45466	0.35814
500	0.52324	0.40386
600	0.57912	0.44196
700	0.63754	0.4826
800	0.69596	0.52324
900	0.75946	0.56134
1000	0.8255	0.5969
1100	0.88646	0.62738
1200	0.9398	0.65786
1300	0.9779	0.68834
1400	1.016	0.71882
1500	1.05664	0.7493
1600	1.09728	0.77978
1700	1.13792	0.81026
1800	1.1811	0.84328
1900	1.22428	0.87376
2000	1.26746	0.90424
2100	1.31064	0.93726
2200	1.40462	0.96774
2300	1.397	0.99822
2400	1.44018	1.03124
2500	1.4859	1.06172
2600	1.52908	1.09474
2700	1.61798	1.12522
2800	1.61798	1.1557
2900	1.6637	1.18872
3000	1.70942	1.2192
3100	1.7526	1.25222
3200	1.79832	

3300	1.84404	
3400	1.88722	
3500	1.93294	
3600	1.97866	
3765	2.05486	

RESPONSE SPECTRUM ANALYSIS

The comparison of the deflection with height is shown in the Fig. 10. The behaviour of the two silos are analysed; by plotting the graph of displacements along the height of the structure.

Table.II Displacement variation along height STAAD output for RSA

Height in cm	Deflection-18m	Defelection-15m
100	0.03048	0.04064
200	0.06604	0.07366
300	0.13208	0.10414
400	0.1524	0.12446
500	0.17018	0.1397
600	0.18796	0.1524
700	0.20828	0.16764
800	0.2286	0.18034
900	0.24892	0.19558
1000	0.26924	0.20828
1100	0.28702	0.2159
1200	0.29972	0.22606
1300	0.31496	0.23622
1400	0.32766	0.24638
1500	0.3429	0.25654
1600	0.3556	0.2667
1700	0.37084	0.27686
1800	0.38354	0.28956
1900	0.39878	0.29972
2000	0.41402	0.30988
2100	0.42672	0.32004
2200	0.45466	0.3302
2300	0.4572	0.34036
2400	0.47244	0.35052
2500	0.48514	0.36322
2600	0.50038	0.37338
2700	0.51562	0.38354
2800	0.53086	0.3937
2900	0.54356	0.40386
3000	0.5588	0.41402
3100	0.57404	0.42672
3200	0.58928	
3300	0.60452	
3400	0.61722	
3500	0.63246	
3600	0.65786	
3765	0.31496	

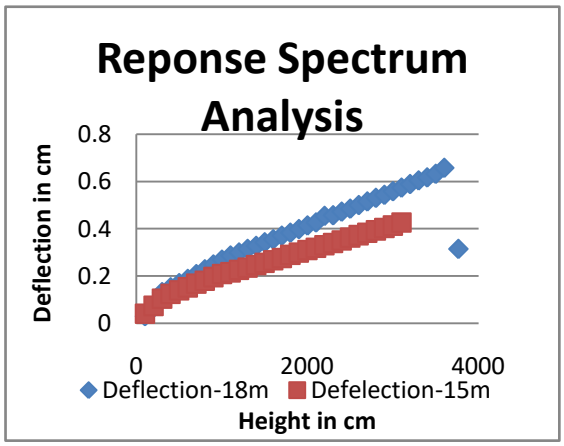


Fig.10. Displacement variation with height.

STAAD OUT PUT

The STAAD output for the base shear calculated for the two models is given below:

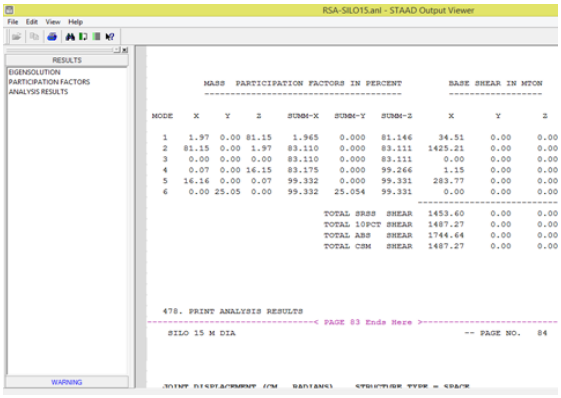


Fig.11. Base share for 15m dia.

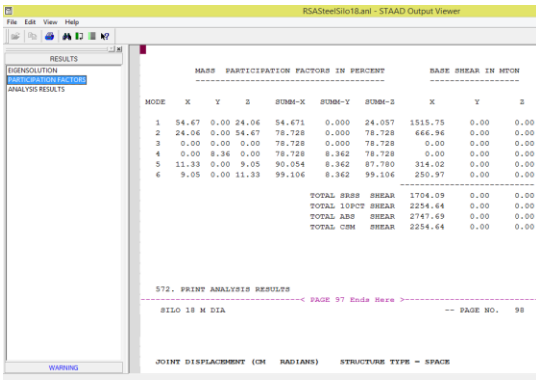


Fig.12: Base share for 18 m dia

IV. CONCLUSION

Dynamic analysis for two types of steel silo is done by using Seismic methods; Time History Method and Response Spectrum method for earthquake zone V as per Indian Standard code.

During the dynamic analysis, behaviour of the structure is evaluated. Permissible displacement of structure is 200mm.
It can be concluded by saying that along the height of the silos the percentage of distortion increases although the thickness of the plate is considered uniform through-out the length of the silos.

REFERENCES

1. Indian Code of practice for design of earthquake loads for buildings and structures IS:1893-2002.
2. Indian Code of practice for design loads (other than earthquake loads) IS: 875-PART-I & PART-II-1987.
3. Indian code of practice criteria for design of steel bins for storage of bulk materials IS:9178-PART-I; General requirements and assessment of loads, IS:9178-PART-II; Design criteria and IS:9178-PART-III; Bins designed for mass flow and funnel flow.
4. Indian Code of practice for silos for grain storage IS:5503(PART-I)-1969.
5. Indian Code of practice for general construction in steel IS : 800-2007
6. Analysis of steel silo structures on discrete supports by Dr.Hongyu Li.
7. Load development and structural consideration in silo design by J.W Carson, Phd and R.T Jenkyn, P.Eng.
8. J. Penzien and M. Watabe, "Characteristics of 3-D Earthquake Ground Motions,"*EarthquakeEngineering and Structural Dynamics*, Vol. 3, pp. 365-373, 1975.
9. Paulo B. Lourenco."Types of analysis :linear static ,linear dynamic and non linear static ". www.civil.uminho.pt/masonary.
10. A.W Roberts "Developments in silo design for the safe and efficient storage and handling of grains".
11. Adem Dogangun1; Zeki Karaca2; Ahmet Durmus3; and HalilSezen, M.ASCE4."Cause of Damage and Failures in Silo Structures".
12. Silos theory and practice by Marcel L. Reimbert and Andre M. Reimbert.
13. SEISMIC BEHAVIOR OF SILOS WITH DIFFERENT HEIGHT TO DIAMETER RATIOS CONSIDERING GRANULAR MATERIALSTRUCTURE INTERACTION,F. Nateghi and M. Yakhchalian; Structural Engineering Research Center, International Institute of Earthquake Engineering and Seismology, Tehran, Iran nateghi@iiees.ac.ir, *m.yakhchalian@iiees.ac.ir*
14. Parametric Study On Dynamic Response Of Silo Anand Adi [1] , Hemant L. Sonawadekar [2] 1-Post Graduate Student, Department of Civil Engineering, KLEMSSCET, Belgaum, Karnataka, India, 590008 ,2-Assistant Professor, Department of Civil Engineering, KLEMSSCET, Belgaum,Karnataka, India, 590008.
15. S.K Duggal; Earthquake Resistant design of structures. Oxford university press 2012.
16. A. K Chopra; Dynamics of structures. Theory and applications to earthquake engineering; second edition.

AUTHOR'S PROFILE

Name: **UZMA BASHIR WANI**

Mobile No: +91 9811653361

Email: uzmawani@gmail.com

Educational Qualifications:

Year	Examination	University
Pursuing	* Ph. D. (Structural Engineering)	Delhi Technological University, Delhi. (formerly Delhi College of Engineering)
2013-15	M. Tech. (Earthquake Engineering)	Jamia Millia Islamia, New Delhi
2005-09	B. Tech. (Civil Engineering)	Jamia Millia Islamia, New Delhi

Ph. D. Thesis:

Topic: Dynamic Analysis of Base Isolated Reinforced concrete structures.

Publications:

"Dynamic Analysis of steel silo using wind load as per Indian Standard'' in International Journal of Engineering Research and Technology (IJERT), vol. 8 issue 11,November2019. ISSN:2278-0181

M. Tech. Thesis

Topic: To perform Dynamic Analysis of Steel Silo as per Indian Standard using Finite Element Method (STAAD) applying linear dynamic time history approach and response spectrum approach

Teaching Experience: (03 Years Approx.)

Position	Organization	Subjects Taught	
		Architecture	**Civil Engineering**
Assistant Professor	Manav Rachna International University, Faridabad	Structural Design –IV Structural Design –V Structural Design –VI	Design of Concrete Structures – I Design of Concrete Structures – II Design of Steel Structure – II Earthquake Resistant Design of Structures STAAD Pro Labs Research Mentor

Industrial Experience: **(04 Years Approx.)**
Mott Macdonald Pvt. Ltd. IndiDesign Engineer (Structures)
Worked as a Structural Engineer on various domestic and international assignments. My role primarily involved Analysis and structural design of RCC Structures for VIP complex, Cargo terminal building and Ancillary Buildings which includes design of slab, beams, columns, grade slab, staircase and foundations using BS codes

- Analysis and structural design of Steel structures for Ancillary buildings including trusses, purlins ,staircase , canopy, beams, columns and Skylight frame using BS codes
- Structural analysis and design of free standing Steel cabins in main Passenger Terminal Building including Promat roof design for fire protection
- Structural analysis and Design of in Retail Cabin in main Passenger Terminal Building.

Projects:

- Master Planning and Detailed Engineering Design Services for Expansion and Modernisation of Malé International Airport (**Location : Maldives**)
- Attended one week Training Programme on Road Safety and Road Safety organized by IAHE

Synergics Hydro Power Pvt. Ltd, India **Design Engineer** (July 2009 – May 2011)

- Prepared preliminary design, BOQ and cost estimates required to analyse cost to benefit of capital expenditures of projects accordingly.
- Hydrology - Involved in Power Potential Studies
- Hydraulic designing of structures using IS Codes
- Cost Estimation/BOQ with the CEA/CWC guidelines

Name: **Dr.Nirendra Dev**
Designation: Head of Civil Engineering Deparment of Delhi Technological University
Phone: 9999399375
Email: nirendradev@dce.ac.in
Qualifications: Ph.D, M. Sc.(Engg) in Building Engg, B. Tech.(Civil)

Areas of Interest: Structural Engg., Structural Dynamics, Finite Element Method, Disaster Management, High Strength Nano Composites, Concrete Technology

Post Graduate and Ph. D. works:
M.Sc. Engg. Dissertation topic
"**ANALYSIS OF L-SHAPED PLATES WITH CONTINUOUS SUPPORT**"
Ph.D. topic
"BEHAVIOR OF TRACK STRUCTURE UNDER STAIONARY AND MOVING VERTICAL WHEEL LOADS"
Software Packages Developed:
SATRACKS - Static Analysis of TRACK Structure
DATRACKS - Dynamic Analysis of TRACK Structure
PLTRACKS - Plotting of TRACK Structure (Pre and post processor)
Professional Training:
Programming & Interfacing the 8086/88 from 12.09.1988 to 16.09.1988 organized by Deptt. of E & CE, Univ. of Roorkee.
ii. Finite Elements in Non-linear Mechanics from 25.09.1989 to 06.10.1989 organized by Civil Engg. Deptt. Univ. of Roorkee

Post – Qualification Experience: 32 years
Teaching: 11 years
Research: 6 years
General & Educational Administration: 15 years
M. Tech. Projects Supervised: 18
Ph. D. Thesis Supervised: 04 (under progress)

Courses taught
(a) Under – Graduate Courses taught: 14
(b) Post– Graduate Courses taught: 04
Memberships of Society:
- Life Member of Indian Society for Wind Engg (ISWE)
- Life Member of Indian Society for Construction Materials and Structures (ISCMS)
- Life Member of The Indian Science Congress Association

Consultancy Projects: 04
Research Reports: 03
Papers in Conference and Seminar: 10

Dynamic Power Consumption and Delay Analysis for Ultra-Low Power 2 to 1 Multiplexer Designs

Nishant KUMAR[1], *Poornima MITTAL*[1], *Bhawna RAWAT*[1], *Mudit MITTAL*[2]

[1]Department of Electronics and Communication Engineering, Delhi Technological University, Main Bawana Road, 110042 Delhi, India
[2]Department of Information Technology, Institute of Technology and Management, 60 Chakarata Road, 248001 Dehradun, India

nishant205kumar@gmail.com, poornimamittal@dtu.ac.in, bhawnarawat12@gmail.com, mushi172@gmail.com

DOI: 10.15598/aeee.v19i2.3821

Article history: Received May 23, 2020; Revised Apr 03, 2021; Accepted Apr 04, 2021; Published Jun 30, 2021.
This is an open access article under the BY-CC license.

Abstract. *This paper highlights a comparative analysis of eight diverse techniques for 2 to 1 multiplexer implementation. The functionality is identical but significant differences in dynamic power consumption and propagation delay are observed. This paper aims to enable the designer to pick out the best fit structure for a specific application in keeping with their design requirement. The multiplexers are designed at 90 nm technology node and simulated at a supply voltage of 1 V.*

Keywords

Average power dissipation, CMOS, dynamic power, leakage power, low power application, power delay product, standard cells.

1. Introduction

Technology is pacing at an exponential rate against time. Devices are being remodelled and improved within short spans to outdo themselves. Consequently, each component of a device is being analysed to improve its performance. Designers are ceaselessly trying to improve the performance of the existing devices or trying to devise a new way to design and improve the performance [1]. A multiplexer is one of the basic building blocks of digital systems [2] and [3]. In this paper, different design techniques of 2 to 1 multiplexer are compared.

The advancements in design automation of ASICs have enabled positioning millions of transistors on a single chip for robust circuitry implementation. Consequently, the density of transistors for a given area has increased significantly, thereby leading to an increase in formidable designing issues [4]. Henceforth, the solutions that have been recommended suggest a reduction of the transistor power supply voltage, switching frequency, and capacitance [5], [6] and [7]. Depending on the application, different types of circuits and design methodologies were proposed which disallowed the formulation of uniform rules for optimum logic types.

The smallest multiplexer that can be designed is a 2 to 1 multiplexer. It forms the building block for other larger multiplexer modules [8] and optimising its configuration enhances its stability [9]. Therefore, to optimise the performance of 2 to 1 multiplexer, different configurations including Gate Diffusion Input (GDI), Pass Transistor (PT), Multiplexer Single with Level Restoration (MSL), Transmission Gate (TG), Static CMOS, Complementary Pass Logic (CPL), Cascode Voltage Switch Logic (CVSL) and Multi Threshold CMOS CVSL (MTCMOS CVSL) are analysed in this paper. All the multiplexers are designed using complementary MOSFET transistors at 90 nm technology node. Further, their performance is analysed and compared by means of output response and dynamic power dissipation using Cadence Virtuoso software. Additionally, power and delay are analysed to find the best multiplexer amongst all the configurations.

This paper is arranged into five sections, including this introductory section. Further, Sec. 2. illustrates schematics diversity and working for different multi-

plexer configurations. The output response for different multiplexer designs is presented and discussed in Sec. 3. . Additionally, dynamic power dissipation and delay response are elaborated in Sec. 4. . Finally, evaluated structures' properties are summarized in Sec. 5. .

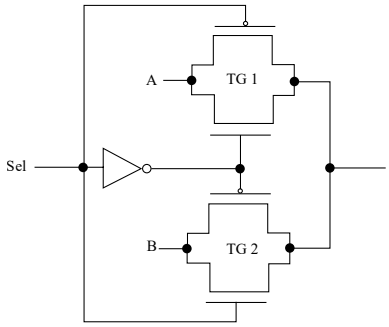
2. CMOS Structures of Multiplexers

Diverse structures of 2 to 1 multiplexer are simulated and analysed. Designs such as GDI and PT multiplexer have simpler schematic and utilize lesser area with low power dissipation but with degraded output response. Other techniques such as MSL, Static CMOS, CPL, CVSL and MTCMOS based multiplexer are liable for producing non-degraded output but consume more power and introduce larger delay, as well.

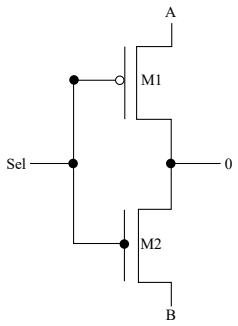
Transmission gate based 2 to 1 multiplexer is designed using a pair of Transmission Gates (TGs). Each TG is a pair of NMOS and PMOS transistors wherein the source and drain terminals of transistors are connected in parallel, as illustrated in Fig. 1(a). Both NMOS and PMOS permit the same input simultaneously. Thus, it is transferred to the output node through this TG without any deterioration [10]. At high input signal, the NMOS gives a weak 1 at the output. However, PMOS provides a strong 1 at the same time, thereby maintaining the output level. Similarly, at low input, the PMOS produces a weak 0 but NMOS supplies a strong 0 at the output [11]. TG configuration is used to isolate the components and signals/data from being transmitted to the other nodes without using any other hardware.

Gate Diffusion Input (GDI) logic allows the user to design complex logic circuits using a smaller number of transistors. The basic structure of GDI resembles a CMOS inverter and inherits characteristics of CMOS and PTL logic [12]. This is an appropriate technique for designing fast and low-power circuits using fewer transistors as compared to CMOS and PTL techniques [13] and [14]. The schematic for the GDI based 2 to 1 multiplexer is given in Fig. 1(b). When Sel = 1, the NMOS transistor operates in ON mode and input signal B will pass to the output; O. On the other side, output O receives the input signal A, when Sel is maintained at 0.

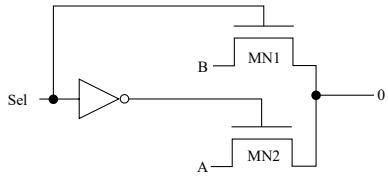
The GDI based multiplexer employs A and B inputs to the multiplexer, while the Sel signal acts as the select line and determines the input that gets transmitted to the output. The logical function implemented by the GDI based SRAM bit cell is nSel.A+ Sel.B. The advantages associated with the GDI technique are minimal transistor requirements, low power dissipation and fast operation. However, the limitations of this structure encounters are - (1) If A = 0, then PMOS being a weak 0 will not pass a perfect 0 at the output, (2) The complementary is applicable for B = 1, as NMOS delivers weak 1 to output.



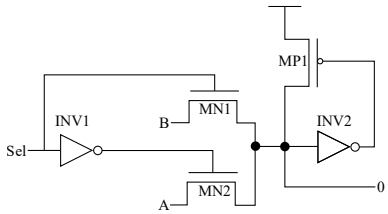
(a) TG based multiplexer.



(b) GDI based multiplexer.



(c) PT based multiplexer.



(d) MSL based multiplexer.

Fig. 1: Schematic of TG, GDI, PT, and MSL based multiplexers.

Pass Transistor Logic (PTL) uses two NMOS transistors in pass transistor configuration. This logic is different from CMOS design as the source side of the logic network is connected to the input signal instead of the power supply [15], [16], [17] and [18], as depicted in Fig. 1(c). When Sel = 0, the MN2 transistor is saturated that leads the value at A to appear at the output terminal. Whereas, for Sel = 1, the value at B gets transferred to the output. The upsides

© 2021 ADVANCES IN ELECTRICAL AND ELECTRONIC ENGINEERING

of PTL are high speed, low power consumption and low interconnect effect. However, the factor that limits the use of PTL technique is reduced signal integrity due to the inability of NMOS to pass a strong 1 [19]. Therefore, at conditions such as Sel = 1, B = 1 or Sel = 0, A = 1, the output driving capabilities are weak.

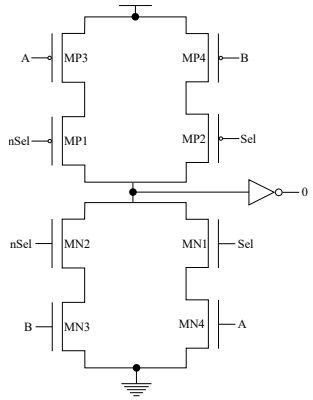
The shortcoming of the PT based multiplexer is modified and MSL based multiplexer is created. In this design, an additional PMOS transistor is used to restore the output level. The gate of the PMOS (MP1) is driven by an inverter (INV1) controlled by the output of the PT based multiplexer [20]. The gate and drain of PMOS (MP1) are connected to the output and input of the second inverter (INV2), respectively, as indicated in Fig. 1(d). Similarly, gate terminals of both NMOS transistors (MN1 and MN2) are connected to the input and the output of the first inverter (INV1). The purpose of designing MSL based multiplexer is to overcome the drawback of output deterioration and complex synthesis methodology for PT multiplexer [21]. When Sel = 1 and B = 1, NMOS produces a weak 1 as explained in the previous subsection. To rectify it, MSL based multiplexer converts this 1 to 0 through the inverter [22]. This inverter then drives the PMOS to ON state, thereby making the output signal a strong 1. The other advantage of MSL based multiplexer is maximum output swing (nearly V_{DD} and 0).

The next multiplexer implementation for analysis is Static CMOS based multiplexer. This multiplexer uses eight transistors, wherein four PMOS transistors form the pull-up network and the other four NMOS transistors form the pull-down network. This multiplexer utilizes inverted inputs and delivers inverted output; OB along with the actual output; O. The schematic for static CMOS based multiplexer is illustrated in Fig. 2(a) [23]. When Sel = 0, B = 0 or nSel = 1, A = 1/0, the MP2, MP4, MN2 are ON and MN1, MN3, MP1 are OFF, thus output becomes 0. There is some leakage due to MN2 and MN4 of the pull-down network. Similar observations are achieved for nSel = B = 1 or Sel = 0, A = 0/1 and 1 is produced at the output O. With similar phenomena, a zero output is produced for nSel = A = 0 or Sel = 1, B = 0/1. In such a condition, the leakage occurs through MN2 and MN4 of the pull-down network. This analysis concludes that the input B/A gets produced at the output for Sel = 0/1.

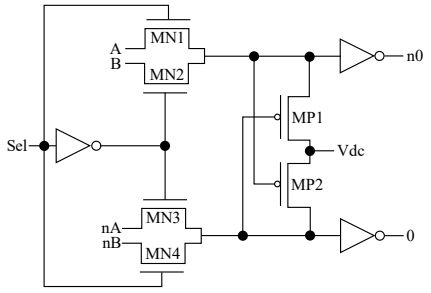
Complimentary Pass Transistor Logic (CPL) comprises three inverters and six transistors of which four are NMOS transistors and two are PMOS transistors. The NMOS transistors (MN1, MN2, MN3, and MN4) are connected in pairs as pass transistors. A pair of NMOS (MN1-MN2) is used to pass the input signal, while the other pair (MN3-MN4) is used to provide inverted input. The PMOS pair (MP1-MP2) is used for level restoration. The inverters are used to invert the select signal and the output. In this circuit, half of the transistors are used to pull up, while the other half is used to pull down the logic, thereby providing both actual output (O) and its complement (nO), as depicted in Fig. 2(b). When Sel = 0, transistors MN2-MN3 are turned ON and pass input signal B and nB (complement of input B) to O and nO, respectively. If input B = 0 (i.e. nB = 1), MP2 is ON. Therefore, V_{DC} gets connected to the output of MN3 and restores the logic level 1. Hence, 0 is produced at the output of the inverter. Now, if B = 1 (nB = 0), transistor MP1 is ON, the V_{DD} gets connected to the output of MN2 and logic level 1 is restored. Accordingly, inverted output (nO) is 0. This technique exhibits advantages like the presence of both output and inverted output, fast operation and restoration of output level. But this structure dissipates significantly more power than other structures, thereby limiting its application.

Cascode Voltage Switch Logic (CVSL) multiplexer is shown in Fig. 2(c). The NMOS logic forms the pull-down network and generates the complementary logic. CVSL is composed of a differential latching circuit and a cascaded complementary logic array [24]. Therefore, this structure is also acknowledged as Differential Cascode Voltage Switch Logic (DCVS or DCVSL) [20] and [23]. When Sel = 0, A = 0, transistors MN3, MN5, and MN6 are ON, henceforth, the output and gate of MP2 get connected to the ground terminal through transistor MN5 and MN6. Thus, the output obtained at O = 0 (nO = 1). For Sel = 0, A = 1, transistors MN3, MN4, and MN5 are ON and the ground gets connected to the output O through MN3 and MN4, thereby making the output nO = 0 and O = 1 as it gets connected to V_{DC} owing to the ON state of MP1. This configuration exhibits an advantage of reduction in number of PMOS transistors from each logic function leading to a significant area reduction.

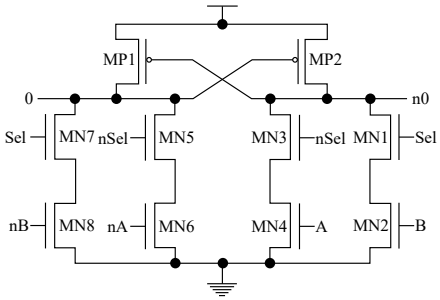
The functionality of Multi-Threshold CMOS Cascode Voltage Switch Logic (MTCMOS CVSL) based multiplexer is similar to the CVSL multiplexer. This technique is used to reduce the circuit leakage in static conditions [24]. In this configuration, one pair of PMOS-NMOS (MP3-MN9) is used, wherein MP3 isolates the logic circuit from V_{DC} and MN9 isolates the ground, as shown in Fig. 2(d). MTCMOS technique separates the circuit from the power supply and ground to prevent power dissipation in static state [25]. Here, two complemented sleep signals; Sleep and nSleep (complement of sleep signal) are used to control the gates of PMOS (MP3) and NMOS (MN9), respectively. When sleep is low, the circuit works as standard CVSL based multiplexer. During sleep state, both outputs are in High Z (high impedance) state and the circuit is non-operational.



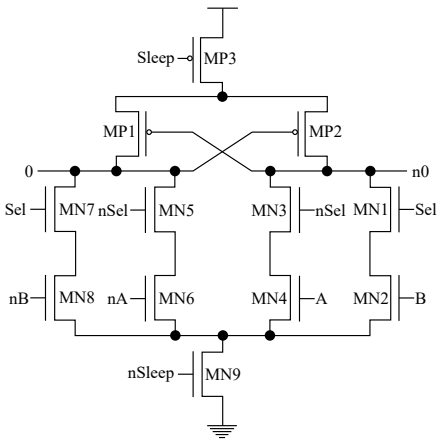
(a) Static CMOS based multiplexer.



(b) CPL based multiplexer.



(c) CVSL based multiplexer.



(d) MTCMOS CVSL based multiplexer.

Fig. 2: Schematic of static CMOS, CPL, CVSL, and MTCMOS CVSL based multiplexers.

3. Output Response and Analysis

In this section, the output responses corresponding to different multiplexer structures discussed in the previous section are analyzed. The rise and fall timing values for A, B and Sel are incorporated as 0.45 ns, 0.35 ns and 1 ns, respectively. To analyze the realistic performance of multiplexers, a load capacitor of 1 fF [26] is used at the output.

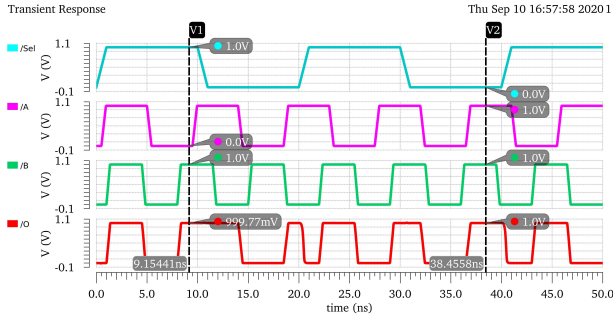
In an ideal scenario, the generated output waveform will attain 0 V and 1 V level as per the input combination and select line configuration. If the output waveform is unable to attain a perfect 0 V or 1 V level, it is a non-ideal waveform. The output waveforms generated for all multiplexer structures discussed are ideal except for GDI, PT and MTCMOS CVSL. The output waveforms for ideal and non-ideal response for different multiplexer techniques are illustrated in Fig. 3.

The structure of TG, MSL, Static CMOS, CPL, and CVSL is such that they can transfer perfect 0 V or 1 V to the output whenever the select lines permit. Therefore, their output waveforms obtained are ideal. The ideal output waveform is given in Fig. 3(a). The ideal output situation is not attained for GDI based multiplexer, as showcased in Fig. 3(b). The GDI based multiplexer is designed using a pair of NMOS and PMOS. When the input at PMOS is 0, the output is not a perfect 0, as PMOS delivers a weak 0. Similarly, when input is 1 at NMOS, the output obtained is not equivalent to 1 as NMOS provides weak 1 at the output. The other multiplexer structure unable to produce ideal output is PT based multiplexer. This is the case as the schematic for PT based multiplexer is purely dependent on NMOS for the generation of output. As a result, the structure can never pass a perfect 1 due to NMOS being a weak 1. This technique produces a perfect 0 V level but the 1 V output level always falls short of its perfect value, as can be observed in Fig. 3(c).

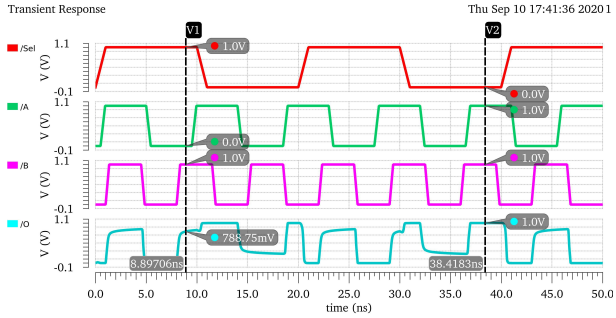
Lastly, the output obtained for MTCMOS CVLS is different in comparison to all other multiplexer structures, as depicted in Fig. 3(d). This technique is dependent on an additional pair of PMOS-NMOS, of which the PMOS transistor is used to isolate the V_{DC} from the circuit and the NMOS transistor is used to isolate the ground. This is done to reduce the power dissipation of the circuit. But this has its implications on the output waveforms. When sleep = 0, the circuit works as expected of a multiplexer circuitry, but when sleep = 1, the multiplexer circuit gets isolated from supply voltage and ground. The output is put in high-impedance state, losing the driving capabilities. This occurs as the entire circuit is disconnected from the supply voltage and ground. On the other side,

© 2021 ADVANCES IN ELECTRICAL AND ELECTRONIC ENGINEERING

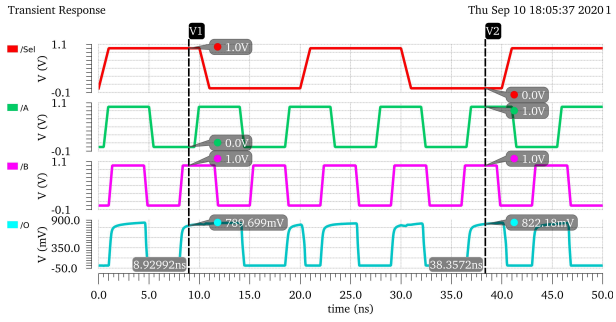
when the sleep signal is low, the connections to the supply voltage and ground are re-established for the multiplexer core. Thus, the circuit produces output in keeping with the functioning of a multiplexer circuit.



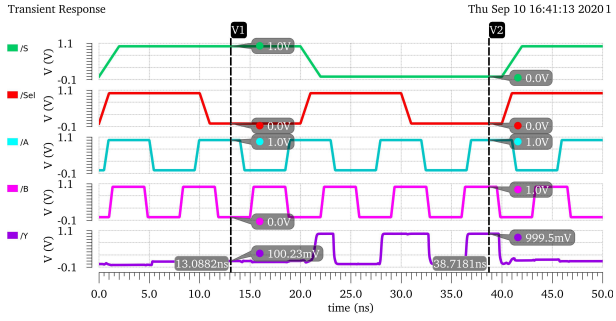
(a) Ideal scenario.



(b) GDI multiplexer.



(c) PT multiplexer.



(d) MTCMOS CVSL multiplexer.

Fig. 3: The output response for ideal scenario, GDI multiplexer, PT multiplexer and MTCMOS CVSL multiplexer.

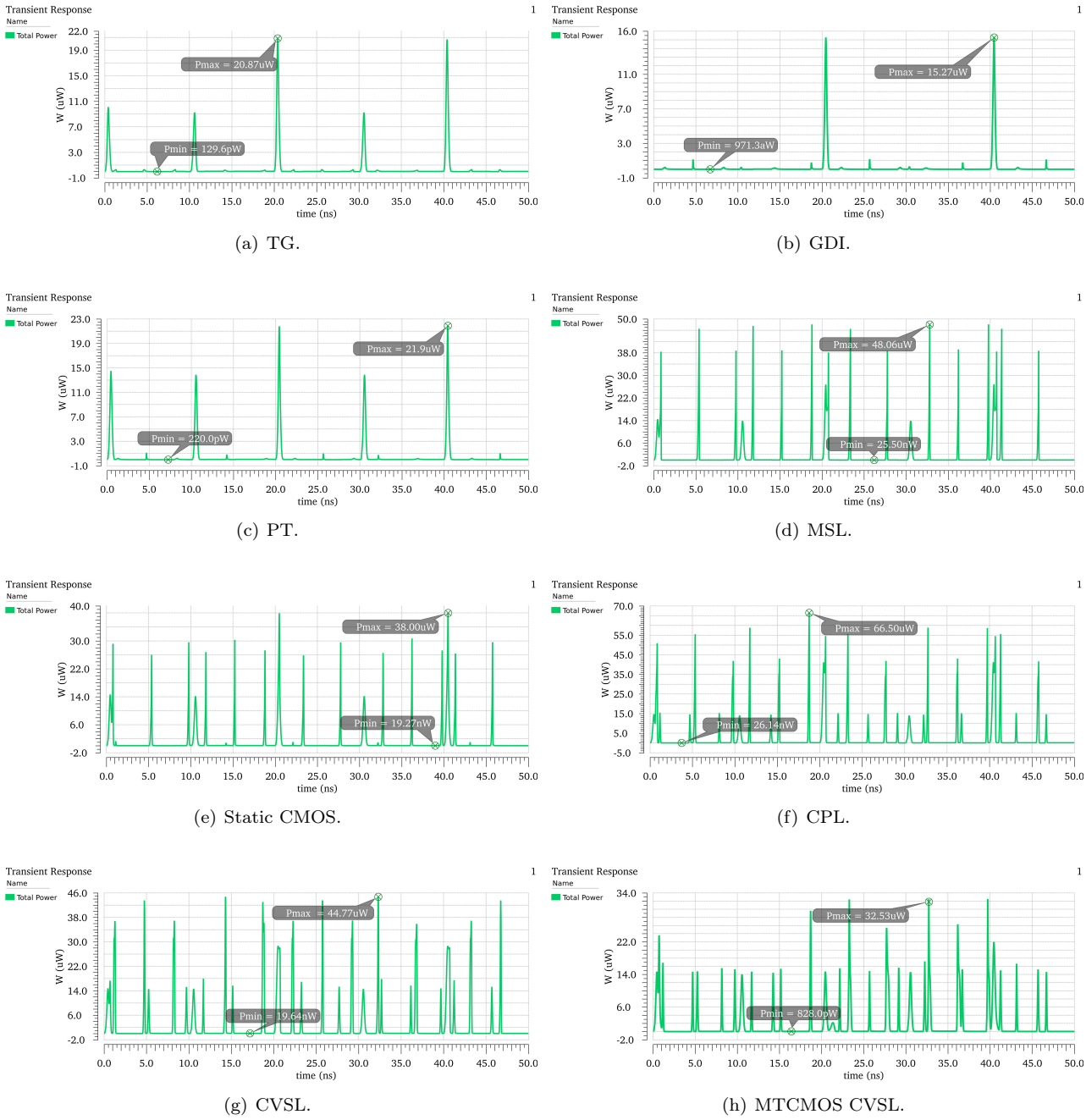
4. Dynamic Power Dissipation Analysis

Power dissipation is a key characteristic of all considered structures. With the increasing popularity of low-power devices, it has become imperative to study the power dissipation of all multiplexer techniques as well. In Fig. 4, the dynamic power dissipation pulses corresponding to the output wave pulse are presented for different multiplexer techniques. Dynamic power consumption is caused by current flow in circuit, when the transistors of the devices are switching from one logic state to another. The frequency at which the device is switching, plus the rise and fall times of the input signal, as well as the number of internal nodes on the critical path for the device, have a direct effect on the duration of the current spike [27].

To measure the power dissipation for the multiplexers, a test wave was applied to all and the power dissipation was measured from the output waveform generated at the output node. The maximum and minimum power dissipation of TG, GDI, PT, MSL, Static CMOS, CPL, CVSL and MTCMOS CVSL based multiplexer are depicted in Tab. 1. The maximum power dissipation of MSL based multiplexer is obtained as 463.59 µW, which is observed least amongst all other multiplexers, as demonstrated in Tab. 1. On the other hand, the obtained power dissipation is minimum for GDI based multiplexer with a magnitude of 9.7 fW, which is reduced by order of 5 as compared to TG, PT and MTCMOS CVSL based multiplexer and decreased by order of 7 than that of MSL, Static CMOS, CPL and CVSL based multiplexers. The average power dissipation of PT based multiplexer is achieved as 537.1 nW, which is the lowest amongst all other multiplexers, as shown in Tab. 1.

Having explained the maximum, minimum and average power consumption for different configurations for the ease of inter-technique comparison, the values are tabulated in Tab. 1 and Tab. 2. As can be noticed from Tab. 1, the maximum power consumption is least for the MSL technique at 463.59 µW. However, the minimum power consumption for this technique is significantly high. The least minimum power consumption is observed for GDI based technique at 9.71 fW, but its maximum power consumption is the highest of all.

Another parameter, average power dissipation, is used to characterize the multiplexer structures. All the inputs are switching at different time instance, so the current will also be different at each time, which leads to different power dissipation value [28]. The average power is proportional to the energy required to charge and discharge the circuit capacitance. This power dissipation parameter is dependent on



(a) TG. (b) GDI.

(c) PT. (d) MSL.

(e) Static CMOS. (f) CPL.

(g) CVSL. (h) MTCMOS CVSL.

Fig. 4: Dynamic power dissipation for TG, GDI, PT, MSL, Static CMOS, CPL, CVSL and MTCMOS CVSL.

Tab. 1: Total and average power dissipation.

Multiplexer configuration	Maximum power dissipation (μW)	Minimum power dissipation (nW)	Average power dissipation (μW)
TG	523.36	0.1296	0.642
GDI	587.34	0.00000971	2.065
PT	493.74	0.22	0.5371
MSL	463.59	25.50	1.73
Static CMOS	478.63	19.27	1.305
CPL	579.63	26.14	3.01
CVSL	556.32	19.64	2.88
MTCMOS CVSL	527.8	0.828	1.93

© 2021 ADVANCES IN ELECTRICAL AND ELECTRONIC ENGINEERING

Tab. 2: Percentage analysis.

Multiplexer configuration	Percentage reduction in maximum power dissipation (μW) of MSL based multiplexer as compared to other multiplexer	Percentage reduction in average power dissipation (μW) of PT based multiplexer as compared to other multiplexer
TG	11.42 %	16.33 %
GDI	21.06 %	73.99 %
PT	6.106 %	—
MSL	—	68.95 %
Static CMOS	3.14 %	58.54 %
CPL	20.01 %	82.15 %
CVSL	16.66 %	81.35 %
MTCMOS CVSL	12.13 %	72.17 %

Tab. 3: Percentage analysis.

Multiplexer configuration	Delay (ps)	Average power (μW)	PDP (aJ)	Static power dissipation (nW)
Conventional TG	14.34 (A to O) 15.3 (B to O)	0.642	0.00981	98.63
GDI	18.49 (A to O) 15.38 (B to O)	2.065	0.0794	0.689
PT	14.83 (A to O) 14.79 (B to O)	0.5371	0.00796	102.86
MSL	55.12 (A to O) 56.64 (B to O)	1.73	0.0962	104.22
Static CMOS	51.24 (A to O) 54.68 (B to O)	1.305	0.0713	105.98
CPL	103.0 (A to O) 95.66 (B to O)	3.01	0.31	210.75
CVSL	166.3 (A to O) 151.8 (B to O)	2.88	0.478	162.62
MTCMOS CVSL	303.9 (A to O) 295.7 (B to O)	1.93	0.586	167.68 (S = 0) 78.95 (S = 1)

the switching frequency but is independent of the device parameters.

The average power dissipation is found least for PT based structure at 537.1 nW. The interesting thing to observe about the PT structure is its moderate maximum and minimum power. They are neither too high nor too low, making it the ideal technique in terms of power dissipation performance.

5. Delay and Power Analysis

The necessities of multiplexer for better performance include short propagation time and total power consumption, while static power dissipation ought to be minimum. The selection of multiplexer depends on the requirement of either propagation delay or power consumption. Table 3 explains the analysis of Delay, Average Power, PDP and Static Power Dissipation. By analysing the delay and power of different multiplexers, it is evident that the performance of each multiplexer is different, moreover, two of them presenting better performance compared to others in terms of propagation time, power consumption and leakage current.

Table 3 reveals that the propagation delay of TG based multiplexer is the shortest. For static condition, GDI based multiplexer outperforms with less amount of minimum power dissipation. Additionally, it dissipates minimum static power of 689 pW. On the other side, PT based multiplexer is slower by the value of 0.49 ps than TG multiplexer but it shows the lowest average power dissipation of 537.1 nW. In some condition, PDP of the circuit is considered as an important characteristic which ought to be minimum. Subsequently, the PDP of PT based multiplexer is minimum with a magnitude of 0.00796 aJ, thereby PT based multiplexer shows the optimum performance.

6. Conclusion

Eight structures of multiplexer implementation have been analysed for dynamic power consumption, delay, static power and power delay product. The maximum and minimum power dissipation are illustrated in Tab. 1. If average power dissipation is considered, then PT based multiplexer has achieved the best performance at 537.1 nW. The delay analysis recommends TG based multiplexer as the fastest multiplexer as it has the least delay recorded at 14.34 ps. Additionally, GDI based multiplexer registers the minimum static power dissipation of 689 pW. While in terms of average power dissipation, PT based multiplexer outperforms

TG based multiplexer but it is slower by the value of 0.49 ps than the TG multiplexer. The results obtained above strongly suggest that each multiplexer technique has its own merits. Some techniques demonstrate excellent power performance, while the other have faster operations, or are better at static performance. We believe that this comparison will be helpful for right choice of multiplexers in the design structure according to switching requirements, power consumption and occupied area.

Author Contributions

N.K. and P.M. conceived the idea and planned the complete work. N.K. has carried out all the simulations and plotted/analyzed all the results. M.M. and P.M. performed the calculations and analysed the data. B.R. and M.M. contributed to the interpretation of the results and modified the graphs. M.M. and B.R. designed the model and the computational framework. M.M. and N.K. drafted first write-up of the manuscript. P.M. and B.R. prepared the first revised manuscripts. B.R. and P.M. prepared the detailed Response to the Reviewer comments. M.M. and N.K. carried out the implementation and amended the modifications in the manuscript. M.M., N.K., B.R. and P.M. prepared second, third and fourth revised manuscripts. All authors provided critical feedback and helped shape the research, analysis and manuscript.

References

[1] SINGH, U., L. LI and M. M. GREEN. A 34 Gb/s Distributed 2:1 MUX and CMU using 0.18 µm CMOS. *IEEE Journal of Solid State Circuits*. 2006, vol. 41, iss. 9, pp. 2067–2076. ISSN 1558-173X. DOI: 10.1109/JSSC.2006.880630.

[2] CHEN, F.-T., J.-M. WU and M.-C. FRANK CHANG. 40-Gb/s 0.7-V 2:1 MUX and 1:2 DEMUX with Transformer-Coupled Technique for SerDes Interface. *IEEE Transactions on Circuits and Systems I: Regular Papers*. 2015, vol. 62, iss. 4, pp. 1042–1051. ISSN 1558-0806. DOI: 10.1109/TCSI.2015.2395634.

[3] MITTAL, P., Y. S. NEGI and R. K. SINGH. Impact of source and drain contact thickness on the performance of organic thin film transistors. *Journal of Semiconductors*. 2014, vol. 35, iss. 12, pp. 1–7. ISSN 1674-4926. DOI: 10.1088/1674-4926/35/12/124002.

[4] HAFIZ, M. A. A., L. KOSURU, M. I. YOUNIS and H. FARIBORZI. A 2:1 MUX Based on Multiple MEMS Resonators. *Procedia Engineering*. 2016, vol. 168, iss. 1, pp. 1642–1645. ISSN 1877-7058. DOI: 10.1016/j.proeng.2016.11.480.

[5] ETEZADI, S. and S. A. HOSSEINI. Novel Ternary Logic Gates Design in Nanoelectronics. *Advances in Electrical and Electronic Engineering*. 2019, vol. 17, iss. 3, pp. 294–305. ISSN 1804-3119. DOI: 10.15598/aeee.v17i3.3156.

[6] YAN, Z., W. ZHIGONG and L. WEI. 80 Gb/s 2:1 Multiplexer in 0.13-µm SiGe BiCMOS Technology. *Journal of Semiconductors*. 2009, vol. 30, iss. 2, pp. 1–4. ISSN 1674-4926. DOI: 10.1088/1674-4926/30/2/025008.

[7] CHANGCHUN, Z., W. ZHIGONG, S. SI, M. PENG and T. LING. 5-Gb/s 0.18-µm CMOS 2:1 multiplexer with integrated clock extraction. *Journal of Semiconductors*. 2009, vol. 30, iss. 9, pp. 1–6. ISSN 1674-4926. DOI: 10.1088/1674-4926/30/9/095009.

[8] MAJUMDER, A., M. DAS, S. K. SAW and B. K. BHATTACHARYYA. An energy efficient PVT aware novel CML-TG based mux-latch circuit serializes high rate data. *Microsystem Technologies*. 2021, vol. 27, iss. 2, pp. 555–568. ISSN 1432-1858. DOI: 10.1007/s00542-018-4093-x.

[9] JAIN, P. and S. AKASHE. Power delay optimization of nanoscale 4x1 multiplexer using CMOS based voltage doubler circuit. *Radioelectronics and Communications Systems*. 2016, vol. 59, iss. 11, pp. 477–488. ISSN 1934-8061. DOI: 10.3103/S0735272716110017.

[10] DIXIT, A., S. KHANDELWAL and S. AKASHE. Power Optimization of 8:1 MUX using Transmission Gate Logic (TGL) with Power Gating Technique. *International Journal of Computer Applications*. 2014, vol. 99, iss. 5, pp. 37–42. ISSN 2250-1797. DOI: 10.5120/17373-7911.

[11] MISHRA, M. and S. AKASHE. High performance, low power 200 Gb/s 4:1 MUX with TGL in 45 nm technology. *Applied Nanoscience*. 2014, vol. 4, iss. 3, pp. 271–277. ISSN 2190-5517. DOI: 10.1007/s13204-013-0206-0.

[12] PONNIAN, J., S. PARI, U. RAMADASS and O. C. PUN. A New Systematic GDI Circuit Synthesis Using MUX Based Decomposition Algorithm and Binary Decision Diagram for Low Power ASIC Circuit Design. *Microelectronics Journal*. 2021, vol. 108, iss. 1, pp. 1–23. ISSN 0026-2692. DOI: 10.1016/j.mejo.2020.104963.

[13] MORGENSHTEIN, A., V. YUZHANINOV, A. KOVSHILOVSKY and A. FISH. Full-Swing Gate Diffusion Input logic—Case-study of low-power CLA adder design. *Integration*. 2014, vol. 47, iss. 1, pp. 62–70. ISSN 0167-9260. DOI: 10.1016/j.vlsi.2013.04.002.

[14] MORGENSHTEIN, A., A. FISH and I. A. WAGNER. Gate-diffusion input (GDI): a power-efficient method for digital combinatorial circuits. *IEEE Transactions On Very Large Scale Integration (VLSI) Systems*. 2002, vol. 10, iss. 5, pp. 566–581. ISSN 1557-9999. DOI: 10.1109/TVLSI.2002.801578.

[15] YANO, K., T. YAMANAKA, T. YISHIDA, M. SAITO, K. SHIMOHIGASHI and A. SHIMIZU. A 3.8-ns CMOS 16*16-b multiplier using complementary pass-transistor logic. *IEEE Journal of Solid State Circuits*. 1990, vol. 25, iss. 2, pp. 388–395. ISSN 1558-173X. DOI: 10.1109/4.52161.

[16] PASTERNAK, J. H., A. S. SHUBAT and C. A. T. SALAMA. CMOS differential pass-transistor logic design. *IEEE Journal of Solid State Circuits*. 1987, vol. 22, iss. 2, pp. 216–222. ISSN 1558-173X. DOI: 10.1109/JSSC.1987.1052705.

[17] SUZUKI, M., N. OHKUBO, T. SHINBO, T. YAMANAKA, A. SHIMIZU, K. SASAKI and Y. NAKAGOME. A 1.5-ns 32-b CMOS ALU in double pass-transistor logic. *IEEE Journal of Solid State Circuits*. 1993, vol. 28, iss. 11, pp. 1145–1151. ISSN 1558-173X. DOI: 10.1109/4.245595.

[18] YANO, K., Y. SASAKI, K. RIKINO and K. SEKI. Top-down pass-transistor logic design. *IEEE Journal of Solid-State Circuits*. 1996, vol. 31, iss. 6, pp. 792–803. ISSN 1558-173X. DOI: 10.1109/4.509865.

[19] HU, X., A. S. ABRAHANM, J. A. C. INCORVIA and J. S. FRIEDMAN. Hybrid Pass Transistor Logic with Ambipolar Transistor. *IEEE Transactions on Circuits and Systems I: Regular Papers*. 2021, vol. 68, iss. 1, pp. 301–310. ISSN 1558-0806. DOI: 10.1109/TCSI.2020.3034042.

[20] GUPTA, I., N. ARORA and B. P. SINGH. Simulation and analysis of 2:1 multiplexer circuits at 90nm technology. *International Journal of Modern Engineering Research*. 2011, vol. 1, iss. 2, pp. 642–647. ISSN 2249-6645.

[21] YANG, C. and M. CIESIELSKI. Synthesis for mixed CMOS/PTL logic. In: *Proceedings Design, Automation and Test in Europe Conference and Exhibition*. France: IEEE, 2000, pp. 750. ISBN 0-7695-0537-6. DOI: 10.1109/date.2000.840883.

[22] PATEL, S. K., B. GARG and S. K. RAI. A power and area mux efficient approximate carry skip adder for rror resilient application. *Turkish Journal of Electrical Engineering & Computer Sciences*. 2019, vol. 528, iss. 1, pp. 443–457. ISSN 0167-9260. DOI: 10.3906/elk-1907-72.

[23] PAL, P. K. and R. K. NAGARIA. A Low-Power Low-Supply MOS-Only Subthreshold Voltage Reference for Wide Temperature Range. *Advances in Electrical and Electronic Engineering*. 2019, vol. 17, iss. 2, pp. 167–178. ISSN 1804-3119. DOI: 10.15598/aeee.v17i2.3197.

[24] HATANO, H. Single Event Effects on Static and Clocked Cascade Voltage Switch Logic (CVSL) Circuits. *IEEE Transactions on Nuclear Science*. 2009, vol. 56, iss. 4, pp. 1987–1991. ISSN 1558-1578. DOI: 10.1109/TNS.2009.2015663.

[25] JIAO, H. and V. KURSUN. Mode transition timing and energy overhead analysis in noise-aware MTCMOS circuits. *Microelectronics Journal*. 2014, vol. 45, iss. 8, pp. 1125–1131. ISSN 0026-2692. DOI: 10.1016/j.mejo.2014.05.006.

[26] DAS, M., A. MAJUMDER, A. J. MONDAL and B. K. BHATTACHARYYA. A 90nm Novel MUX-Dual Latch Design Approach for Gigascale Serializer Application. In: *2017 IEEE International Symposium on Nanoelectronic and Information Systems (iNIS)*. Bhopal: IEEE, 2017, pp. 210–214. ISBN 978-1-5386-1356-6. DOI: 10.1109/iNIS.2017.46.

[27] MISHRA, N., P. MITTAL and B. KUMAR. Analytical modeling for static and dynamic response of organic pseudo all-p inverter circuits. *Journal of Computational Electronics*. 2019, vol. 18, iss. 4, pp. 1490–1500. ISSN 1572-8137. DOI: 10.1007/s10825-019-01400-9.

[28] NADOOSHAN, R. S. and M. KIANPOUR. A novel QCA implementation of MUX-based universal shift register. *Journal of Computational Electronics*. 2014, vol. 13, iss. 1, pp. 198–210. ISSN 1572-8137. DOI: 10.1007/s10825-013-0500-9.

About Authors

Nishant KUMAR was born in Firozabad, India in 1993. He received his B.Tech. Degree in Electronics and Communication Engineering with 67.04 %

from Ajay Kumar Garg Engineering College, Ghaziabad, India in 2015. He has received his M.Tech. degree with CGPA of 8.46 from Delhi Technological University, New Delhi, India. His research interests include low power circuit design, memory circuits in emerging technologies.

Poornima MITTAL (Ph.D., M.Tech., B.Tech.), Member IEEE has published 110+ research papers in international journals and conferences of repute. Her research interest includes design/modeling of flexible electronic devices, thin film fabrication, memory and low power very large scale integration circuits. She has published one patent and a Text Book titled "Organic Thin-Film-Transistor Applications: Materials to Circuits" by CRC Press, Taylor & Francis in 2016. She is the reviewer of Institute of Electrical and Electronics Engineers (IEEE) transactions and other reputed international journals of IEEE, Institution of Engineering and Technology (IET), Elsevier, Institute of Physics Publishing (IOP), Wiley and Taylor & Francis. She has received the research awards in 2012, 2015, 2019, 2020 and 2021. She has delivered many expert talks and chaired sessions in the reputed international conferences. She is the life member of many professional societies. She has more than 15 years of academic and research experience. Presently, she is working as Professor in the Department of Electronics and Communication Engineering at Delhi Technological University, Delhi, India.

Bhawna RAWAT (M.Tech., B.Tech.) was born in Delhi, India in 1993. She received her B.Sc. Degree in Electronics and Communication Engineering from Shiv Nadar University, Greater Noida, India in 2015. She received M.Sc. degree in VLSI Design from Indira Gandhi Delhi Technical University for Women, Delhi, India. She is presently pursuing Ph.D. at Delhi Technical University, New Delhi, India. Her research interests include low power circuit design, memory circuits in emerging technologies.

Mudit MITTAL (M.Tech., MCA) is presently working as the Head of CS/IT Department, Institute of Technology and Management, Dehradun (Uttarakhand). He has 12 years rich experience of academics. His area of interest includes Digital Electronics, Automata Technology, Programming, Algorithm Analysis, Graph Theory and Data Structures. He has many international and national research publications.

© 2021 ADVANCES IN ELECTRICAL AND ELECTRONIC ENGINEERING

Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Effect of Rice Husk Ash, silica fume & GGBFS on compressive strength of performance based concrete

Mohit Gupta *, Dr. Ritu Raj, Dr. Anil Kumar Sahu

Department of Civil Engineering, Delhi Technological University, Delhi, India

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Performance based concrete
Mineral admixtures
Hardened concrete
Super plasticizers
Compressive strength

ABSTRACT

This investigation exhibits a study on the effect of silica fume, Rice husk ash and Ground granulated blast furnace slag on mechanical property of Performance based concrete. The Performance based concrete with various proportion of different mineral admixtures were tested including compressive strength at 7,14 & 28 days of curing. The ordinary Portland cement by weight was partially replaced with Rice Husk Ash (10%, 15%, 20%), Silica Fume (10%, 15%, 20%) and Ground granulated blast furnace slag (10%, 20%,40%). Estimations were completed subsequent to restoring at 20 °C and 65% relative humidity. Total 90 cubes were casted for different mixes for compressive strength of concrete. The test results reported that optimum % of Rice Husk Ash is for the Mix(C90RHA10) and the compressive strength increases by 6.62% compare to control mix at 28 days. Also, the optimum % of Silica Fume is for the Mix (C80SF20) and the compressive strength increases by 13.4%. Similarly, the optimum % of GGBFS is for the Mix (C60GGBFS40) and the compressive strength increases by 6.81%.
© 2021 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Integration of Advanced Technologies for Industry 4.0 : Materials Science.

1. Introduction

Concrete is the most broadly utilized development material everywhere throughout the world taking into account its strength aspect, deformed ability in any shape & stability against structural loads. In the present study the main focus is to determine the optimum use of different mineral admixtures with partial replacement of ordinary Portland cement in various proportions to assess the effect on performance based concrete. Industry across the world, concrete has major role in the effects of construction to our environment. High performance concrete is generally having a high strength comparative to the normal strength concrete. The most prominent characteristics of HPC is modulus of elasticity is very high along with durability i.e long lasting service life in built environment. Admixtures are generally classified as chemical admixture and mineral admixture. The chemical admixture further classified as Napthalene based water reducing admixture and polycarboxylate Ether based superplasticizer. Moreover, mineral admixtures are RHA, Metakaolin, SF, FA etc. When GGBFS content rises the temperature of HPC reduced along with cracking stress and cracking age decreased. Moreover, Silica fume generated from silicon industry and it is noncrystalline or amporphous powder form. Rice husk ash contains 85–90% of silica in it obtained from rice milling industry. GGBFS is having lower heat of hydration when used in concrete. The optimal content of GGBFS should not exceed 20% to assess early-age cracking and the proportion of GGBFS including is 0%, 20%, 35%, 50% and early-age cracking of HPC goes on increasing upto 20% under adiabatic condition [1]. By using local materials available U-HPC was developed and there was no increase in compressive strength by using silica fume and fine sand [2]. Resistance to sea water was accessed by influencing different mineral admixtures and microstructure paste is denser than without mineral admixture cement paste [3]. Max. compressive strength at 28 days increases by 0.1% when 20% SF & 10% glass fibre added with 70% natural sand [4]. As compared to flyash; ground granulated blast furnace showed superior mechanical performance at 28 and 91 days [5]. Flyash is generated from industry like thermal power plant as a waste which is generally used in concrete to reduce the carbon dioxide emission as it is replaced with cement and also improves the workability of concrete. Flyash is classified as calcium content such as type C & type F. On partial replacement of 10% and 20% with Flyash and SF the foam concrete was tested for 7 & 28 days and found the most superior property with less water absorption and high compressive strength [6]. Rice Husk ash also improving the mechanical property and effective for

* Department of Civil Engineering, Delhi Technological University, Delhi, India
E-mail address: mohit.g@live.com (M. Gupta).

https://doi.org/10.1016/j.matpr.2021.06.343
2214-7853/© 2021 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Integration of Advanced Technologies for Industry 4.0 : Materials Science.

Please cite this article as: M. Gupta, Dr. Ritu Raj and Dr. Anil Kumar Sahu, Effect of Rice Husk Ash, silica fume & GGBFS on compressive strength of performance based concrete, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2021.06.343

the sustainable environment on replacement of 8 to 10% cement [7]. Huge decrease of Ca(OH)2 in cement and as its greatest pore sizes and normal pore size become smaller & the improvement of its pore structure are the primary purposes behind the compressive strength increment of RHA concrete. [8]. The best way of utilising the rice husk ash in concrete based on optimized proportion as approximately 18–19% RHA can be yield from raw husk [9]. Impacts of GGBFS substitution proportion from 50% to 80% and compressive strength diminished with increase of slag content. [10]. Around 50% & above the proportion of GGBFS there is no improvement in the compressive strength of concrete [11]. Partial replacement of SF upto 10% does not reduce the workability and also upto 20 percent SF content the maximum strength is obtained [12]. The cement was replaced by Silica fume in proportion of 5% & 10% and it is observed that maximum compressive strength obtained at 10% replacement of SF [13]. Scope of the present investigation is limited to determine the effect of different mineral admixtures like RHA (10%,15%,20%), SF (10%,15%,20%) & GGBFS (10%,20%,40%) on performance based M40 grade of concrete and compare the test results with control mix for 7,14 & 28 days of curing. The pozzolanic reaction can be represented as;

Calcium hydroxide + silica = tricalcium silicate + water

$$3Ca(OH)2 + SiO2 = 3CaO \cdot SiO2 + 3H2O \quad (1)$$

A. Objectives of the study

i. To find the optimum % of Rice Husk Ash, Silica Fume & Ground granulated blast furnace Slag by partial substitution of ordinary Portland cement.

ii To determine the effect of different mineral admixtures and compare it with control mix on performance based concrete.

B. Research significance

Research on the replacement of cement as the cementitious material is gradually increasing these days because of the enormous scope creation of concrete has forced numerous ecological issues. Another issue in this quickly developing world is to include the toughness and the quality of the solid structures. Partial replacement of the cement has twin advantages, as it utilizes the industrial waste products like RHA, GGBFS and silica fume which otherwise requires land and resources for effective dumping as well as it prevents the excessive use of conventional cement. By incorporating supplementary cementitious materials in the concrete, CO2 emissions related with the concrete are decreased (Figs. 1–3).

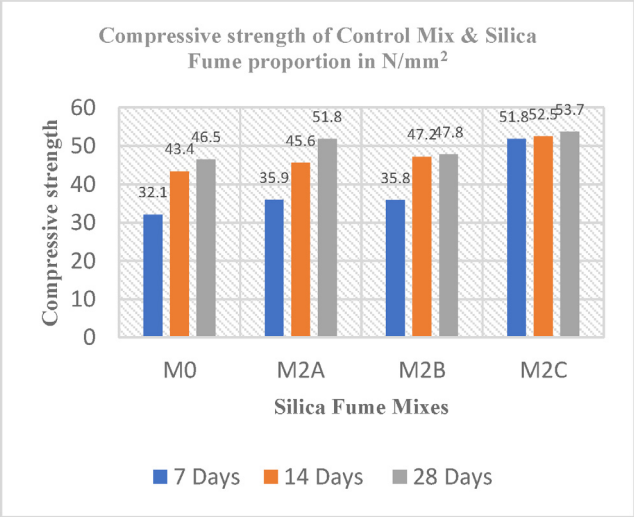


Fig. 2. Average compressive strength of control Mix & Silica Fume proportion.

2. Materials and experimental Procedure

Important materials used for making Performance based concrete are characterize and tested for different properties.

A. Cement: It was tested according to IS 4031:1988. It confirmed to IS 12269:1987.

B. Water: Potable water is used throughout the investigation.

C. Fine Aggregate: as per IS: 383-1970 was determined.

Sand belongs to Zone-II. Specific gravity of fine aggregate was determined as 2.65 and fineness modulus is 3.15.

D. Coarse Aggregate as per IS: 2386 was determined and important properties are found.

E. Superplasticizer: Polycarboxylate Ether based superplasticizer is used in this investigation and optimum dosage is quantify with Marsh cone.

3. Experimental Procedure

Table 3 shows the mix proportion of various mixes. M0 represents the control mix. Mix M1A, M1B, M1C represents

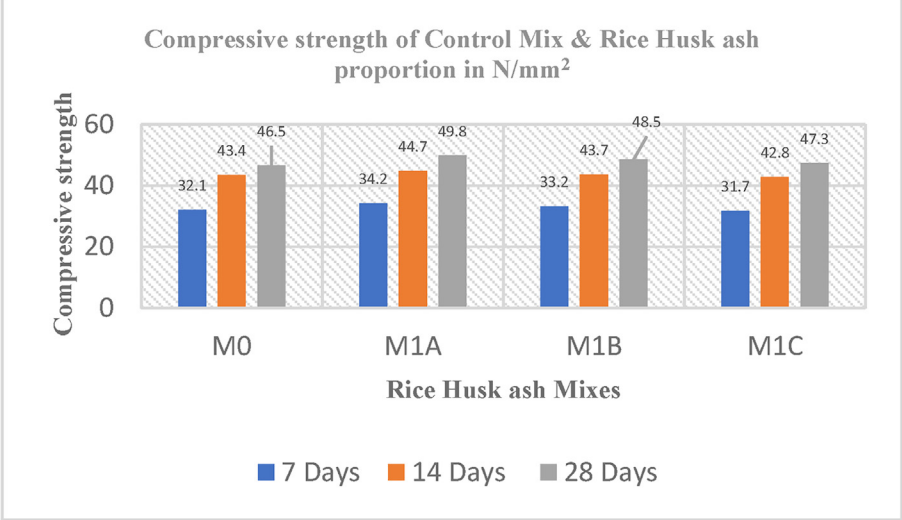


Fig. 1. Average compressive strength of control Mix & RHA proportion.

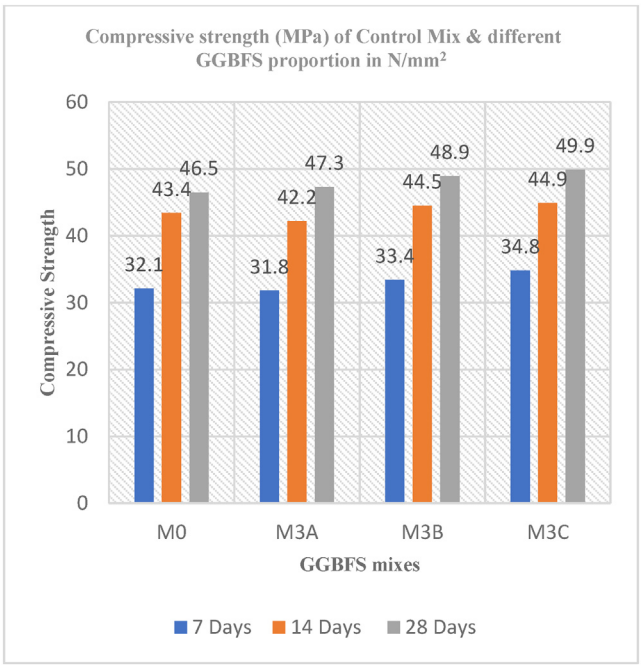


Fig. 3. Average compressive strength of control Mix & GGBFS proportion.

incorporation of RHA proportion. Also, M2A, M2B, M2C & M3A, M3B, M3C shows Silica Fume & GGBFS proportion respectively. The proportions of the mixture, of all of the period of cement are shown in Table 3. All blend ratios have been adapted with the aid of the particles, the density of each component of the mixture, to obtain a yield of 1 m3. The ordinary Portland cement by weight was partially replaced with Rice Husk Ash (10%, 15%, 20%), Silica Fume (10%, 15%, 20%) and Ground granulated blast furnace slag (10%, 20%,40%). Estimations were completed subsequent to restoring at 20 °C and 65% relative humidity.

4. Cube compressive strength procedure

Three of the samples (cubes with dimensions of 150 × 150 × 150 mm) were prepared for each mixture. After casting, the specimens were stored for 24 h at room temperature, and cover it with a wet gun sacks. The specimens were cured in water for 7, 14 and 28 days, after which they were tested for compressive strength. Concrete was poured into molds, and, correspondingly, tempered, so that there is no air voids. Within 24 h, the mold is removed and the sample is placed in the water to curing. The top surface of the samples must be flat and smooth. This is done through the placement of the concrete paste, and distributes it evenly over the entire surface of the sample. These samples have been studied by compression testing machine after 7 days, 14 days, and curing or 28 days curing. The load shall be gradually applied at a rate of 140 kg/cm2 for one minute, until the samples are fractured. As the load divided by the specimen surface and the strength of the concrete.

5. Results and discussions

A. Effect of Rice Husk Ash on Performance based concrete

Average compressive strength was determined by various partial substitution of RHA in proportion of 10%, 15%, 20% by weight of cement as shown in Table 1 . The Mix M1B represents the control mix with no replacement of RHA and average compressive strength is 32.18 MPa, 43.4 MPa, 46.5 MPa for 7,14 & 28 days of curing respectively. Mix M1A (C90RHA10) represents 10% RHA & 90% cement and average compressive strength is 34.2 MPa, 44.7 MPa, 49.8 MPa for 7, 14 & 28 days of curing. Mix M1B (C85RHA15) represents 15% RHA & 85% cement and average compressive strength is 33.25 MPa, 43.7 MPa, 48.5 MPa respectively for 7, 14 & 28 days of curing. Mix M1C (C80RHA20) represents 20% RHA & 80% cement and average compressive strength is 31.7 MPa, 42.8 MPa, 47.3 MPa for 7, 14 & 28 days of curing. The optimum % Rice husk ash is 10% for Mix M1A(C90RHA10) and average compressive strength is 4.12% higher than control mix M0. The similar observations were found by authors [7,8]. The amorphous phase content in RHA will also affect the activity index of the strength the ratio of the compressive strength of standard mortar cubes 80% made up with cement and 20% of additives, by mass, compressive strength of mortar is the standard cube operational information, the cement will be tested on the same age, etc.); however, there was no increase strength the index in connection with the growth of the amorphous phase, index above 55% [14].

Table1
Tests on Cement.

S. No.	Cement property	Experimental values	Acceptable Limits
1	Specific gravity	3.15	3.15
2	Standard Consistency	33	30%-34%
3	Initial setting time	32	Not less than 30 min.
4	Final setting time	600	Not more than 600 min.

B. Effect of Silica Fume on Performance based concrete

Average compressive strength was determined by various partial substitution of SF in proportion of 10%, 15%, 20% by weight of cement as shown in Table 2 . The Mix M0 represents the control mix with no replacement of SF and average compressive strength is 32.18 MPa, 43.4 MPa, 46.5 MPa for 7,14 & 28 days of curing respectively. Mix M2A (C90SF10) represents 10% Silica Fume & 90% cement and average compressive strength is 35.9 MPa, 45.6 MPa, 51.8 MPa for 7, 14 & 28 days of curing. Mix M2B (C85SF15) represents 15% Silica Fume & 85% cement and average compressive strength is 35.8 MPa, 47.2 MPa, 52.5 MPa respectively for 7,14 & 28 days of curing. Mix M2C (C80SF20) represents 20% Silica Fume & 80% cement and average compressive strength is 36.2 MPa, 47.8 MPa, 53.7 MPa for 7, 14 & 28 days of curing. The optimum % of Silica Fume is 20% for Mix M2C(C80SF20) and average compressive strength is 13.4% higher than control mix M0. As the specific surface area of silica fume is large as compared to other mineral admixtures therefore it involves in hydration process more actively and increases the pozzolonic reaction and forms more Calcium silicate hydrate gel. The similar results were found on effect of silica fume [4,6]. Increasing the compressive strength of cement slurry with silica fume, as a partial replacement for cement, helps to strengthen the bonds between the cement paste and the filler. Partial replacement of cement with silica and addition of superplasticizer increases the strength of the mortar, but does not affect the strength of the paste cement [15,10,11].

Table 2
Coarse Aggregate properties.

S.No.	Coarse Aggregate test	Values	Acceptable limit as per IS:2386
1	Specific gravity	2.74	2.7 to 2.8
2	Sieve analysis	–	It should be good.
3	Water absorption	1.15%	It should be less than 1.50%
4	Impact value	20%	It should be less than 30%
5	Crushing value	22%	It should be less than 30%

Table 3
Mix Proportion in Kg/m$^{3.}$

S.No.	Mix Designation	Cement	Fine Aggregate	Coarse aggregate	Water	Chemical admixture	RHA	SF	GGBFS
1	M0	435	772	1098	154	4.31	–	–	–
2	M1A(C90RHA10)	435	772	1098	154	4.31	43.5	–	–
3	M1B(C85RHA15)	435	772	1098	154	4.31	65.2	–	–
4	M1C(C80RHA20)	435	772	1098	154	4.31	87	–	–
5	M2A(C90SF10)	435	772	1098	154	4.31	–	43.5	–
6	M2B(C85SF15)	435	772	1098	154	4.31	–	65.2	–
7	M2C(C80SF20)	435	772	1098	154	4.31	–	87	–
8	M3A(C90GGBS10)	435	772	1098	154	4.31	–	–	43.5
9	M3B(C80GGBS20)	435	772	1098	154	4.31	–	–	87
10	M3C(C60GGBS40)	435	772	1098	154	4.31	–	–	174

C. Effect of Ground granulated blast Furnace Slag on Performance based concrete.

Average compressive strength was determined by various partial substitution of GGBFS in proportion of 10%, 20%, 40% by weight of cement as shown in Table 3 . The Mix M0 represents the control mix with no replacement of GGBFS and average compressive strength is 32.18Mpa, 43.4Mpa, 46.5Mpa for 7, 14 & 28 days of curing respectively. Mix M3A (C90GGBFS10) represents 10% GGBFS & 90% cement and average compressive strength is 31.8 MPa, 42.2 MPa, 47.3 MPa for 7, 14 & 28 days of curing. Mix M3B (C80GGBFS20) represents 20% GGBFS & 80% cement and average compressive strength is 33.4Mpa, 44.5 MPa, 48.9 MPa respectively for 7,14 & 28 days of curing. Mix M3C (C60GGBFS40) represents 40% GGBFS & 60% cement and average compressive strength is 34.8 MPa, 44.9 MPa, 49.9 MPa for 7, 14 & 28 days of curing. The optimum % of GGBFS is 40% for Mix M3C(C60GGBFS40) and average compressive strength is 6.81% higher than control mix M0. The similar results were reported for GGBFS [1,10,11]. Ground granulated blast furnace slag (GBS) has been used in the concrete mix for the ages. The majority of the work in made on GGBS concrete has demonstrated that it will give you an advantage of the fresh and reliable and qualities like durability aspects [1–6]. In addition, it may be to ensure the protection of the environment, and the economic benefits of the construction in [5,7,8]. However, it is a most common issue associated with GGBS is a slower rate of strength growth at an early age, such as concrete, consisting of POFA [5,7,8]. Conversely, some researchers have found that the triple-cement significantly improved properties in comparison with the binary cements [12].

D. Comparison of optimum % of RHA, SF, GGBFS with Control mix

The Compression strength test was performed to assess the effect of different mineral admixtures on performance based concrete. Table 4 shows the optimum % of different mineral admixtures in which Rice Husk Ash in Mix M1A(C90RHA10) shows average compressive strength is 4.12% higher than control mix M0. Similarly, optimum % of Silica Fume in Mix M2C(C80SF20) shows that average compressive strength is 13.4% higher than control mix M0. Also, the optimum % of GGBFS in Mix M3C (C60GGBFS40) shows average compressive strength is 6.81% higher than control mix M0. Moreover, Table 4 & Fig. 4 represents the maximum effect of Silica Fume at 20% replacement with cement & the maximum compressive strength of 53.7 MPa at 28 days than all other different mineral admixtures (Figs. 5–7 and Tables 5–7).

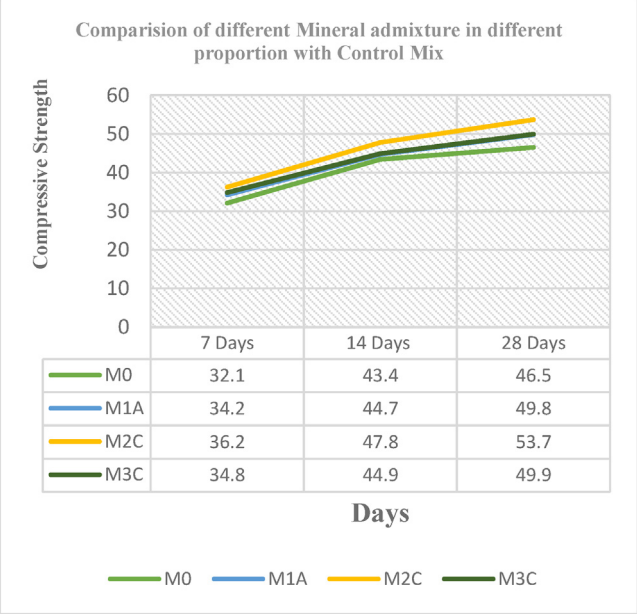


Fig. 4. Comparison of optimum % of RHA, SF, GGBFS with Control mix.

Table 4
Average compressive strength (MPa) of control mix & various proportion of RHA.

S.No.	Mix Designation	7 Days	14 Days	28 Days
1	M0(Control Mix)	32.18	43.4	46.5
2	M1A(C90RHA10)	34.2	44.7	49.8
3	M1B(C85RHA15)	33.25	43.7	48.5
4	M1C(C80RHA20)	31.7	42.8	47.3

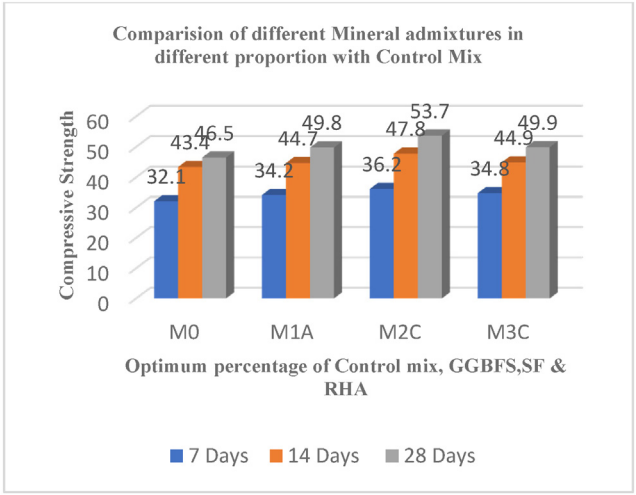


Fig. 5. Optimum % of RHA, SF, GGBFS with Control mix.

Fig. 6. Cube fracture of size (150 mm × 150 mm × 150 mm).

Fig. 7. (150 mm) size cube under Compression testing machine.

Table 5
Average compressive strength (MPa) of control mix & various proportion of SF.

S.No.	Mix Designation	7 Days	14 Days	28 Days
1	M0(Control Mix)	32.18	43.4	46.5
2	M2A(C90SF10)	35.9	45.6	51.8
3	M2B(C85SF15)	35.8	47.2	52.5
4	M2C(C80SF20)	36.2	47.8	53.7

Table 6
Average compressive strength (MPa) of control mix & various proportion of GGBFS.

S.No.	Mix Designation	7 Days	14 Days	28 Days
1	M0(Control Mix)	32.18	43.4	46.5
2	M3A(C90GGBFS10)	31.8	42.2	47.3
3	M3B(C80GGBFS20)	33.4	44.5	48.9
4	M3C(C60GGBFS40)	34.8	44.9	49.9

Table 7
Average compressive strength (MPa) of optimum % of RHA, SF, GGBFS with Control mix.

S.No.	Mix Designation	7 Days	14 Days	28 Days
1	M0(Control Mix)	32.18	43.4	46.5
2	M1A(C90RHA10)	34.2	44.7	49.8
3	M2C(C80SF20)	36.2	47.8	53.7
4	M3C(C60GGBFS40)	34.8	44.9	49.9

6. Conclusion

The following key points were observed during the experiment on the study of effect of different mineral admixtures on performance based concrete.

- There is a positive effect on the different mineral admixtures on the performance based concrete as compared to control mix.
- It has been observed that the mineral admixtures like rice husk ash, silica fume, GGBFS can be used as partial replacement of cement and helps to make environment sustainable by reducing CO2 emission to a certain extent.
- The optimum % of Rice Husk ash is 10% for Mix M1A (C90RHA10) and average compressive strength is 4.12% higher than control mix M0.
- The optimum % of Silica Fume is 20% for Mix M2C(C80SF20) and average compressive strength is 13.4% higher than control mix M0.
- The optimum % of GGBFS is 40% for Mix M3C(C60GGBFS40) and average compressive strength is 6.81% higher than control mix M0.
- The maximum effect of Silica Fume at 20% replacement with cement was observed & the maximum compressive strength of 53.7 MPa at 28 days obtained as compared to other different mineral admixture like RHA & GGBFS.

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.

Acknowledgements

The author would like to thank Head of Department of Civil Engineering, Delhi Technological University for providing Lab facilities and materials which were used during the research work.

References

[1] D. Shen, Y. Jiao, Y. Gao, S. Zhu, G. Jiang, Influence of ground granulated blast furnace slag on cracking potential of high performance concrete at early age, Constr. Build. Mater. 241 (2020) 117839, https://doi.org/10.1016/j.conbuildmat.2019.117839.

[2] A. Alsalman, C.N. Dang, W. Micah Hale, Development of ultra-high performance concrete with locally available materials, Constr. Build. Mater. 133 (2017) 135–145, https://doi.org/10.1016/j.conbuildmat.2016.12.040.

[3] C. Han, W. Shen, X. Ji, Z. Wang, Q. Ding, G. Xu, Z. Lv, X. Tang, Behavior of high performance concrete pastes with different mineral admixtures in simulated seawater environment, Constr. Build. Mater. 187 (2018) 426–438, https://doi.org/10.1016/j.conbuildmat.2018.07.196.

[4] S. Hemavathi, A. Sumil Kumaran, R. Sindhu, An experimental investigation on properties of concrete by using silica fume and glass fibre as admixture, Mater. Today Proc. 21 (2020) 456–459, https://doi.org/10.1016/j.matpr.2019.06.558.

[5] R. Yu, P. Spiesz, H.J.H. Brouwers, Development of an eco-friendly Ultra-High Performance Concrete (UHPC) with efficient cement and mineral admixtures uses, Cem. Concr. Compos. 55 (2015) 383–394, https://doi.org/10.1016/j.cemconcomp.2014.09.024.

[6] H.S. Gökçe, D. Hatungimana, K. Ramyar, Effect of fly ash and silica fume on hardened properties of foam concrete, Constr. Build. Mater. 194 (2019) 1–11, https://doi.org/10.1016/j.conbuildmat.2018.11.036.
[7] S. Hesami, S. Ahmadi, M. Nematzadeh, Effects of rice husk ash and fiber on mechanical properties of pervious concrete pavement, Constr. Build. Mater. 53 (2014) 680–691, https://doi.org/10.1016/j.conbuildmat.2013.11.070.
[8] R.S. Bie, X.F. Song, Q.Q. Liu, X.Y. Ji, P. Chen, Studies on effects of burning conditions and rice husk ash (RHA) blending amount on the mechanical behavior of cement, Cem. Concr. Compos. 55 (2015) 162–168, https://doi.org/10.1016/j.cemconcomp.2014.09.008.
[9] S.N. Raman, T. Ngo, P. Mendis, H.B. Mahmud, High-strength rice husk ash concrete incorporating quarry dust as a partial substitute for sand, Constr. Build. Mater. 25 (7) (2011) 3123–3130, https://doi.org/10.1016/j.conbuildmat.2010.12.026.
[10] E. Özbay, M. Erdemir, H.I. Durmuş, Utilization and efficiency of ground granulated blast furnace slag on concrete properties - A review, Constr. Build. Mater. 105 (2016) 423–434, https://doi.org/10.1016/j.conbuildmat.2015.12.153.
[11] A. Oner, S. Akyuz, An experimental study on optimum usage of GGBS for the compressive strength of concrete, Cem. Concr. Compos. 29 (6) (2007) 505–514, https://doi.org/10.1016/j.cemconcomp.2007.01.001.
[12] R. Duval, E.H. Kadri, Influence of silica fume on the workability and the, Cem. Concr. Res. 28 (4) (1998) 533–547.
[13] S. Singh, G.D. Ransinchung, P. Kumar, Effect of mineral admixtures on fresh, mechanical and durability properties of RAP inclusive concrete, Constr. Build. Mater. 156 (2017) 19–27, https://doi.org/10.1016/j.conbuildmat.2017.08.144.
[14] A. Shvarzman, K. Kovler, G.S. Grader, G.E. Shter, The effect of dehydroxylation/amorphization degree on pozzolanic activity of kaolinite, Cement Concr. Res. 33 (3) (2003) 405–416.
[15] Toutanji, H. A, Tahar El-Korchi. (1995). the Influence of Silica Fume on the Compressive. Cement Concr. Res., 25(7), 1591–1602.

Trans Indian Inst Met
https://doi.org/10.1007/s12666-021-02326-2

ORIGINAL ARTICLE

Effect of Welding Parameters on Mechanical Properties of Cold Metal Transfer Welded Thin AISI 304 Stainless-Steel Sheets

Jayanta Ghosh Roy[1] · N. Yuvaraj[1] · Vipin[1]

Received: 12 April 2021 / Accepted: 6 June 2021
© The Indian Institute of Metals - IIM 2021

Abstract Cold metal transfer (CMT) allows thin plates to be joined with minimal distortion and without spattering. In this paper, the CMT process has been used to weld AISI 304 stainless steel. To investigate the weld strength properties, various welding process parameters such as welding current, welding speed, and contact-to-work distance (CTWD) were chosen. The results show that there is no fracture on the weld joint; the tensile residual stress levels of all the samples are in a controlled manner. FESEM images reveal a dimpled morphology and crack-free tensile fracture surfaces in welded joints. Welding speed is the most significant welding process parameter, followed by welding current and CTWD on the tensile strength of the welded joint. The joint welded at a welding current of 95 A, a welding speed of 4 mm/sec and CTWD 7 mm exhibits higher weld strength compared with other welded joints.

Keywords CMT · AISI 304 · Tensile strength · Hardness · Residual stresses

✉ N. Yuvaraj
yuvraj@dce.ac.in

Jayanta Ghosh Roy
jghoshroy@gmail.com

Vipin
vipin@dce.ac.in

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

1 Introduction

AISI 304 is widely popular among all austenitic stainless steels, also known as 18/8 steel, and consists of 18% chromium, which makes it highly corrosion-resistant, with good toughness and ductility. The role of 8% Ni is to stabilize the austenite, which promotes ductility, which, in the absence, can lead to forming a magnetic martensitic structure. AISI 304 stainless steel is being used in several areas such as food processing, dairy equipments, architectural structures, chemical, rail, transportation, aerospace industries, nuclear power plants, pressure vessels, cryogenic, petrochemical units and sanitary appliances because of their excellent properties like strength, non-magnetic, lower thermal and electrical conductivity in comparison with other carbon steels [1–4]. The consumption of 304 and 304L stainless steel is approximately 20,000 tons per year according to the global market size [5]. Traditionally carbon steels are welded by using MIG/MAG, TIG and SMAW process providing high heat input, which is the cause for the creations of coarse grains and wider heat-affected zone [6]. Maintaining constant heat energy is difficult for these conventional welding processes. During the welding process, the temperature variation in the parent metal's heat-affected zone affects the dimensional accuracy of the welded material and micro-characteristics of the weld joint [7]. GMAW-modified CMT welding process provides low heat input phenomena during welding process, which is very much suitable for welding of thin materials with a smallest amount of distortion, less dilution rate, less structural stress with lower residual stress in the weld area [8, 9]. The difference between conventional GMAW and CMT is the liquid droplet mode, which is detached by the electrode wire [10]. In conventional GMAW, Lorentz forces are causing droplet detachment.

Published online: 04 July 2021

Wire retraction motion with low current input is responsible for droplet detachment in CMT with spatter less welding [11, 12]. In the CMT process, a stable arc is formed due to precise control of arc length and wire feed rate. Koli et al. [13] found lower bead penetration in pulse MIG and manual MIG in comparison with the CMT process with the same current. Jan et al. [14] compared CMT mode arc with standard and pulsed arc modes in laser hybrid welding of thicker steel sheets and observed that CMT mode produces a narrower HAZ in comparison with other processes. The reduction of power with the higher stability of beads and lesser HAZ is advantageous in use on CMT arc mode. Kannan et al. [15] observed that AISI 316L thin sheets welded by CMT process with lower welding speed (300 mm/min) and higher welding current (105A) yield the best weld profile with sound joints for Tailor Welded Blanks. Stanciu et al. [16] reported that in CMT welding of 1-mm low S235JR low carbon steel sheet, by increasing the welding speed, improper penetration take place and by reducing the wire speed, reinforcement also get reduced. Luchten-Berg et al. [17] studied different heat inputs of S32205 duplex stainless steel with CMT process; the change in heat input changes the ferrite/austenite balance. Higher heat energy increases the intergranular austenite formation and decreases the formation of ferrite. Zhang et al. [18] investigated Q235 zinc-coated low-carbon steel with 6061 aluminum alloys through the CMT process. The strength of the joint increases with a lower heat input.

Kumar and Shahi [19] studied the dissimilar GTAW of 316L SS with S32205 duplex SS with different heat inputs. Larger grain size of the weld metal is observed with a higher heat input. Lesser heat input is selected to reduce HAZ and grain coarsening to improve mechanical properties. Kannan et al. [20] reported CMT welding of 2-mm 316L SS sheets with a different arc correction factor for finding the tensile properties. When the arc length correction increases, the filler reinforcement also increases. Tensile strength of the weldment rises corresponding to the arc length correction increase from 0 to 20%. Ahsan et al. [21] studied the CMT GMAW process with three different filler wires with varying deoxidizing elements such as silicon and manganese for detecting the oxygen content in the slag and weld pool. A lower quantity of deoxidizers produces a good-quality weld joint with greater than 0.006 wt% of oxygen in the weldment. An increase in oxygen changes the temperature gradient of surface tension with lower deoxidizers. Dissimilar stainless-steel joint was performed by Balbande et al. [22] with CMT for 3-mm butt weld. Fractography analysis reveals that fine dendritic Mo content structure is resulted in the crack-free weld joint. Zhou et al. [23] used ER 308L filler wire to weld the AISI 430 ferritic stainless steel and studied the effect in the size of grain in the weldment area and heat-affected zone. The tensile strength of welded samples does not show any remarkable changes as compared with the base metal. The strength of the HAZ mainly depends upon the formation of intergranular martensite and dispersion of carbide generates a strengthening effect, comparatively than the formation of ferrite grain size. Dharmik et al. [24] found that CMT-welded CRNGO 0.5-mm thin electrical sheets have a lesser heat-affected zone in comparison with MIG and GTAW welding due to lesser heat input and lower grain size. Madhavan et al. [25] found CMT welding of Al 6061 alloy and AZ31B magnesium alloy and revealed that pitting corrosion resistance has been found with high heat input, resulting in lower tensile strength due to compressive residual stress in the weld zone. In welding of carbon steel and stainless steel, Akbari and Sattari [26] found that with reduction of heat input, the residual stress distributions in the substrate of stainless-steel side reduces much in comparison with the carbon-steel side. Yagati et al. [27] studied the effect of Al 4043 and Al 4047 filler wire on joining of Al 6061-T6 to IF steel by CMT brazing and found that joint made with Al 4047 filler wire has been recorded in a superior quality. Babu et al. [28] investigated the lap joining of AA2219 to AISI 321 steel by CMT process. The weld joint strength depends on appropriate Al coating thickness on the steel prepared by friction surfacing before welding. Girinath et al. [29] studied the effect of three different torch angles (-10^0, 0^0, $+10^0$) on the formability of AA5052 alloy by CMT welding. Weld blank prepared by -10^0 torch angle has more formability than other torch angles due to arc preheating.

Residual stress is measured by two techniques, namely destructive and non-destructive. The most common destructive methods are the hole-drilling technique, deep hole method, sectioning technique, ring-core method and contour method. Hole-drilling is an easy and convenient process. It is capable of determining the direction, magnitude, sense of residual stress distribution across a thickness. The ring core method is utilized for measuring the deformation of the material in a central area. This method has the superiority than the hole-drilling method, which provides better surface strains. To measure the deep interior stresses, the deep hole technique is utilized; it has combined features of hole-drilling and ring-core methods. The contour method is simpler to use because it requires almost no calculations [30–33]. The non-destructive techniques are based on diffraction methods. The most common non-destructive methods are X-ray diffraction, neutron diffraction, Barkhausen noise method and ultrasonic method. In the X-ray diffraction method, the inter-planar atomic spacing can be precisely measured, and overall stress on the metal can then be calculated. The neutron diffraction technique can measure large thickness material with larger penetration in comparison with the X-ray

Table 1 Chemical composition of base material and filler wire (% weight)

Materials	Fe	Cr	Mn	Ni	Si	C	P	S	Mo	V	Ti
AISI304	Bal	17.90	1.88	8.38	0.254	0.043	0.046	0.016	0.224	0.072	0.006
ER308L	Bal	19.15	1.42	10.03	0.422	0.029	0.022	0.012	0.256	0.032	0.003

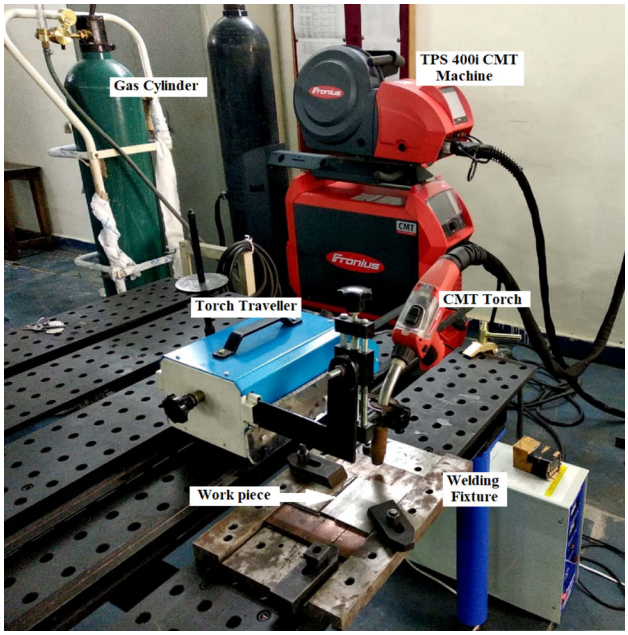


Fig. 1 Welding experimental setup

Table 2 Welding parameter and their levels

Process parameter	Current	Welding speed	CTWD
Unit	Ampere	mm/sec	mm
Symbols	I	S	L
Level 1	85	4	3
Level 2	90	6	5
Level 3	95	8	7

diffraction method. The magnetic Barkhausen noise method is used to measure surface residual stress of ferromagnetic materials. The ultrasonic method is employed to measure residual and applied stresses of the interior depth of material. In addition, this technique is simple, portable, economical and free from radiation hazards [33–35].

Pandey et al. [36] used the blind hole-drilling process for measuring the axial and transverse residual stresses of various P91 steel welded plates. Taraphdar et al. (2020) [37] introduce a novel deep hole contour method to measure residual stress. This method can measure through

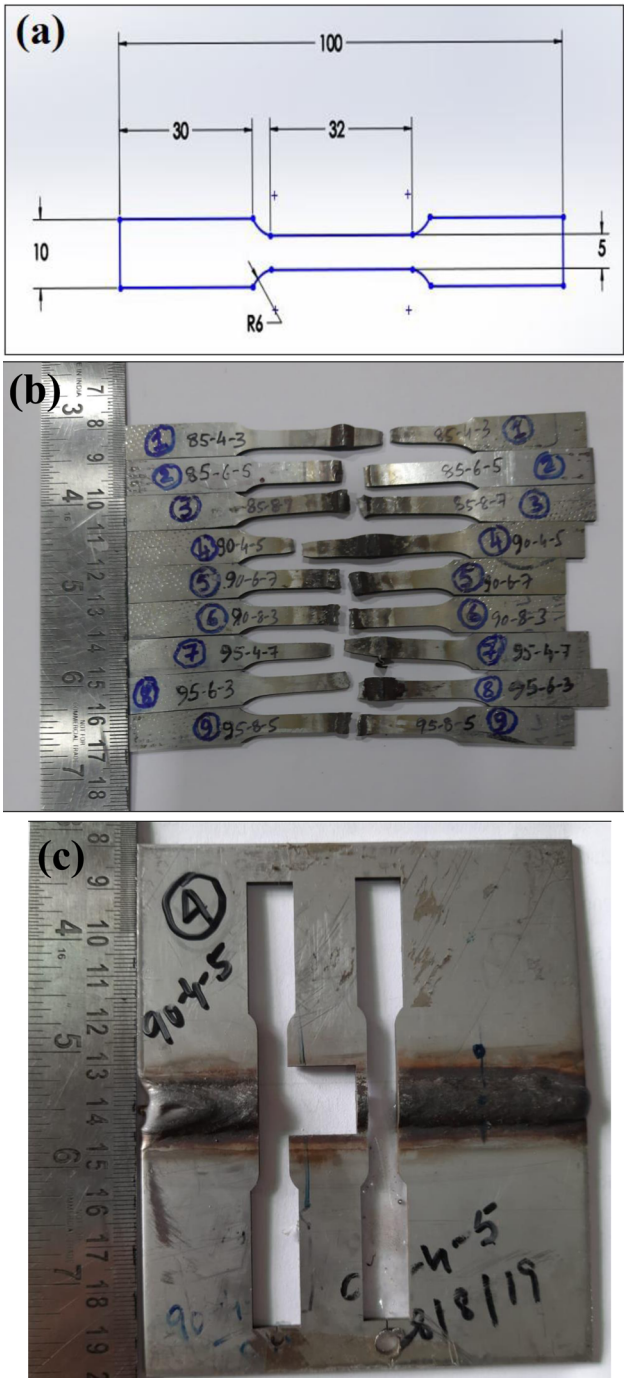


Fig. 2 (**a**) Tensile specimen dimensions, (**b**) CMT tensile samples after fracture and (**c**) tensile specimen extracted from welded joint

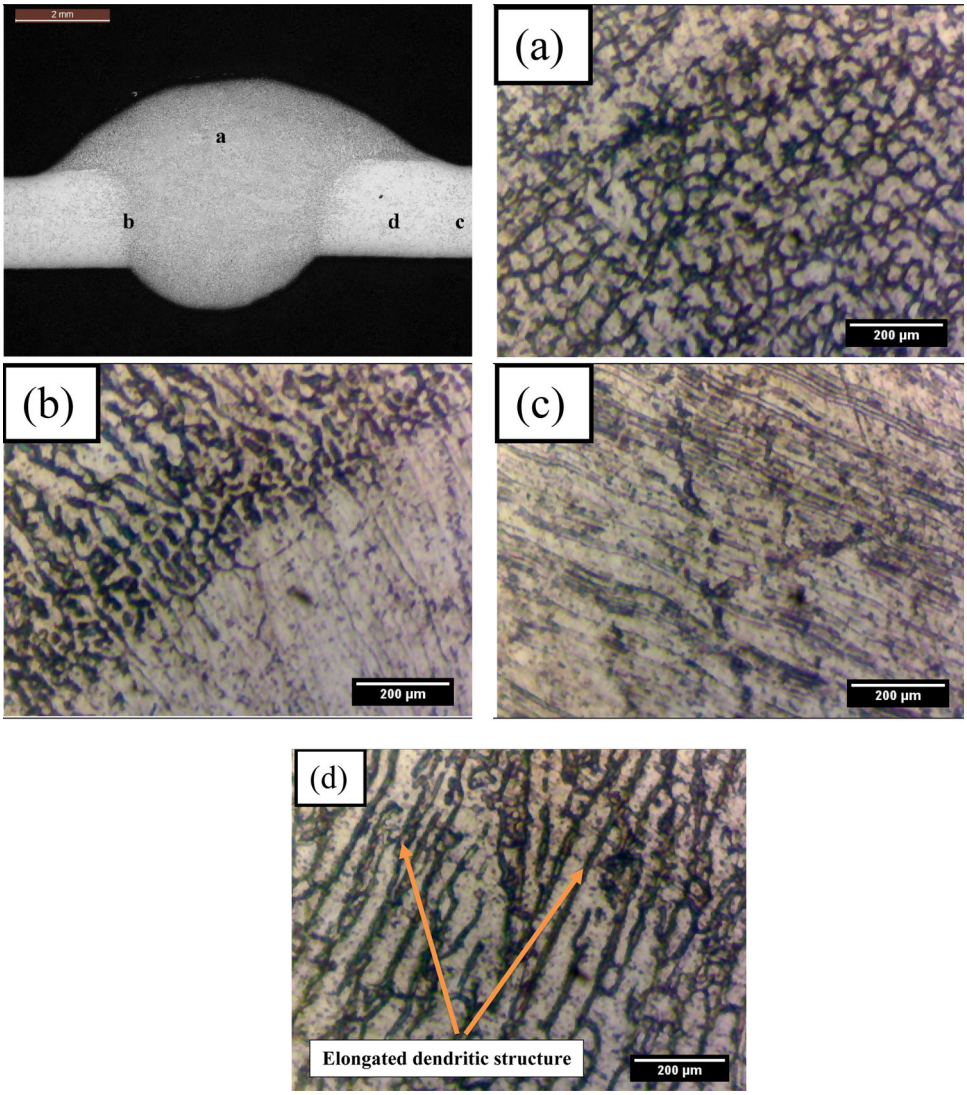


Fig. 3 Microstructure of sample 7 at various positions: (**a**) weld bead, (**b**) fusion line, (**c**) base metal zone and (**d**) HAZ

thickness residual stress distribution with a lower degree of specimen damage than the deep hole-drilling technique and contour process methods. Banik et al. [38] evaluated the residual stresses of thick welded austenite stainless-steel plates by blind hole and deep hole-drilling techniques. The residual stresses are higher throughout the thickness of higher strength welded material in comparison with lower strength welded material. Lai and Wu [39] investigated residual stresses of three different welded low-carbon steels by two different X-ray stress analyzers, namely μ-X360 and iXRD diffractometer. Both techniques showed identical results with an "M"-shaped common profile. It is evident with reference to the literature that welding of thin stainless-steel sheets had consistent demand in the past, which will be having good demand in the future due to their excellent mechanical properties. Limited research work has been carried out with low thickness AISI 304 stainless-steel sheets with various CMT process parameters. The present work aims to examine the influence of optimum process parameters to achieve maximum mechanical properties. Taguchi L9 orthogonal array technique is applied to find out optimum welding parameters. Input welding parameters, namely welding current, speed and CTWD, are considered to obtain maximum tensile properties. In the welded joint, microhardness, residual stresses and NDT radiographic technique have also been investigated for finding the quality of the joint.

2 Materials and Methods

The rolled sheet of AISI 304 SS of sizes 100 mm × 60 mm × 2 mm is considered as a base material. The filler wire ER 308L SS diameter of 1.2 mm is used to join. The chemical composition of the AISI 304 SS and the ER 308L SS is given in Table 1. TPS 400i CMT manual welding

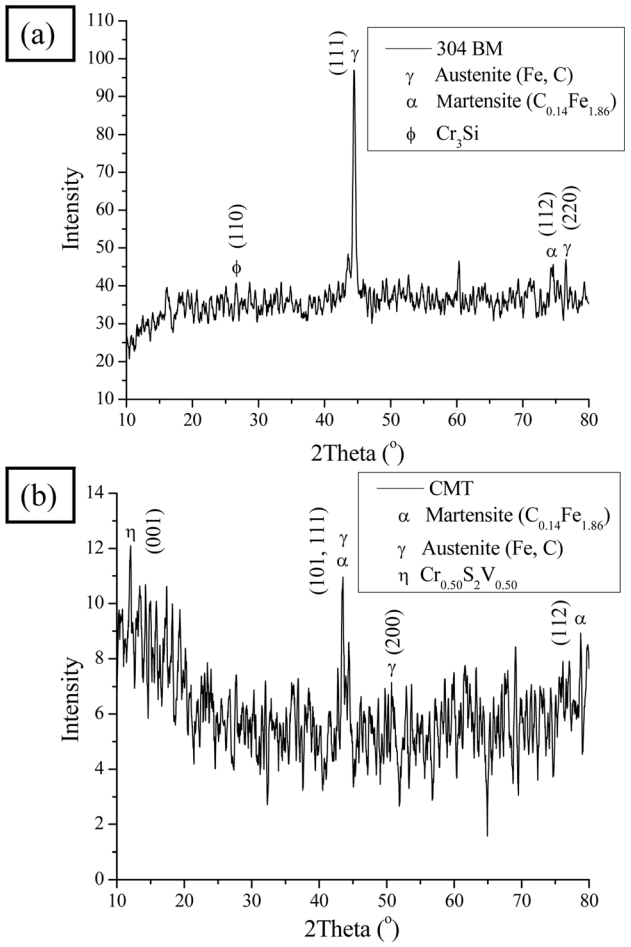


Fig. 4 XRD images of (**a**) base material and (**b**) weld bead

machine is used to conduct the experiments. Figure 1 shows the experimental setup. A large number of preliminary experiments were carried out to find the upper and lower limits of welding parameters. The selected process parameters and their levels are presented in Table 2 as per Taguchi L9 orthogonal array. With the set parameters of current, welding speed and CTWD, the machine automatically decided the voltage, which was in the range of 10–20.8 V. Pure argon gas (99.99%) of 14 L/min flow rate was considered for all the experiments. The microstructure was observed with the help of an optical microscope (Make OLYMPUS). The standard metallographic procedure was followed for preparing the test specimens. Etchant composition of 30 ml HCl + 120 ml distilled water + 10 g $FeCl_3$ was used for observing the microstructure [5]. Vicker's microhardness tester (FM-700 Future Tech) was used to estimate the hardness of the welded joint samples with an applied load of 500 gm for 20 s. Tensile tests were performed at a speed of 1 mm/min by UTM (Tinius H50KS). The specimens were prepared as per ASTM: E8/E8M-011 standards with a gauge length of 25 mm as shown in Fig. 2a. The broken tensile specimens are shown in Fig. 2b. Figure 2c shows the tensile specimen extracted from welded joint. For comparing tensile test results with pulse MIG welding, selective samples were welded with the same parameters. The UTS (MPa), yield strength (MPa) and elongation (%) of the base material were 732.3 MPa, 555.6 MPa and 60% respectively. The hardness of the parent metal was found at 238 HV. Residual stresses of the welded joints of various regions were measured by μ-X360n Pulstec High-Resolution Full 2D X-ray Diffraction (HR-XRD) machine. A standard Cr x-ray tube (30 kV and 1 mA) with a collimator size of 1 mm and measuring conditions are included in the device specification. This equipment works on the Cosα method, and it obtains a full Debye–Scherrer ring to collect the entire diffraction through a 2D detector [40].

3 Discussions and Results

3.1 Microstructure

Figure 3 depicts the microstructure of various regions of a welded specimen, (a) weld bead, (b) fusion line, (c) base metal region and (d) HAZ. The hardness and tensile strength of SS 304 mainly depend upon the microstructural evaluation and solidification behavior of welded material [41, 42]. During welding, the filler wire fuses with parent metal in a liquid state as a result of refined grain size that occurs in weld bead, which is seen from Fig. 3a. In the HAZ, the size of the grain size has increased due to the softening of material during welding, which is seen from Fig. 3b, d. The solidification of material after welding causes the formation of dendrites. Pandey [43] also observed the dendrite and inter-dendritic structure in HAZ of SS304 welded material. The presence of Cr- and Ni-rich elements in the HAZ is accountable for the formation of dendrite phases. Kumar and Shahi [1, 19] reported that different heat inputs in GTAW welding of AISI 304 SS change the mechanical properties and microstructural behavior of the welded joints. Larger heat input is experienced for the formation of the dendrite and inter-dendrite structure in welded joint. The XRD confirms the austenitic γ (111) and martensite α (110) in various phases of base material 304SS [44]. The weld bead consists of austenite (Fe, C) and martensite ($C_{0.14}$ $Fe_{1.86}$), phases as seen in Fig. 4. The intensity of γ (111) plane in welded material has differed in comparison with the base material. The γ (111) phase of 304SS has crystal growth in the preferred orientation direction [45]. Figure 5 shows the FESEM with EDX results of the CMT weld bead zone. FESEM image shows the combination of austenite and martensite phases. EDX results confirm the presence of Cr and Ni elements in

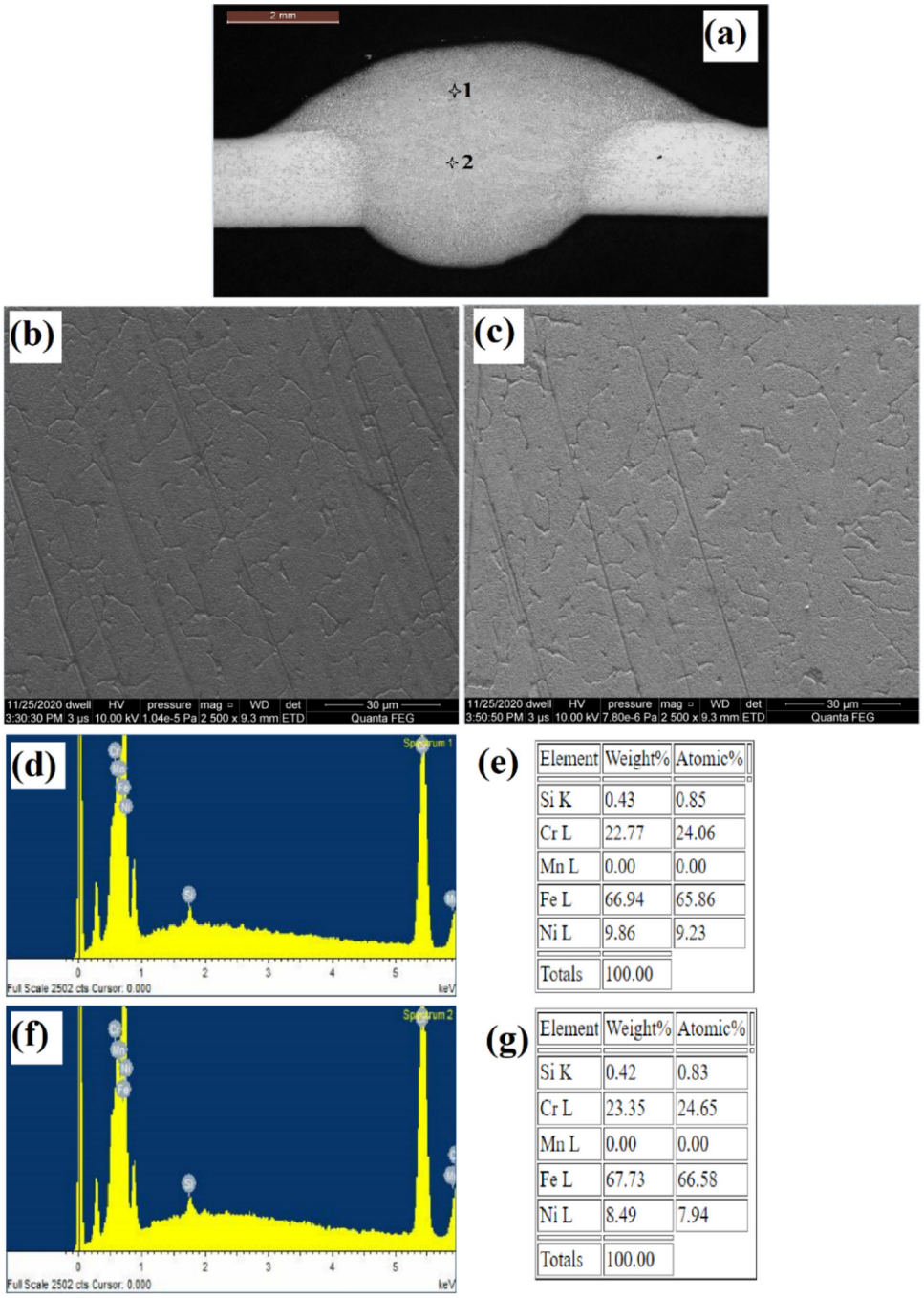


Fig. 5 FESEM and EDX results of CMT at weld bead (**a**) The points indicate FESEM image taken on weld bead (**b**) and (**c**) FESEM image at point 1 and point 2, respectively, (**d**) and (**e**) spectrum image of point 1 with its composition in weight %, (**f**) and (**g**) spectrum image of point 2 with its composition in weight %

the weld region, which is transformed from filler material as required by the chemical composition weight % of 304 SS.

3.2 Microhardness

The microhardness value of the welded samples taken at various points is shown in Fig. 6. It is seen from the figure that the hardness of the HAZ region and weld bead is less in comparison with the base metal. Approximately 16% and 25% decrease in the microhardness is experienced in the HAZ and weld bead, respectively, as compared with the base metal region. The decrease in hardness in the weld region and HAZ is due to the formation of dendrites, which is seen from microstructural image as shown in Fig. 3a, d. The weld zone experiences a small dendrite, whereas HAZ experiences an elongated dendritic structure. In the interface region, the grain size becomes coarser due to softening behavior of higher heat input, which is seen from Fig. 3b. At higher input, austenitic stainless steel undergoes

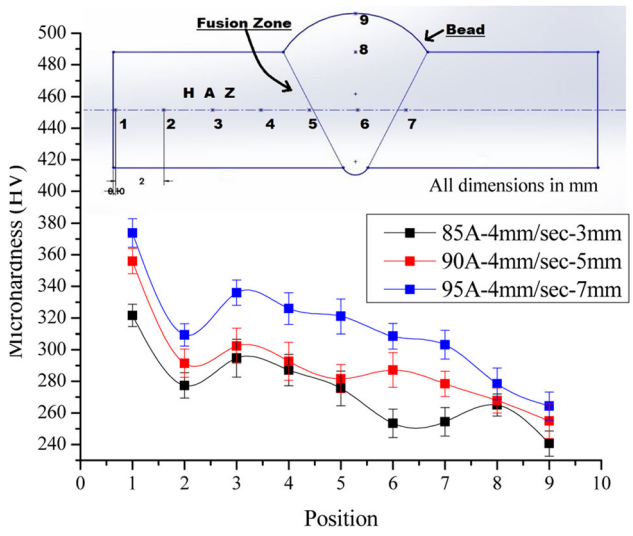


Fig. 6 Microhardness variation of welded joints

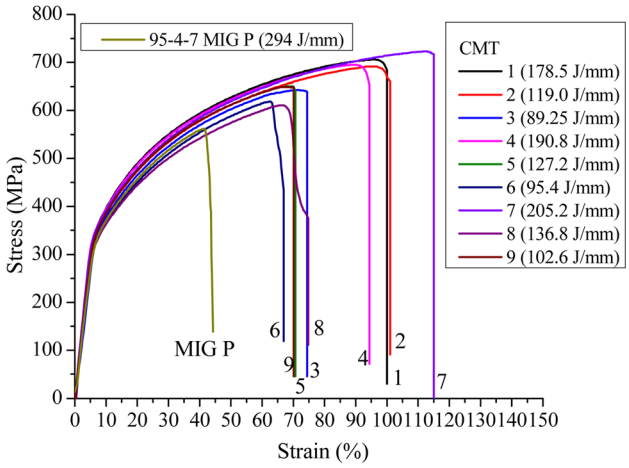


Fig. 7 Stress vs strain graphs of CMT and MIG P welded joint

softening behavior, which deteriorates the tensile strength and microhardness [19].

3.3 Tensile Properties

The tensile tested results are exhibited in Table 3. Higher tensile strength and ductility are observed in welded samples due to appropriate heat input during welding. Welded joint strength mainly depends upon the heat input during the welding process. The heat input equation is given below [46].

$$\text{Heat Input } (Q) = \eta \frac{VXI}{S} \quad (1)$$

where efficiency of the welding (η); voltage (V); current (I); welding speed (S). The welding efficiency in this study has been assumed as 80% [13]. Lesser strength welded samples occur due to the formation of higher brittle intermetallic compounds in the fusion region, which is seen from the XRD and EDX plots of the weld bead.

Heat input is mainly dependent on welding current, speed and voltage. Welding of larger thickness material requires higher heat input. However, welding speed is also dependent on thickness. Too high or too low welding speed results in inadequate weld penetration. The microstructure of the weld bead is affected by heat input, and microstructural properties of steels are greatly influenced by the austenite/ferrite balance in the welded joint. Further, the refinement of grain is also governed due to controlled heat input, observed from the microstructure. In this process, swift retraction of filler wire during short-circuit process takes place to manage the heat input. With the welding parameters (95 A current, 4 mm/sec welding speed and 7 mm CTWD), one sample is also welded with MIG pulse welding process in comparison with CMT sample 7 as shown in tensile graph in Fig. 7. The results reveal that due to higher heat input, the tensile properties of MIG welded samples are decreased. Higher heat inputs

Table 3 Welding parameters and experimental test results

S.No.	I (A)	S (mm/sec)	CTWD (mm)	Q (J/mm)	UTS (MPa)	Elong. (%)	Fracture location	Joint efficiency (%)	MH_b ($HV_{0.5}$)	S/N ratio
1	85	4	3	178.50	706	100	HAZ	96.40	253.0	44.0661
2	85	6	5	119.00	691	101	WM	94.36	273.1	44.2200
3	85	8	7	89.25	643	74.4	WM	87.80	267.4	41.8252
4	90	4	5	190.80	695	94.4	BM	94.90	296.3	43.7787
5	90	6	7	127.20	651	70.7	WM	88.89	289.1	41.4593
6	90	8	3	95.40	619	66.9	WM	84.52	291.4	41.0087
7	95	4	7	205.20	723	115	HAZ	98.73	279.4	45.2125
8	95	6	3	136.80	611	74.8	WM	83.43	281.3	41.8922
9	95	8	5	102.60	649	70.1	WM	88.62	289.3	41.3902

I Welding current; *S* welding speed; *CTWD* contact tip-to-workpiece distance; *Q* heat input; *UTS* ultimate tensile strength; *Elong.* elongation; MH_b microhardness at weld bead

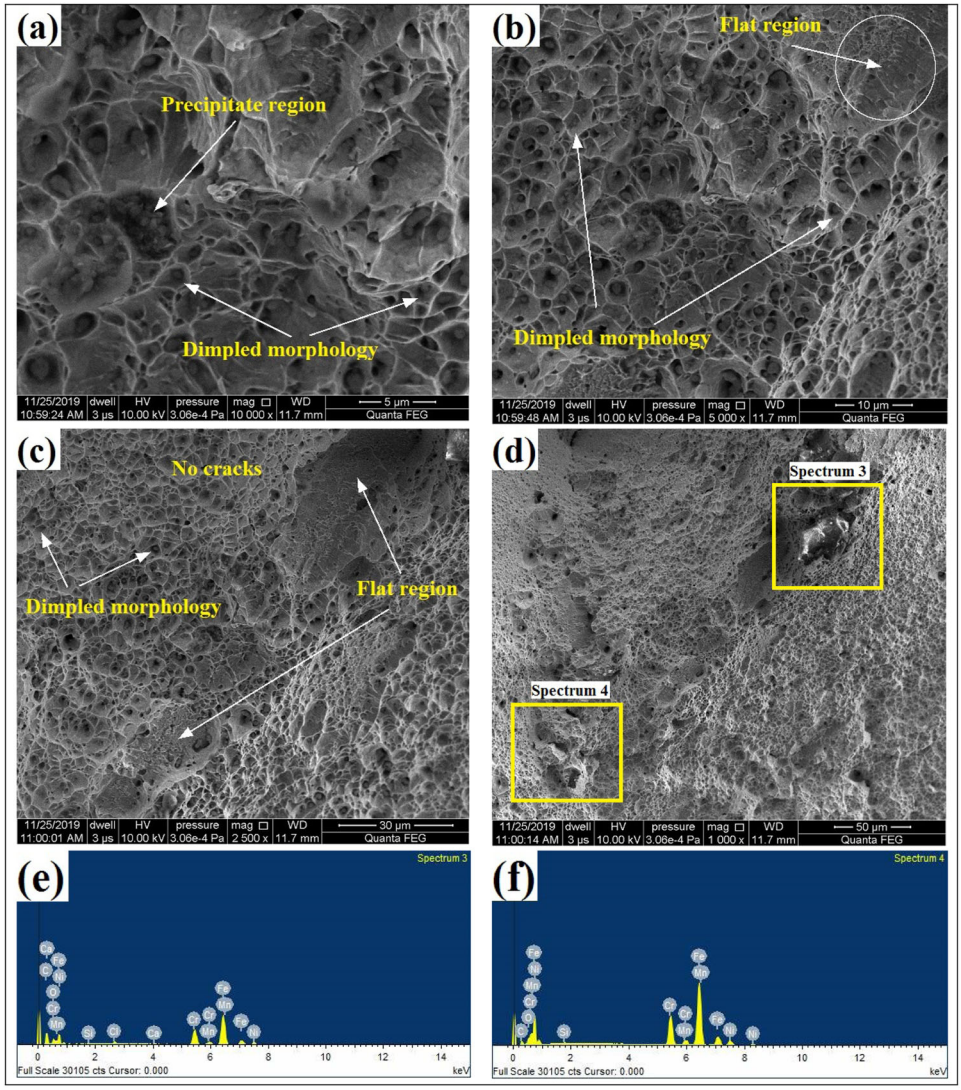


Fig. 8 FESEM images of the fracture surface of sample 7 taken at various magnifications and EDX was taken at spectrum 3 and 4 points

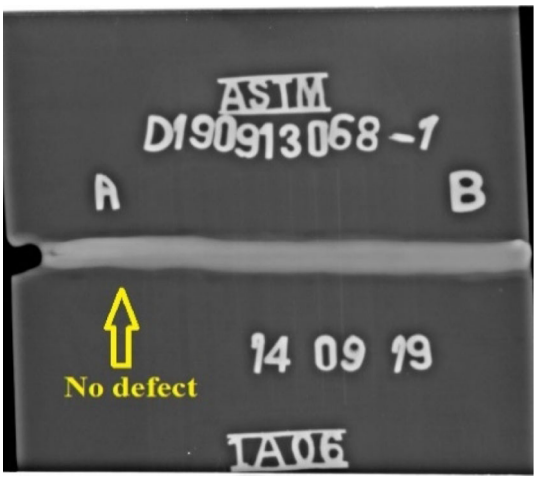


Fig. 9 Radiography test as per ASME Sec.-V:2017 of sample 7

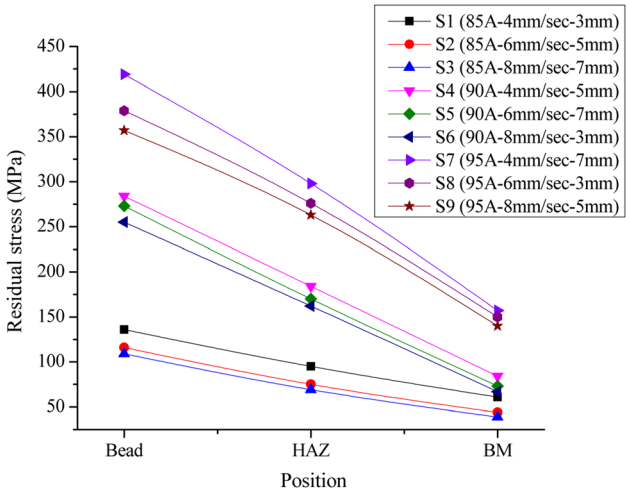


Fig. 10 Residual stress graph of welded samples taken at various points

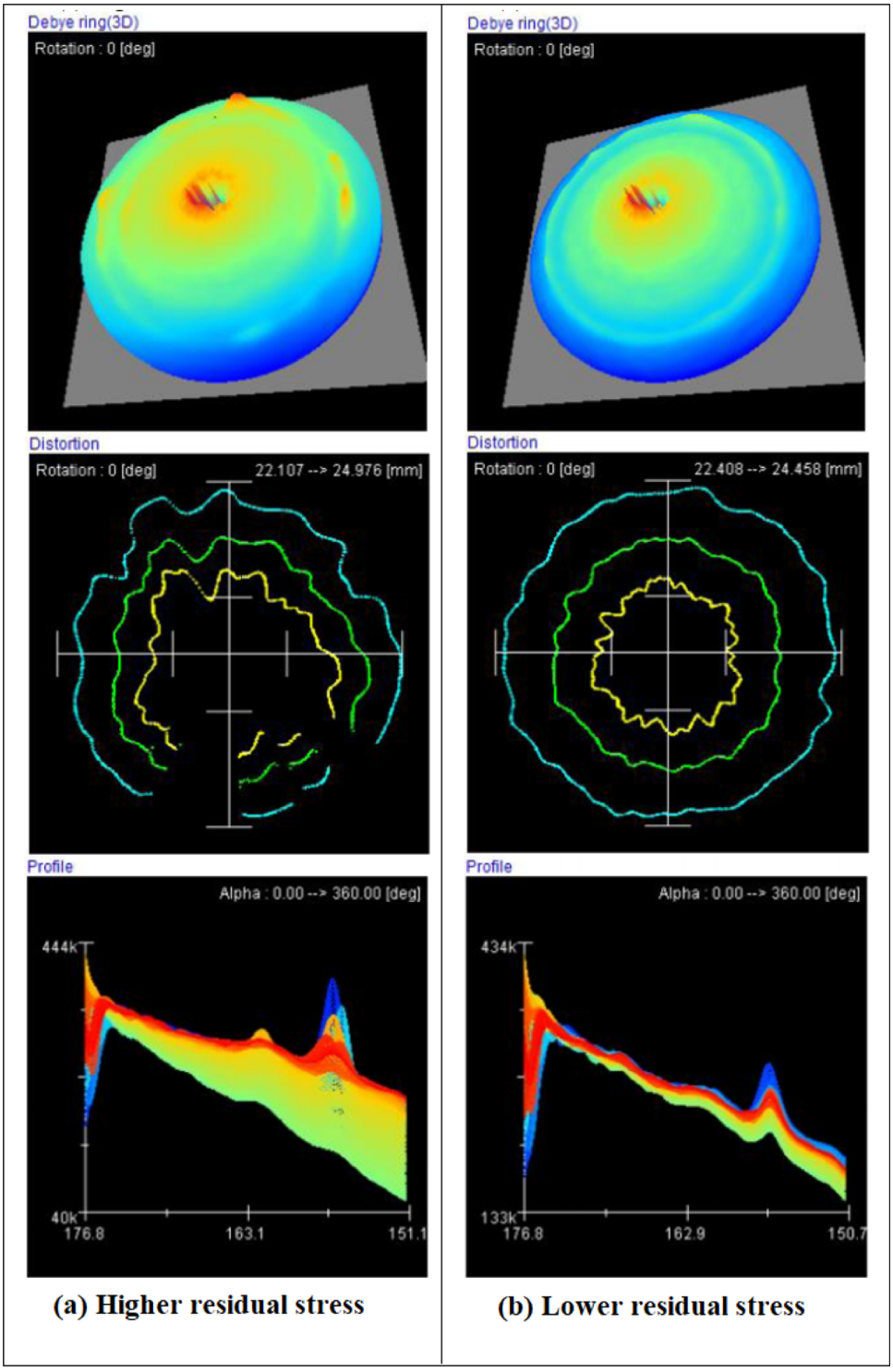


Fig. 11 Debye ring (3D), distortion and profile curve of residual stress for sample 7 (higher residual stress) and sample 3 (lower residual stress), respectively

than the required rate deteriorate the mechanical properties and the microstructure of the welded joint due to material softening.

The tensile test fractured surface FESEM images of sample 7 taken at various magnifications are shown in Fig. 8. It clearly shows dimpled morphology without any cracks, and minimum carbide precipitation is noticed. Very fine dimples (microscopic pits) with various sizes of voids are observed on the tensile broken fracture surface. Due to high stress concentration, the growth of large size micro-voids and sub-microvoids takes place in the welded region [47–49]. The ductile fracture is observed in Fig. 8c, which

Table 4 Response table for different S/N ratios

Different levels	Current (I)	Welding speed (S)	CTWD
1	43.37	44.35	42.32
2	42.08	42.52	43.13
3	42.83	41.41	42.83
Delta	1.29	2.94	0.81
Rank	2	1	3

Table 5 Confirmation test

	Initial factor setting	Predicted values	Experimental
Level	I3-S1-L3	I1-S1-L2	I1-S1-L2
UTS	723		729
%E	115		117
HV	279.4		282.4
S/N ratio	45.2125	45.3011	45.3296
Improvement in S/N ratio		0.1171	

shows that the material experiences higher elongation and enhances the weld joint strength. Selective regions from fractured surface EDX are taken at spectrum 3 and 4 points in tensile fracture surfaces, which is exhibited in Fig. 8e, f. EDX analysis shows the presence of Cr, Ni, Mn and Fe as main elements and carbide precipitation as a secondary phase. Kumar and Shahi [1] observed a similar dimpled morphology with a high degree of precipitation and flat featureless region in GTAW welding of 304L SS. The high degree of precipitation affects the grain boundaries separation, which leads to the degradation of mechanical properties.

Radiography testing is conducted to check the weld joint quality. Welding defects such as cracks, porosity, and lack of fusion are detected through radiography test as per ASTM Sec.-V:2017 standards. Figure 9 clearly shows that there is no defect observed in the weld bead, which indicates the stronger welded joint.

3.4 Residual Stress

A higher rate of solidification takes place in the upper and lower portion of the fusion region due to fast cooling of the weld material as a result of which residual stresses are produced. When unexpectedly any material fails, it occur quite often due to many reasons, but tensile residual stress also creates a possibility of unexpected failure. Residual stress results taken at various positions of the welded joints are presented in Fig. 10.

Uneven thermal distribution takes place in the welded joint due to contraction and expansion in HAZ and fusion zone. The different cooling rate in the weld zones influences the growth of residual stresses due to the change in microstructure pattern. The residual stresses are also affected by the phase transformation in the fusion zone and HAZ [10]. The larger size of the fusion area and HAZ takes place with expansion in heat input during welding. Welded sample 7 has higher residual stresses than other samples. This helps to keep away from any defects in the welded region, which are confirmed from the radiography test (Fig. 9). In the HAZ, tensile residual stress (positive) is observed, which results in fatigue failure and might be the possibility of crack commencement, which leads to

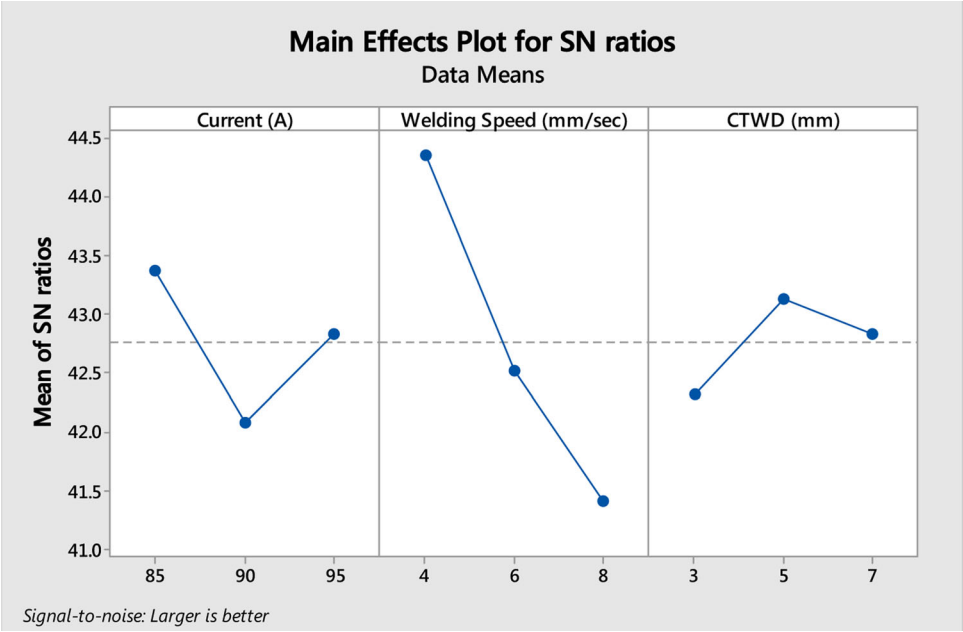


Fig. 12 Optimal parameter by Taguchi method is 85 A current, 4 mm/sec of welding speed and 5 mm of CTWD

weakening of the mechanical properties. The distortion is seen from Fig. 11a for sample 7 at weld bead position and a smaller amount of distortion is observed for sample 3 at a base metal region, which is seen from Fig. 11b. This distortion is caused due to a higher heat input, which is experienced by sample 7. Profile curve of residual stress at weld bead shows the number of peaks that are incurred in the material in comparison with the base material. The red color shade peaks indicate a higher absorption of residual stresses, whereas blue color shade peaks indicate a negligible concentration of residual stresses. This can be seen in the profile curve of Fig. 11 in both cases. The UTS of sample 7 is higher than the tensile residual stresses, including the maximum tensile residual stress. The appropriate heat input enhances the joint efficiency so that the material breaks from the base metal.

3.5 Optimization of Welding Parameters

Taguchi L9 orthogonal array design matrix with multiple response and signal-to-noise ratio is shown in Table 3. Minitab software has been used to check the optimal parameters for finding larger better responses. The response table for the S/N ratio concludes that the most dominant welding parameter is the welding speed (S) followed by the current and then CTWD, which is shown in Table 4.

The slight change in the dominant parameter (welding speed) will drastically change the responses. The optimal parameters 85 A current, 4 mm/sec of welding speed and 5 mm of CTWD have been found from this technique. The main effect plots of the S/N ratio are shown in Fig. 12.

The highest S/N ratio in the experimental data is 45.2125 for sample 7 (95A-4 mm/sec-7 mm) as shown in Table 3. The S/N ratio for the optimal parameter (85A-4 mm/sec-5 mm) is found to be 45.3296. For confirming the predicated results, three samples are welded with optimal parameters and the average value of three samples test results is presented in Table 5. The improvement in the S/N ratio is about 0.1171, which shows closer to the predicted values.

4 Conclusions

The present work investigated the mechanical properties of AISI 304 SS joints welded with ER 308L SS filler wire by the CMT welding process. The main conclusions observed from this study are listed below.

Approximately 16% and 25% decrement in the microhardness is experienced in the heat-affected zone and weld zone, respectively, in comparison with the base material. This is possible due to the softening effect of material and the formation of courser grains in the heat-affected zone.

The highest tensile strength of 723 MPa is achieved with 95 A current, 4 mm/sec welding speed and 7 mm of contact tip to workpiece distance of welded sample. Tensile tests pertain to the joint efficiency in the range of 83.43 to 98.73% for appropriate heat input welded samples.

The tensile fracture surface has dimpled morphology with lesser precipitation, which results in significant improvement in mechanical properties. The radiography test result confirms the sound welded joint.

Higher heat input with a rapid cooling rate leads to a rise in tensile residual stresses, which results in the possibility of crack commencement.

The optimal welding parameters 85 A of current, 4 mm/sec of welding speed and 5 mm of CTWD were obtained by Taguchi. The obtained optimal responses are 729 MPa of UTS, 117% of elongation and 282.4 HV of microhardness.

References

[1] Kumar S and Shahi A S, *Mater Des* 89 (2016) 399.
[2] Jha A K, Diwakar V and Sreekumar K, *Eng Failure Anal* 10 (2003) 699.
[3] Baek J H, Kim Y P, Kim W S and Kho Y T, *Int J Press Vessel and Pipe* 78 (2001) 351.
[4] Taraphdar P K, Pandey C and Mahapatra M M, *Archives Civil and Mech Eng* 20 (2020) 54.
[5] Singh J and Shahi A S, *J Manuf Process* 33 (2018) 126.
[6] Ragu Nathan S, Balasubramanian V, Malarvizhi S and Rao A G, *Def Tech* 11 (2015) 308.
[7] Huan P, Wang X, Zhang J, Hu Z, Chen W, Nagaumi H and Di H, *Mater Sci Eng A*, 790 (2020) 139713.
[8] Selvi S, Vishvaksenan A and Rajasekar E, *Def Tech* 14 (2018) 28.
[9] Cao R, Wang T, Wang C, Feng Z, Lin Q and Chen J H, *J Alloys Comp* 605 (2014) 12.
[10] Zhang H, Wang Y, Han T, Bao L, Wu Q and Gu S, *J Manuf Process* 51 (2020) 95.
[11] Koli Y, Yuvaraj N, Aravindan S and Vipin, *Trans Ind Inst Met* 73 (2020) 645.
[12] Tian Y, Shen J, Hu S, Wang Z and Gou J, *J Mater Process Technol* 259 (2018) 282.
[13] Koli Y, Yuvaraj N, Vipin and Aravindan S, *Mater Res Exp* 6 (2020) 1265e5.
[14] Jan F, Kaplan A F H and Lamas J, *Welding World* 58 (2014) 649.
[15] Kannan A R, Shanmugam N S and Vendan S A, *Inter J Adv Manuf Technol* 103 (2019) 4265.
[16] Stanciu E M, Pascu A and Gheorghiu I, In IOP Conference Series, *Mater Sci Eng* 209 (2017) 012051.
[17] Luchten-Berg P, Campos P T, Soares P, Laurindo C A H, Maranho O and Torres R D, *Surf Coat Technol* 375 (2019) 688.

[18] Zhang H T, Feng J C and He P, *Mater Sci Technol* 24 (2008) 1346.
[19] Kumar S and Shahi A S, *Mater Des* 32 (2011) 3617.
[20] Kannan A R, Rajesh N Shanmugam N S and Naveenkumar S, *Mater Today Proc*. 18 (2019) 3916.
[21] Ahsan M R, Cheepu M, Ashiri R, Kim T H, Jeong C and Park Y D, *Weld World* 61 (2017) 1275.
[22] Balbande S, Taiwade RV Patil A P and Pathak U, *Mater Res Exp* 6 (2019) 106542.
[23] Zhou J, Shen J Hu S, Zhao G and Wang Q, *Mater Res Exp* 6 (2019) 116536.
[24] Dharmik Y B and Lautre N K, *Materials Letters* 2020; 127901.
[25] Madhavan S, Kamraj M, Vijayaraghavan L and Rao K S, *Trans Ind Inst Met* 70 (2017) 1047.
[26] Akbari D and Sattari F, *Int J Press Vessel Pip* 86 (2009) 769.
[27] Yagati K P, Bathe R, Joardar J, Phaniprabhakar K V and Padmanabham, *Trans Ind Inst* Met 72 (2019) 2763.
[28] Babu S, Panigrahi S K, Janaki Ram G D, Venkitakrishnan P V and Suresh Kumar R, *J Mater Process Tech* 266 (2019) 155.
[29] Girinath B, Siva Shanmugam N and Sankaranarayanasamy K, *Trans Ind Inst Met* 72 (2019) 1551.
[30] Taraphdar P K, Kumar R, Pandey C and Mahapatra M M, *Metals Materials Int* (2021). https://doi.org/10.1007/s12540-020-00921-4.
[31] Pandey C, Mahapatra M M and Kumar P, *Arch Civil Mech Eng* 18 (2018) 1000.
[32] Díaz A, Cuesta I I, Alegre J M, Jesus A M A and Manso J M, *Thin Walled Struct* 159 (2021) 107335.
[33] Rossini N S, Dassisti M, Benyounis K Y and Olab A G, *Materials Design*, 35 (2012) 572.
[34]] Gadallah R, Tsutsumi S, Aoki Y and Fujii H, *J Manuf Processes* 64 (2021) 1223.
[35] Kumar B, Bag S, Mahadevan S, Paul C P, Das C R and Bindra K S, *CIRP J Manuf Sci Tech* 33 (2021) 158.
[36] Pandey C, Mahapatra M M, Kumar P and Saini N. *Metall Mater Trans B*, *49* (2018) 2881.
[37] Taraphdar P K, Thakare J G, Pandey C and Mahapatra M M, *Mater Letters 277* (2020) 128347.
[38] Banik S D, Kumar S, Singh P K, Bhattacharya S and Mahapatra M M, *J Mater Processing Tech* 289 (2021) 116944.
[39] Lai H H and Wu W, *J Mater Res Techol*, 9 (2020) 2717.
[40] Kumar S, Grover S and Walia R S, *Int J Inter Des Manuf* 12 (2018) 1119.
[41] Dak G and Pandey, C, *J Manuf Process* 58 (2020) 377.
[42] Sridhar P V S S, Biswas P and Mahanta P, *J Brazilian Society Mech Sci Eng* 42 (2020) 551.
[43] Pandey C, *Metall Mater Trans A*, 51A (2020) 2126.
[44] Loayza C R L, Borges D J A, Cardoso D C S, Assuncao P D C, Ferreira V O, Baia P E C, Castro A A F, Reis M A L and Braga E M, *Diamond Related Mater* 92 (2019) 138.
[45] Hsieh C C, Lai C H and Wu W, *Mater Int* 19 (2013) 835.
[46] Han L, Han T, Chen G, Wang B, Sun J and Wang Y, *Mater Characterization* 175 (2021) 111052.
[47] Pandey C, Saini N, Mahapatra M M, and Kumar P, *Eng Failure Anal* 71 (2017) 131.
[48] Pandey C, Mahapatra M M and Kumar P, *Mater Sci Eng A* 731 (2018) 249.
[49] Singh D K, Sahoo G, Basu R, Sharma V and Mohtadi-Bonab M A, *J Manuf Processes* 36 (2018) 281.

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

Wireless Personal Communications
https://doi.org/10.1007/s11277-021-08692-w

Effective Capacity Analysis Over Fisher-Snedecor *F* Fading Channels with MRC Reception

Manpreet Kaur[1] · **Rajesh Kumar Yadav**[1]

Accepted: 17 June 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract
In contrary to Shannon's ergodic capacity, the delay-constrained effective rate is used to define the maximum data rate of the real-time applications in 5G and beyond networks. The effective capacity (EC) performance of the delay sensitive communication system over Fisher-Snedecor *F* fading channel with maximum ratio combining is studied. The closed-form mathematical expressions for the EC are derived in terms of Meijer G-function and the effect of different fading parameters on the effective throughput of the system is demonstrated. The simplified asymptotic expression for high signal-to-noise ratio (SNR) and low SNR regimes are provided to gain more insight into the system. The effect of different system parameters on the EC performance is also demonstrated. Simulation results are corroborated along with the numerical results to verify the correctness of the proposed formulations.

Keywords Effective capacity · MRC reception · Fisher-Snedecor F fading · Delay constraint · Real-time applications

1 Introduction

In traditional communication systems, throughput was considered the only quality of service (QoS) parameter and ergodic capacity was used to characterize the effective data rate. Now with the emergence of real-time applications like cellular communication, device-to-device (D2D) communications, peer-topeer video streaming, visible light communication, full-duplex communications, ultra-reliable low latency communications (URLLC), reconfigurable intelligent surface (RIS) assisted communication system, video conferencing, smart grid network and online interactive gaming etc., not only the throughput but the delay sensitivity of the communication system has gain importance for the system designer. Shannon capacity is not able to characterize the effective rate of a system, when delay constraint is considered as a QoS parameter. To account for the delay performance of

✉ Manpreet Kaur
manpreettiet@gmail.com

Rajesh Kumar Yadav
rkyadav@dtu.ac.in

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

Published online: 13 July 2021

the real-time systems, the effective rate was proposed in [1]. Effective capacity (EC) can be defined as the maximum rate of data transfer that a wireless communication system can support with a guaranteed delay bound QoS constraint. The concept of effective rate has been widely applied over the last few years to investigate the trade-off among latency, reliability, energy efficiency, and security.

The fading effect of the channels result in fluctuations of the instantaneous effective rate of the system. For this reason, many researchers have studied the EC performance of various delay-sensitive applications over different fading channels. In [2], the EC model over Rayleigh fading channel is used for optimal channel allocation in cognitive radio networks (CRN), and in [3], the model is applied for the performance evaluation of voice applications over Ricean fading channel. The achievable capacity of the CRN in a delay-sensitive asymmetric fading channel is studied in [4], while a comprehensive survey of EC over the CRN for various multimedia applications is presented in [5]. Taking into consideration, the multicast receiver cross-layer, the EC analysis for the Nakagami-*m* fading channel is provided in [6]. In [7], EC model over the voice over IP (VoIP) applications for the end-to-end routing and link scheduling scheme for multi-hop wireless sensor networks (WSN) is explored. In [8], the authors have examined the attainable rate of multi-hop mobile networks with different delay-bounds over VOIP application.

Video applications are also challenging to analyse due to strict delay bound and burst flow of the networks. Regarding this, some excellent work using the concept of EC is performed for the video streaming application over the different fading channels. To enable high and low-speed video transmission the authors in [9], have provided a QoS-driven resource allocation scheme to upper-bound the overall delay for the two-hop delay-and-forward (DF) relay networks over Rayleigh fading channel. The probability distribution of a delay process with video streaming support over the Rician fading channels is investigated in [10]. EC concept is utilized in [11], to support the real-time video streaming in a heterogeneous network and the results are utilised for cooperative spectrum sensing to observe the QoS guarantee in CRNs. Similarly in [12], QoS-driven power allocation scheme in the CRN is proposed and the EC performance for different sensing scenarios in the Nakagami-*m* fading environment are studied. Due to the requirement of critical delay constraints in the wireless tele-ultrasonography medical systems, the end-to-end EC-based delay distribution is studied in [13]. In another application, EC model is used to evaluate the support level of communication network for the smart grids technologies [14].

In [15], authors have obtained EC over multiple antenna systems which indicates a significant capacity gain using multiple antennas. In [16], a general probability density application (PDF)-based approach over the multipath and composite multiple-input–single-output (MISO) fading channels have been utilized for analyzing the EC performance. Moment generating function (MGF)-based framework is proposed in [17] to study the effect of the delay QoS parameter over the generalized fading channel. Further, to get more insight into the system, the performance for low and high signal-to-noise ratio (SNR) approximations are also presented. The variation of EC performance over correlated MISO Nakagami-*m* channel has been proposed in [18]. In practical scenarios, the effect of both multipath fading and shadowing are present simultaneously. Composite fading models being the more realistic channel are used for accurate modeling of wireless channels [19]. The EC performance over the α–η–μ/Lognormal and α–κ–μ/Lognormal distribution has been studied in [20]. Recently, the EC analysis over gamma shadowed α–η–μ and α–κ–μ fading channel has been carried out in [21] and [22] using mixture gamma distribution.

In [23], the practical channel measurements for device-to-device (D2D) communication is taken and based on the empirical model it was concluded that Fisher-Snedecor

F fading channel provides the accurate modeling of the scenario. It was also shown that Fisher-Snedecor *F* distribution provide a better fit to the experimental data in comparison to the generalized-*K* channel. Recently the closed-form expressions for different power adaptation channel capacity over Fisher-Snedecor *F* fading Channel have been derived in [24]. Diversity as a method to overcome the effect of fading, the (maximum ratio combining) MRC diversity over Fisher-Snedecor *F* fading channel has been discussed in [25]. As per the author's knowledge, none of the previous works have addressed the effective rate analysis over Fisher-Snedecor *F* fading channel with MRC reception. Motivated by this, we have presented the performance analysis of the EC over the said channel by analytical determination of the closed-form expressions for the exact and asymptotic (high SNR and low SNR regimes) cases.

The remaining paper is organized as follows: Sect. 2 discusses the system and channel models in detail. Effective rate analysis over Fisher-Snedecor *F* fading channel is presented in Sect. 3. In Sect. 4, the methodology for deriving the asymptotic expressions (high power and low power) of the EC is illustrated. Section 5, shows the variation of fading parameters and their impact on the EC performance. Finally, concluding points are given in Sect. 6.

2 System and Channel Model

To encourage the effective support of QoS in next-generation wireless networks, it is essential to present a wireless channel in terms of connection-level QoS metrics. Using the physical-layer channel models, the performance evaluation of the delay-limited applications can be complicated and unreliable in some cases [1]. With the support of EC, the channel can be modeled in terms of link-layer QoS-metrics. To evaluate the performance of real-time applications of the wireless communication system under the QoS requirements such as delay, reliability, and efficiency, the EC concept is introduced in [1]. EC model can be used to find the maximum data rate that a channel transmits under the required delay constraint. So, EC can be used to analyse the performance of the system with delay bounds. The QoS guarantees heavily rely on the queuing model as shown in Fig. 1. Here, *A(t)* is the amount of the source data coming to the buffer of size Q at time t and *S(t)* represents the stochastic service provided by the channel. The length of the queue at time t is q.

Consider a real-time system of bandwidth B where data packets are divided into frames having same time duration T and with the QoS exponent θ as the delay factor. For the fading channel, the maximal arrival rate that can be supported under a statistical QoS constraint can be expressed by [22] as

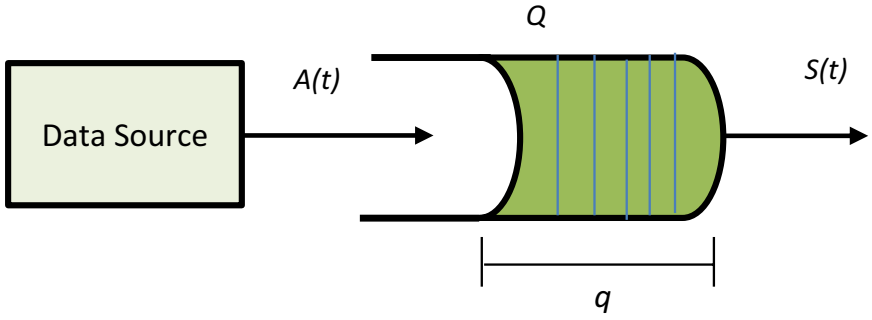


Fig. 1 Queuing system model

$$R(\theta) = -\frac{1}{A}\log_2\left[E\left\{(1+\gamma)^{-A}\right\}\right] = -\frac{1}{A}\log_2\left[\int_o^{\infty}(1+\gamma)^{-A}f_\gamma(\gamma)d\gamma\right] \tag{1}$$

where $A = \theta TB/ln(2)$, $E\{.\}$ is the expectation operator and $f_\gamma(\gamma)$ is the distribution function of the fading channel. The delay exponent is mathematically represented as

$$\theta = -\lim_{q\to\infty}\frac{\log\left[\Pr\{Q > q\}\right]}{q} \tag{2}$$

Here θ parameter specifies the maximum constant arrival rate that *S(t)* can sustain under certain QoS constraints and $\Pr\{.\}$ is the probability of occurrence of an event. It is clear from the above equations that the EC depends upon the channel parameters and system parameters and when $\underset{\theta\to 0}{EC} = Shannon's\ Ergodic\ capacity$. Fisher-Snedecor F distribution is used to model the combined effect of multipath fading and shadowing. This model is formed by the composition of Nakagami-*m* and inverse Nakagami-*m* distributions. The PDF of the received SNR of the l_{th} branch of MRC combiner over Fisher-Snedecor F distribution is given by [23] as

$$f_{\gamma_l}(\gamma) = \frac{m_l^{m_l}\left(m_{s_l}\bar{\gamma}_l\right)^{m_{s_l}}\gamma_l^{(m_l-1)}}{B\left(m_l, m_{s_l}\right)\left(m_l\gamma_l + m_{s_l}\bar{\gamma}_l\right)^{m_l+m_{s_l}}} \tag{3}$$

where $\bar{\gamma}_l$ is the average received SNR of the l_{th} branch and $B(.,.)$ denotes the beta function. m_l is the fading parameter and m_{s_l} is shadowing parameter of the l_{th} branch. The channel severity is inversely proportional to the shadowing parameter. The L branch MRC combiner co-phased, weighted and summed the individual branch SNRs and get the instantaneous SNR (γ) at the output of the combiner using

$$\gamma = \sum_{l=1}^{L}\gamma_l \tag{4}$$

With the assumptions of L independent and identically distributed (i.i.d.) branches of MRC with equal average SNR in all the branches i.e. $\bar{\gamma}_l = \bar{\gamma}$ for $l \in \{1, 2, 3, ...L\}$, the mathematical expression for the PDF of γ is given by [25] as

$$f_\gamma(\gamma) = \frac{\gamma^{(mL-1)}}{B\left(mL, m_sL\right)}\left(\frac{m}{Lm_s\bar{\gamma}}\right)^{mL}{}_2F_1\left(\left(m+m_s\right)L, mL; mL, -\frac{m}{Lm_s\bar{\gamma}}\gamma\right) \tag{5}$$

where ${}_2F_1(.,.;.;.)$ is the Gauss hypergeometric function.

3 Effective Rate

3.1 Fisher-Snedecor *F* Channel

We derive the mathematical formulation of the effective rate of the system over Fisher-Snedecor F fading channel with L branch MRC combiner. Putting (5) in (1) and using [26] the EC expression is written as

$$R(\theta) = -\frac{1}{A}\log_2\left(\begin{array}{c}\dfrac{1}{\Gamma(A)\Gamma(mL)\Gamma\left(m_sL\right)}\left(\dfrac{m}{Lm_s\bar{\gamma}}\right)^{mL}\displaystyle\int_0^{\infty}\gamma^{(mL-1)} \\ \times G_{2\,2}^{1\,2}\left[\dfrac{m}{Lm_s\bar{\gamma}}\middle|\begin{matrix}\left(1-mL-m_sL\right) & (1-mL)\\ 0 & (1-mL)\end{matrix}\right]d\gamma \\ \times G_{1\,1}^{1\,1}\left[\gamma\middle|\begin{matrix}(1-A)\\ 0\end{matrix}\right]\end{array}\right) \tag{6}$$

where $G_{p\ q}^{m\ n}[.]$ is the Meijer G-function and $\Gamma(.)$ is the gamma function. Using equation (/07.34.21.0011.01) of [27] and after some mathematical manipulations, (6) can be further simplified as

$$R(\theta) = -\frac{1}{A}\log_2\left(\frac{1}{\Gamma(A)\Gamma(mL)\Gamma\left(m_sL\right)}G_{2\,2}^{2\,2}\left[\frac{Lm_s\bar{\gamma}}{m}\middle|\begin{matrix}(1-A) & (1-mL)\\ 0 & m_sL\end{matrix}\right]\right) \tag{7}$$

3.2 Special Cases of Fisher-Snedecor *F* Channel

Nakagami-*m* and Rayleigh fading channels are special cases of Fisher-Snedecor *F* distribution. The EC of Nakagami-*m* can be obtained from (7) by putting $m_s \to \infty$ as

$$R(\theta) = -\frac{1}{A}\log_2\left(\frac{1}{\Gamma(A)\Gamma(mL)}\underset{m_s\to\infty}{Lim}\frac{1}{\Gamma\left(m_s\right)}G_{2\,2}^{2\,2}\left[\frac{Lm_s\bar{\gamma}}{m}\middle|\begin{matrix}(1-A) & (1-mL)\\ 0 & m_sL\end{matrix}\right]\right) \tag{8}$$

Using Eq. (8.2.2.12) of [28], the closed-form expression of effective rate over Nakagami-*m* channel with diversity reception can be written as

$$R_{Nak}(\theta) = -\frac{1}{A}\log_2\left(\frac{1}{\Gamma(A)\Gamma(mL)}G_{2\,1}^{1\,2}\left[\frac{\bar{\gamma}}{m}\middle|\begin{matrix}(1-A) & (1-mL)\\ 0\end{matrix}\right]\right) \tag{9}$$

Putting $m=1$ in (9) one can easily find out the EC expression for Rayleigh fading channels with MRC diversity as

$$R_{Ray}(\theta) = -\frac{1}{A}\log_2\left(\frac{1}{\Gamma(A)\Gamma(L)}G_{2\,1}^{1\,2}\left[\bar{\gamma}\middle|\begin{matrix}(1-A) & (1-L)\\ 0\end{matrix}\right]\right) \tag{10}$$

4 Asymptotic Analysis

In this section, we present a detailed asymptotic analysis of the effective throughput in both high and low SNR regimes. The aim of providing asymptotic analysis for system performance is to gain more physical insights and optimize system parameters accordingly. In addition, the asymptotic result reduces the analytical complexity while designing the system.

4.1 High SNR Approximation

High power analysis provides intuitive insight as how the system performance behaves with respect to channel fading parameters at high SNR. The expression of EC at high SNR can be written using $(1+\gamma)^{-A} \cong \gamma^{-A}$ approximation in Eq. (1) as

$$R^{\infty}(\theta) = -\frac{1}{A}\log_2(I) \tag{11}$$

where I is defined as

$$I = \int_0^{\infty} \frac{\gamma^{(mL-A-1)}}{B\left(mL, m_s L\right)} \left(\frac{m}{Lm_s\bar{\gamma}}\right)^{mL} {}_2F_1\left(\left(m+m_s\right)L, mL; mL, -\frac{m}{Lm_s\bar{\gamma}}\gamma\right) d\gamma \tag{12}$$

Using Eq. (2.21.1.1) of [28] and simplifying, the equation can be written as

$$I = \left(\frac{m}{Lm_s\bar{\gamma}}\right)^{A} \frac{\Gamma(mL-A)\Gamma\left(m_sL+A\right)}{\Gamma(mL)\Gamma\left(m_sL\right)} \tag{13}$$

Putting (13) in (11), we can get the final expression of EC. Similarly, the high power EC expressions for Nakagami-m and Rayleigh can be derived by using $\underset{a\to\infty}{Lim}\left[a^{-b}\Gamma(a+b)/\Gamma(a)\right] = 1$ in Eq. (11) and simplified as

$$R^{\infty}_{Nak}(\theta) = -\frac{1}{A}\log_2\left(\left(\frac{m}{\bar{\gamma}}\right)^{A}\frac{\Gamma(mL-A)}{\Gamma(mL)}\right) \tag{14}$$

$$R^{\infty}_{Ray}(\theta) = -\frac{1}{A}\log_2\left(\left(\frac{1}{\bar{\gamma}}\right)^{A}\frac{\Gamma(L-A)}{\Gamma(L)}\right) \tag{15}$$

4.2 Low SNR Approximation

In many communication networks, such as cellular networks, systems often operate in low-SNR situations. Hence it is beneficial to derive an approximation of (5) in the low-SNR regime. The effective rate performance for the fading channel is investigated for a low-power regime by considering that channel information is known to the receiver only where transmitter may have either no channel information or imperfect channel information. By following the second order Taylor expansion of the EC at $\rho \to 0^{+}$, as per below equation:

$$R(\rho,\theta) = R^{'}(0,\theta)\rho + R^{''}(0,\theta)\frac{\rho^2}{2} + o\left(\rho^2\right) \tag{16}$$

where $R^{'}(0,\theta)$ and $R^{''}(0,\theta)$ are the first and second-order derivatives of $R(\rho,\theta)$ with respect to average transmitting SNR, ρ at $\rho=0$. Note that in the low-SNR regime, analysis based on ρ would lead to misleading conclusions. The capacity analysis should be better conducted using the transmitted normalized energy per information bit $\left(E_b/N_0\right)$ rather than the

average SNR ρ [29]. Following the same method as in [29], the effective rate in the low-SNR regime can be approximated by

$$R\left(\frac{E_b}{N_0}\right) \approx S_0 \log_2\left(\frac{E_b/N_0}{E_b/N_{0\ \min}}\right) \tag{17}$$

where

$$\begin{aligned} S_0 &= \frac{-2\left(R'(0,\theta)\right)^2 \ln(2)}{R''(0,\theta)} \\ \left(\frac{E_b}{N_{0\ \min}}\right) &= \frac{1}{R'(0,\theta)} \end{aligned} \tag{18}$$

where $\left(E_b/N_0\right)_{\min}$ is defined as the minimum energy required to transmit a bit which is achieved as the average transmit SNR, ρ approaches to zero, whereas the wideband slope, S_0 is defined as the slope of the spectral efficiency curve which helps to study the energy efficiency at low but non-zero power levels. The first and second order derivatives, $R'(0,\theta)$ and $R''(0,\theta)$ are defined by [21] as

$$\begin{aligned} R'(0,\theta) &= \frac{E[\gamma]}{\ln(2)} \\ R''(0,\theta) &= \frac{1}{\ln(2)}\left[A(E[\gamma])^2 - (A+1)E\left[\gamma^2\right]\right] \end{aligned} \tag{19}$$

Putting (5) in the definition of $E[\gamma]$ and $E\left[\gamma^2\right]$, applying the relationship between hypergeometric function and Meijer G function and using (7.811.4) of [30] we get

$$E[\gamma] = \frac{\Gamma(mL+1)\Gamma\left(m_sL-1\right)}{\Gamma(mL)\Gamma\left(m_sL\right)}\left(\frac{m}{m_sL\bar{\gamma}}\right)^{-1} \tag{20}$$

$$E\left[\gamma^2\right] = \frac{\Gamma(mL+2)\Gamma\left(m_sL-2\right)}{\Gamma(mL)\Gamma\left(m_sL\right)}\left(\frac{m}{m_sL\bar{\gamma}}\right)^{-2} \tag{21}$$

Putting the value of $E[\gamma]$ and $E\left[\gamma^2\right]$ from (19) and (20) into (18), the expressions for $R'(0,\theta)$ and $R''(0,\theta)$ can be obtained as $R'(0,\theta) = \frac{\Gamma(mL+1)\Gamma\left(m_sL-1\right)}{\ln(2)\Gamma(mL)\Gamma\left(m_sL\right)}\left(\frac{m}{m_sL\bar{\gamma}}\right)^{-1}$ (22)

$$R''(0,\theta) = \frac{1}{\ln(2)}\left[\begin{array}{l} A\left\{\dfrac{\Gamma(mL+1)\Gamma\left(m_sL-1\right)}{\Gamma(mL)\Gamma\left(m_sL\right)}\left(\dfrac{m}{m_sL\bar{\gamma}}\right)^{-1}\right\}^2 \\ -(A+1)\left\{\dfrac{\Gamma(mL+2)\Gamma\left(m_sL-2\right)}{\Gamma(mL)\Gamma\left(m_sL\right)}\left(\dfrac{m}{m_sL\bar{\gamma}}\right)^{-2}\right\} \end{array}\right] \tag{23}$$

Now putting (22) and (23) in (18), we get the expressions of $\left(E_b/N_0\right)_{\min}$ and S_0. Further, using (18) in (17), we get the final expression of EC for low SNR regime.

5 Numerical Discussion

In this section, we have presented the analytical results to study the impact of various system parameters on the performance of the effective rate. All the computations and simulations were carried out in MATLAB (version R2014a). To validate the accuracy of the derived expressions, Monte-Carlo simulations are also included with 10 numbers of samples for generating the Fisher-Snedecor F distribution. It can be observed that the simulation results are in close match with the numerical results. The variation of effective rate versus received SNR for different diversity order and fading parameters are shown in Fig. 2. The effective rate increases with an increase in m and L values. It is noted that the effect of diversity order on the system performance is more pronounced for lower values of the fading parameters. The impact of delay constraint and average SNR on the EC as a function of diversity order is shown in Fig. 3. It is clearly observed from the plots that the effect of higher delay constraints can be compensated with the higher diversity order. It is noted that the difference between the plots of delay constraints is more at lower diversity order and the plots come closer with the further increase in diversity order.

The effective throughput versus the QoS exponent θ is plotted in Fig. 4. Overall, the results show that the performance of the system degrades with a large QoS exponent, which implies that as the delay constraints become large, the less throughput system can handle. However it is also seen that, when QoS exponent is smaller, increasing θ within the corresponding region have more impact on the effective throughput performance of the system for higher values of T; and the effective throughput performance degrades significantly with the increase inθ and T values. The asymptotic performance of the system in terms of high SNR approximation for different diversity order and shadowing parameters is depicted in Fig. 5. The asymptotic plots merge quickly with the analytical

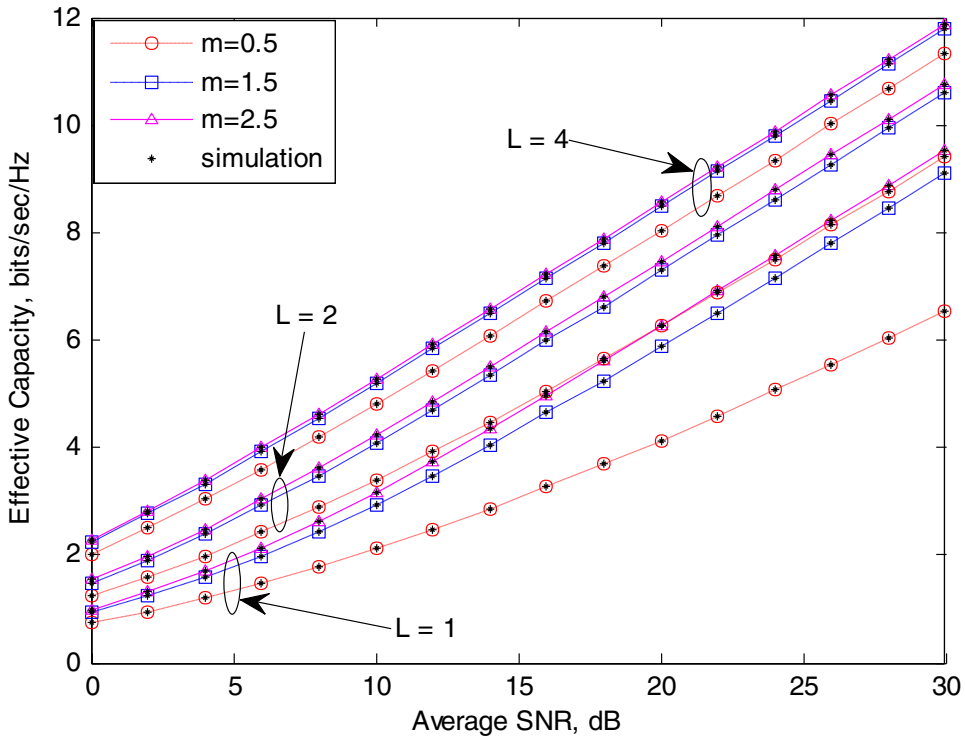


Fig. 2 Effective capacity versus average SNR for various values of fading parameters and diversity order

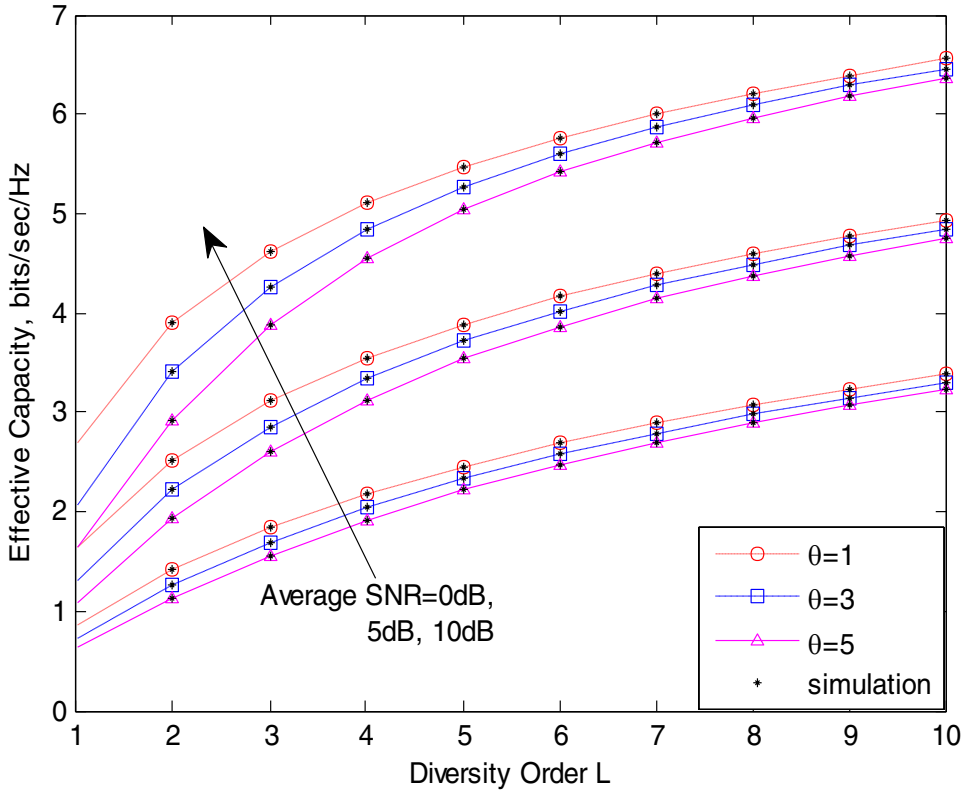


Fig. 3 Effective capacity versus L for different values of delay constraint and average SNR

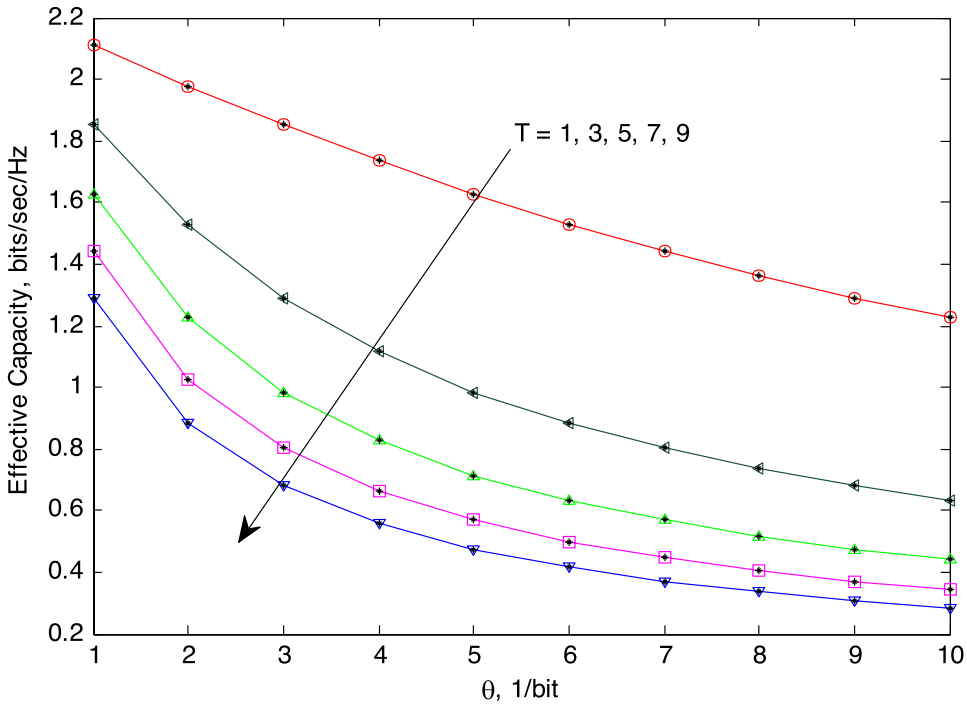


Fig. 4 Effective throughput versus the QoS exponent parameter θ

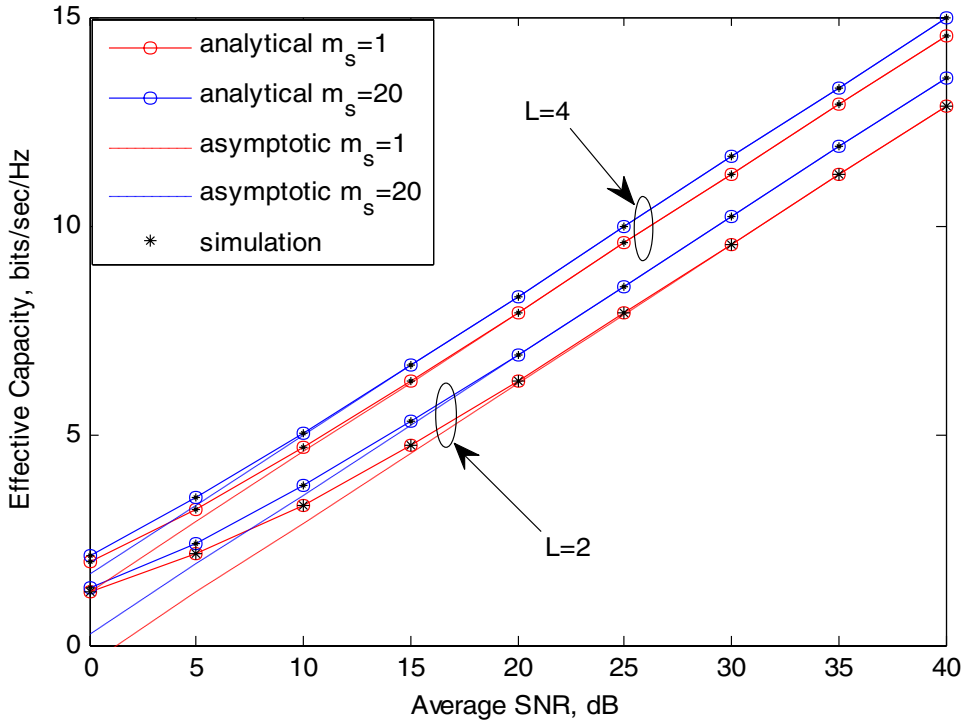


Fig. 5 Impact of diversity order on the asymptotic performance for various values of shadowing parameter

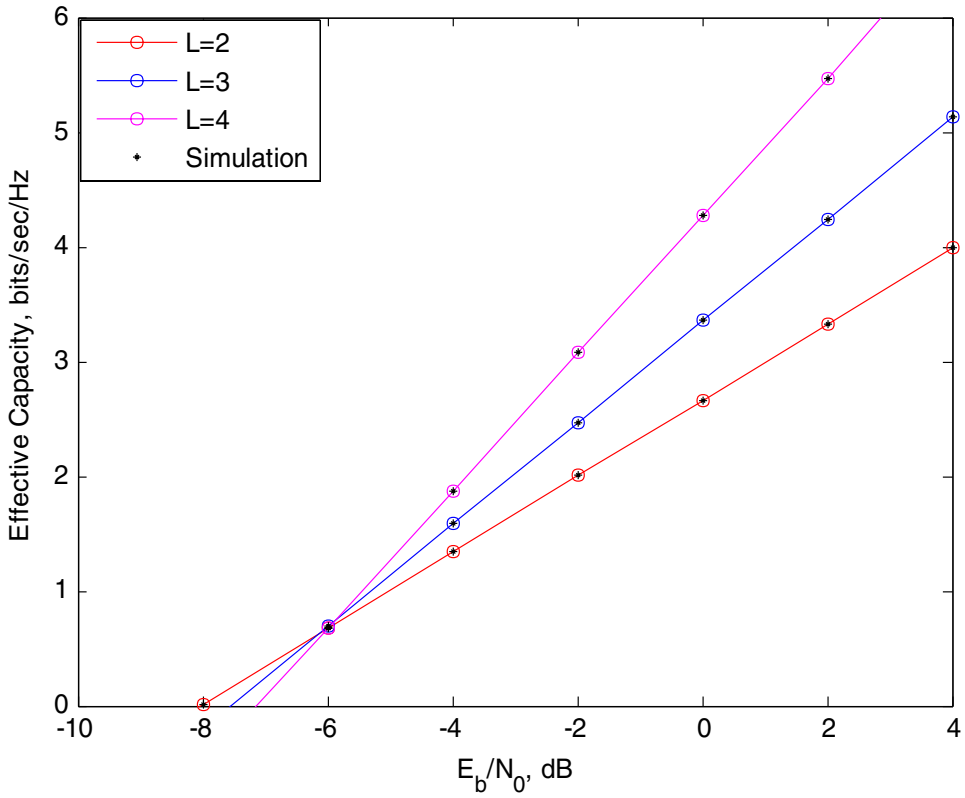


Fig. 6 Low SNR effective capacity versus average transmit SNR

plots for higher values of shadowing parameter and diversity order. In Fig. 6, the results are drawn for effective capacity in case of low SNR against E_b/N_0 for various values of diversity order and fixed values of $A = 1$. The results produced here can be directly applied for designing the communication systems for the real time applications where the channel characteristics are modeled in the form of Fisher-Snedecor F distribution.

6 Conclusions

We have derived the mathematical expression of effective rate performance of the system over Fisher-Snedecor *F* fading channel with MRC reception. The impact of system parameters on the EC performance is studied in details. These results have extended and complimented the existing research of effective rate analysis. High power and low power asymptotic expressions of the maximum data rate are provided and the importance of the approximation for MRC diversity is depicted. It is concluded that the growing demand of the high data rate with tighter delay constraints can be achieved by increasing the number of diversity branches in the system. Simulation results are to validate the accuracy of the proposed methodology and it is observed that the simulation results are in close agreement with the numerical results.

Authors' Contributions All the authors have equally contributed in this manuscript.

Funding No funding was received for this work.

Declarations

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

References

1. Wu, D., & Negi, R. (2003). Effective capacity: A wireless link model for support of quality of service. *IEEE Transaction on Wireless Communications, 2*(4), 630–643.
2. C. C. Tseng, K. C. Chen, and C. W. Su S. Y. Lien, "Cognitive radio resource management for QoS guarantees in autonomous femtocell networks," in *IEEE International conference on communications (ICC)*, 2010.
3. Z. Nikolic, N. Milosevic, and P. Pirinnen S. Glisic, "Effective capacity of advanced wireless cellular ," *IEEE 16th International symposium on personal indoor and mobile radio communications*, vol. 4, pp. 2771–2780, 2005.
4. Xiangquna, S., & Qingxin, C. (2015). Effective capacity of cognitive radio systems in asymmetric fading channels. *The Journal of China Universities of Posts and Telecommunications, 22*(3), 18–25.
5. Piran, M. J., et al. (2020). Multimedia communication over cognitive radio networks from QoS/QoE perspective: A comprehensive survey. *Journal of Network and Computer Applications*. https://doi.org/10.1016/j.jnca.2020.102759
6. Zhang, X., & Du, Q. (2007). Cross-layer modeling for QoS-driven multimedia multicast/broadcast over fading channels in advances in mobile multimedia. *IEEE Communications Magazine, 45*(8), 62–75.
7. Fan, P., Wu, D. O., K. B. Letaief, K. B., & Wang, Q. (2011). "End-to-end delay constrained routing and scheduling for wireless sensor networks," in *IEEE International conference on communications (ICC)*.
8. Chen. Y., & I. Darwazeh, I. (2011). "End-to-end delay performance analysis in IEEE 802.16j mobile multihop relay (mmr) networks," *International conference on telecommunications,* pp. 488–492 .
9. Huang, Y., Ren, P., Zhang, C., & Du, Q. (2011). "Statistical delay control and QoS-driven power allocation over two-hop wireless relay links," in *IEEE Global telecommunications conference - GLOBECOM* , Houston, TX, USA.
10. Chen, Y., & Darwazeh, I. (2013). "An estimator for delay distributions in distributions in packet-based wireless digital communication systems," in *An estimator for d IEEE wireless communications and networking conference (WCNC)*, Shanghai.

11. Zhang, X., & Wang, J. (2017). "Heterogeneous QoS-Driven Resource Allocation over MIMO-OFDMA Based 5G Cognitive Radio Networks," in *IEEE Wireless communications and networking conference (WCNC)*, San Francisco, CA.
12. Wang, J., & Zhang, X. (2017). "Statistical QoS-driven cooperative power ," in *Wireless communications and networking conference (WCNC), 2017,pp. 1–6*.
13. Chen, Y., Darwazeh, I., Philip, N., & Istepanian, R. (2013). "End-to-end delay distributions in wireless tele-ultrasonography medical systems," in *IEEE Global communications conference*, GA, USA, pp. 2592–2597.
14. You, M., Mou, X., & Sun, H. (2015). "Effective capacity analysis of smart grid communication networks," in *International workshop on computer aided modelling and design of communication links and networks*, Guildford, UK, pp. 196–200.
15. Liu, L., & Chamberland, J. (2008). "On the effective capacities of multiple-antenna Gaussian channels," in *IEEE International symposium on information theory*, Toronto.
16. Matthaiou, M., Alexandropoulos, G. C., Ngo, H. Q., & Larsson, E. G. (2012). Analytic framework for the effective rate of MISO fading channels. *IEEE Transaction on Communication, 60*(6), 1741–1751.
17. Ji, Z., Dong, C., Wang, Y., & Lu, J. (2014). "On the analysis of effective capacity over generalized fading channels," in *IEEE Int. Conf. Commu.(ICC)*, Sydney, pp. 1977–1983.
18. Guo, X. B., Dong, L., & Yang, H. (2012). Performance analysis for effective rate of correlated MISO fading channels. *Electronics Letters, 48*(24), 1564–1565.
19. Kumar, S. (2018). Energy Detection in Hoyt/Gamma Fading Channel with Micro-Diversity Reception. *Wireless Personal Communications, 101*(2), 723–734.
20. Yadav, P., Kumar, R., Kumar, S. (2020). Effective capacity analysis over generalized lognormal shadowed composite fading channels. *Internet Technology Letters, 3*(5). https://doi.org/10.1002/itl2.171
21. Yadav, P., Kumar, S., & Kumar, R. (2021). "Analysis of EC over Gamma shadowed α–η–μ fading channel," in *IOP Conference series: Material science & engineering* p. 012010.
22. Yadav, P., Kumar, S., & Kumar, R. (2020). "Effective capacity analysis over α–κ–μ/Gamma composite fading channel," in *IEEE Internatational conference on advances in computing,communication control and networking (ICACCCN)*, Greater Noida, India, pp. 587–592.
23. Yoo, S. K., et al. (2017). The Fisher-Snedecor *F* distribution: A simple and accurate composite fading model. *IEEE Communication Letters, 21*(7), 1661–1664.
24. Zhao, H., Yang, L., Salem, A. S., & Alouini, M. S. (2019). "Ergodic Capacity Under Power Adaption Over Fisher-Snedecor F Fading Channels," *IEEE communications letters*.
25. Badarneh, O. S., da Costa, D. B., Sofotasios, P. C., Muhaidat, S., & Cotton, S. L. (2018). On the sum of fisher-snedecor *F* variates and its application to maximal-ratio combining. *IEEE Wireless Communications Letters, 7*(6), 966–969.
26. Adamchik, V., & Marichev, O. (1990). "The algorithm for calculating integrals of hypergeometric type functions and its realization in REDUCE system," in *International conference on symbolic and algebraic computation*, New York (USA), pp. 212–224.
27. Wolfram. (2017). Wolfram function site. [Online]. http://functions.wolfram.com/PDF/MeijerG.pdf
28. Prudnikov, A. P., & Brychkov, Y. A. (1986). *Integrals and Series Volume 3, More Special Functions*, 1st ed. Amsterdam: Gordon and breach science publishers.
29. Matthaiou, M., & Zhong, C. (2012). Low-SNR analysis of MIMO Weibull fading channel. *IEEE Communications Letters, 16*(5), 694–697.
30. Gradshteyn, I. S., & Ryzhik, I. M. (2000). *Table of integrals, series and products* (6th ed.). Academic press.

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

Manpreet Kaur was born in Punjab, India in 1983. She received her B. Tech. in Electronics and Communication Engineering from Punjab Technical University, Punjab, India in 2005 and Master of Engineering in Electronics and Communication from Thapar University, Patiala, India in 2007. She is currently persuing Ph.D. from Delhi Technological University, Delhi, India. Presently she is working as Member (Senior Research Staff) at Central Research Laboratory, Ghaziabad, Bharat Electronics Limited. Her research interest includes the wireless communication, cognitive radios networks and 5G and beyond networks..

Rajesh Kumar Yadav is presently working as an Assistant Professor in the Department of Computer Science and Engineering, Delhi Technological University, New Delhi, India. His areas of research interest are Mobile ad-hoc Networks and Wireless Sensor Networks.

Arabian Journal for Science and Engineering
https://doi.org/10.1007/s13369-021-05923-x

RESEARCH ARTICLE-CIVIL ENGINEERING

Evaluation of Wind Induced Interference Effects on Shape Remodeled Tall Buildings

Supriya Pal[1] · **Ritu Raj**[1]

Received: 26 February 2021 / Accepted: 23 June 2021
© King Fahd University of Petroleum & Minerals 2021

Abstract
The present work focuses on the outcome of experimental investigation of wind-induced pressure on Square, Fish-plan shape-1, and Fish-plan shape-2 building models under various interference conditions. The tests are carried out in a boundary layer wind tunnel with a scale of 1:300 for 100 percent blockage between twin interfering models. The gap between the twin building models is set at 10% of the principal model's height. By proposing equations, an attempt is made to understand the distribution of the average pressure coefficient at each level along with the building height for all surfaces. The study concludes that $-1 >$ Pressure Interference Factor > 1 as experienced by Fish-shape models exhibits increased pressure and suction due to the presence of interfering building model which is unlike Square-plan shape model at present working condition. For complex plan-shaped structures at interference conditions, the study of only the Interference factor (IF) is misleading, and hence, the study of both IF and ID (Interference Difference) is suggested. In interference tests, the Fish-plan shape-1 model shows enhanced overturning moment (along-wind) performance for all interference conditions. For the Fish-plan shape-2 model, diminished overturning moment (along-wind) performance is determined for Back-to-Front and Front-to-Front interference conditions.

Keywords Tall buildings · Wind pressure · Wind force · Overturning moment · Interference factor · Interference difference · Buffeting factor

1 Introduction

Recent advances in construction methods and building materials and techniques have resulted in a new generation of larger, taller, and versatile buildings [1]. Tall buildings are more susceptible to wind excitations and are more sensitive to dynamic loading [2] for which, shape optimization of the cross section is put forward to improve its wind resistance [3]. Wind-induced lateral movement of tall structures can annoy occupants (especially on upper floors) and cause structural damage [4] and is thus regarded as one of the primary serviceability criteria when designing tall buildings [5]. This variation of the wind-induced load is due to the development of high turbulence of wind at the principal building model. Different international standards [6–10] provide a wealth of information under isolated wind incidence conditions for square, rectangular, cylindrical, and other plan shape buildings. However, due to the problem's complexity, which includes a wide range of variables in interference conditions, no information can be found in any international code. Due to interference of nearby buildings, the pressure on the principal building might increase or decrease depending upon many conditions like terrain category, the exterior shape of building for the formation of separation and reattachment regions, height of building, cross-sectional plan of the building, aspect ratio, etc. The consequences of underspecification can be serious and hence, this necessitates more experimental and analytical study to understand more realistic wind-induced conditions. Furthermore, the findings of this study can assist structural designers in selecting innovative solutions to meet the collapse and serviceability requirements of a structure under extreme wind conditions.

With the development of structural systems and building materials, the possibility of construction of unconventional shape tall buildings is very high. Wind analysis is a task of great unconventionality if the building's plan is unconventional, because of the many flow situations arising from

✉ Ritu Raj
rituraj@dtu.ac.in

[1] Deparment of Civil Engineering, Delhi Technological University, Delhi 110042, India

Published online: 02 July 2021

the interaction of the wind with the structures. A series of experimental and numerical work has been carried out to investigate wind-induced effects on various unconventional shape isolated tall building models such as; "+" [11, 12], "H"[13, 14], "Z" [15], "L" and "U" [16], "L" and "T" [17], "C" [18], "L" [19], "E" [20, 21] shape, "Triangle" shape building with various configurations [22–24], "Triangle" and "Y" [25], "Triangle", "Y" and "Circular" [26] shape buildings. In addition to the study of the above unconventional shape tall building models, investigations have also been carried out to understand the concept of Aerodynamic shape optimization [27–30]. According to the available literature, the majority of the researcher's interference study is focused on determining the optimal building configurations and/or distance between the interfering models of square or rectangular plan shape buildings [31–39]. Only a few studies on interference effects on unusual plan-shaped tall buildings are available in the literature [40–44]. Thus, this accounts for the interference study of unconventional plan shape buildings with the same or different plan shape interfering buildings.

The present study emphasizes on comparative experimental investigation between Square-plan shape (SPS) Fig. 1a, Fish-plan shape-1 (FS-1) Fig. 1b and Fish-plan shape-2 (FS-2) Fig. 1c tall building models of the same volume by considering "Shape optimization" approach under urban environment condition. The aerodynamic behavior of buildings in the urban environment is crucial as they reduce the turbulent behavior of wind in this kind of area [45]. The study's goal is to present pressure distribution variations caused by various full blockage interference conditions. The term "full blockage" refers to shielding in which the interfering model completely blocks the direct path of incident wind to the principal model. The objective is to evaluate interference factors for unconventional plan shape tall building models because these results are not incorporated in relevant codes and no analytical formula for evaluating wind effects on such complex plan-shaped tall buildings exists.

Isolated Fish-plan shape buildings are adapted triangular-plan shape buildings that are often built at triangular plots on the face of merging roads; the best example is the New York Flatiron (see Fig. 2). Fish-plan shaped buildings provide better room availability as compared to triangular-plan shape buildings. The functionality of Fish-plan shape structures can be as hotels, museums, libraries, institutional buildings,

Fig. 2 The Flatiron Building. New York City

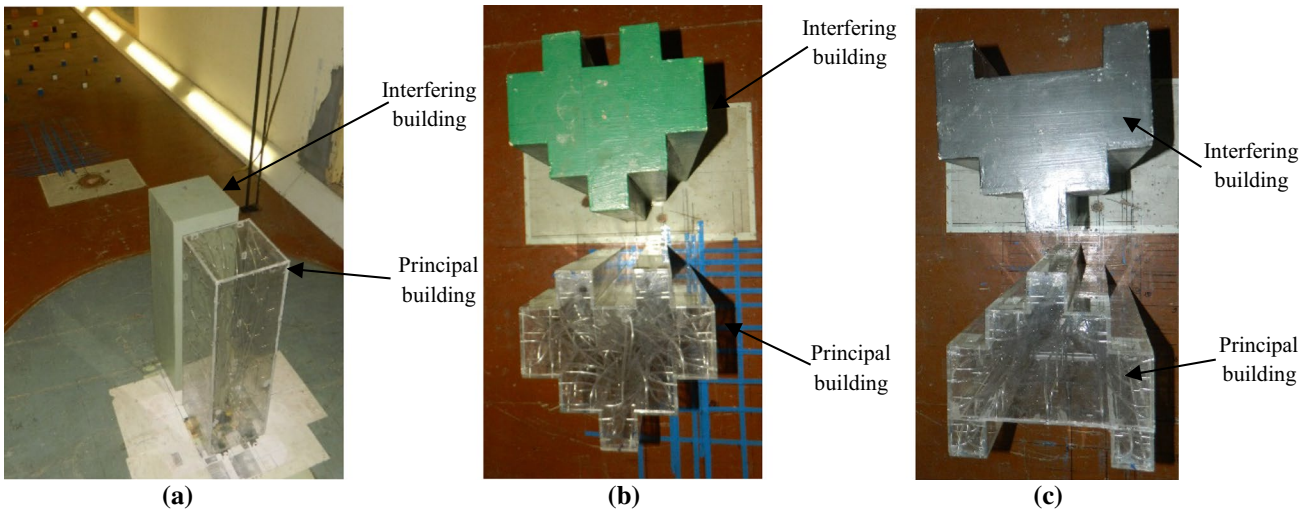


Fig. 1 Downscale model in the wind tunnel of **a** Square model at Full Blockage model; **b** Fish-plan shape-1 model at Front-to-Back and **c** Fish-plan shape-2 model at Front-to-Front interference conditions

offices, hospitals, educational spaces, and other public buildings. The current study's findings emphasize the importance of cladding surface design by determining the position of high pressure and high suction regions, as they significantly cause high external compression and high external tension on the cladding surface, respectively, which leads to cladding failure. Furthermore, the overall performance of the building is determined through investigations of base force and overturning moment.

2 Experimental Program

2.1 Feature of Experimental setup

The experiments were carried out in an open circuit boundary layer wind tunnel with a section of 2 m × 2 m and a length of 15 m. A square holed honeycomb is placed at the wind tunnel's entrance to create a uniform flow of wind throughout the tunnel. Upstream vortex generators are also being used to create boundary layer flow conditions. Figure 3 depicts a pictorial representation of wind tunnel facilities. During the experiment, the wind is continuously flowing through the tunnel owing to suction from a blower fan, which produces a constant mean wind velocity of 10 m/s. The pressure model is placed at the center of a manually controlled turntable, which rotates the model at various angles.

The pressure tapings of 1 mm diameter made up of steel tubes are installed near the edges of each face to study the changes in the variation of pressure due to flow separation. These tapings points and reference pressure points are attached to the pressure transducers for measuring pressure through the Baron instrument attached. Baratron Pressure Transducer used for the present experiment can measure 10/100/1000 cycles per second of each pressure tapping point. In our study, we have recorded the data under the Boundary Layer Wind Tunnel at a frequency of 100 cycles per second. The mean pressure is calculated from fluctuating pressure data with a time length of 170–180 s for each pressure tapping point. Wind pressure on the models is measured by Baratron Pressure Transducer, which is capable of measuring extremely low differential heads. The wind velocity inside the wind tunnel is measured with the help of the instrument "TESTO-480". A probe connected to this instrument to measure the wind velocity at different heights, which has a length of 1 m. The intensity of the turbulence is defined as the ratio of the standard deviation of fluctuating wind velocity to mean wind velocity.

The variation of the mean velocity of wind and turbulence intensity of the same is shown in Fig. 4. The boundary layer wind profile is governed by the power-law equation (Eq. 1):

$$V = V_0\left(\frac{z}{z_0}\right)^{\alpha} \tag{1}$$

where V is the mean velocity at height z above the ground, V_0 is the wind velocity at reference height, z_o is the reference

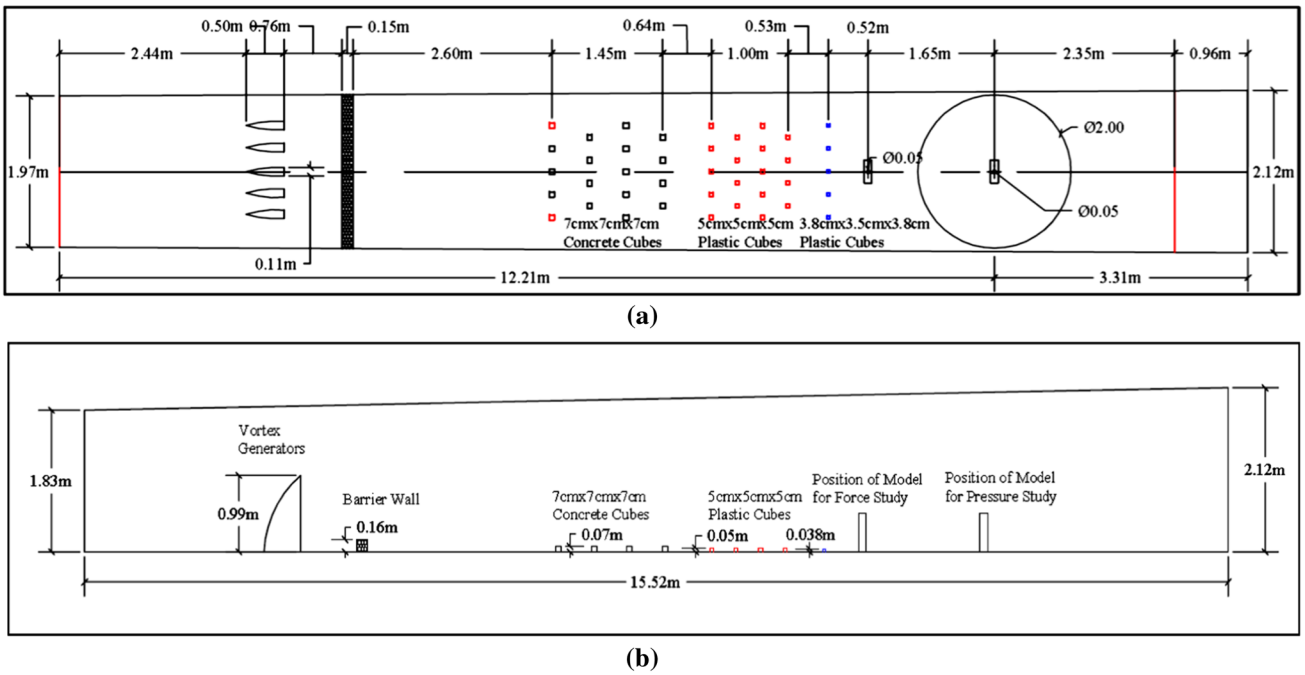


Fig. 3 Pictorial representation for the arrangement of facilities under wind tunnel at **a** Plan view and **b** Elevation view

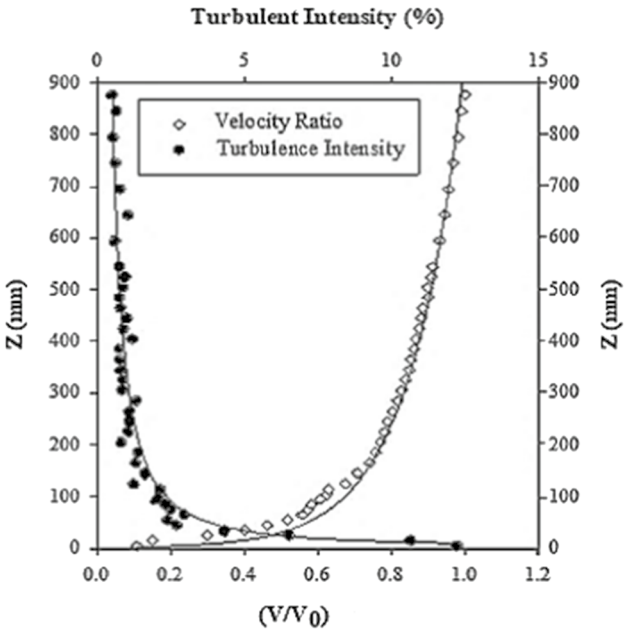


Fig. 4 Mean wind velocity and turbulence intensity profile

height above the ground, i.e., 900 mm for the present experimental work under wind tunnel, and α is the exponent of power-law, varying for different terrains.

2.2 Details of the models

Pressure models of scale 1:300 made up of transparent Perspex sheet have 5 mm thickness, height H of 600 mm and plan area of 40,000mm^2 and other dimensions accordingly. All the dimensions for plan and elevation views of the models are mentioned in Fig. 5a–c for SPS, FS-1, and FS-2 models, respectively.

The pressure tapings are shown in Fig. 5a and b with the help of black spot marks. The pressure tapings are distributed at seven different height levels at 10 mm, 60 mm, 180 mm, 300 mm, 420 mm, 540 mm, and 590 mm from the bottom to acquire a wide and clear picture of the distribution of pressure on all faces and edges of the models.

The High-frequency force balance (HFFB) technique with a five-point load cell is used for base force and overturning moment measurements. For force and moment measurements, rigid models are fixed on the disk, placed at the center of the turntable which is rotated from 0° to 180° at an interval of 15°. The disk is attached to the standard load cell that is mounted on a stable tripod on the floor of the laboratory.

2.3 Validation with international codes

For validation, an experimental study is carried out at the wind tunnel for a square plan shape isolated building model of 600 mm height and 40,000mm^2 plan area under the present working environment. The experimental study is validated with different international codes [6–10]. It is that pressure coefficients of windward, leeward, and sidewalls of the isolating model have appreciable results with the international codes as shown in Table 1.

3 Results and Discussion

For each pressure tapping point, the pressure coefficient is calculated from the formula [10] (Eq. 2):

$$C_{\mathrm{p}} = \frac{P_a}{0.6V^2} \tag{2}$$

where P_a is the pressure at the respective pressure tapping point and V is the mean wind velocity in m/s at the top of the building model i.e., 10 m/s for this experiment.

A thorough study of the distribution of average pressure coefficients (Avg C_{p}) along with the height of each face, interference factor, and interference difference at various wind interference conditions for tall building models of 1:300 scale ratio is carried out. This is a very important step to engineer an understanding of the cross-sectional plan variation of the building on associated wind pressure distribution before designing the buildings for collapse and serviceability conditions.

In interfering buildings, attachment and reattachment of the shear layer are seen at various side surfaces of the instrumented building. In interference studies, blockage in the wind tunnel is an important problem associated with wind tunnel tests. It has been reported that the critical wind direction for interference depends on the geometry and configuration of the buildings but generally, the normal wind direction to the building face will produce significant interaction effects [46]. As a result, the different configurations of twin building models with normal wind incidence directions are taken into consideration in the current studies. Due to interference of nearby buildings, the pressure on principal building (PB) might increase or decrease depending upon many conditions like terrain category, the exterior shape of the building, cross-sectional plan of the building, aspect ratio, etc. All the interference studies in the present experimental work deal with 100% blockage condition as the investigations by Mara [38] concludes that, for configurations where the interfering building (IB) is immediately upstream of the test building, the highest decrease in wind loading occurs.

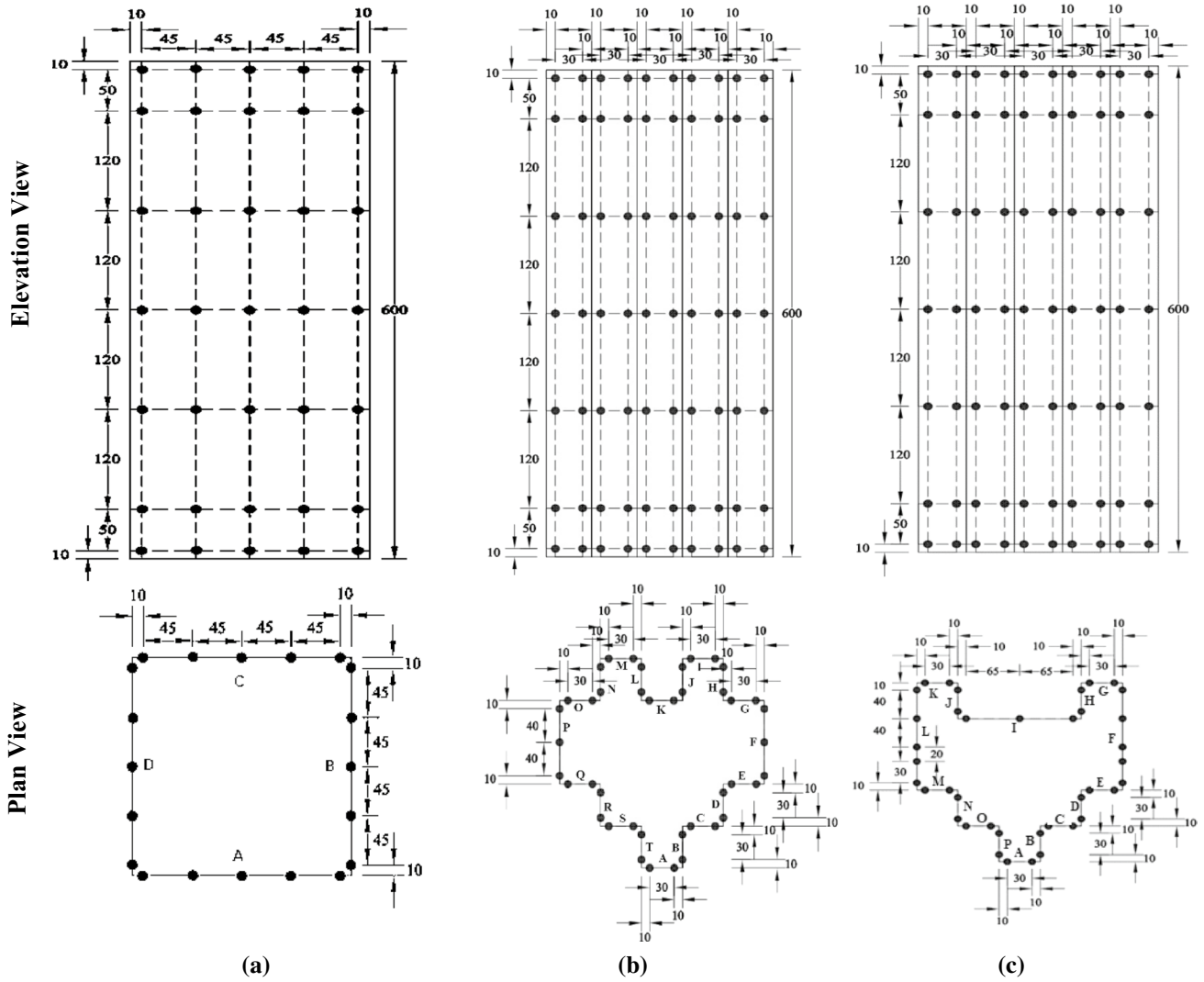


Fig. 5 Pressure taping layout at elevation and plan view of **a** Square Plan Shape Model; **b** Fish-plan shape-1 model and **c** Fish-plan shape-2 model (All units: mm)

Table 1 Comparison of face pressure coefficient (C_p) on the square plan shape tall building

International code	Wind angle	Windward side	Leeward side	Sidewalls
Experimental results	0°	0.71	−0.41	−0.67
	90°	0.73	−0.42	−0.66
AS/NZS: 1170.2:2002	0°	0.80	−0.50	−0.65
	90°	0.80	−0.50	−0.65
ASCE/SEI 7-10	0°	0.80	−0.50	−0.70
	90°	0.80	−0.50	−0.70
EN: 1991-1-4	0°	0.80	−0.55	−0.80
	90°	0.80	−0.55	−0.80
BS: 6399-2	0°	0.76	−0.50	−0.80
	90°	0.76	−0.50	−0.80
IS 875 (part 3)	0°	0.80	−0.25	−0.80
	90°	0.80	−0.25	−0.80

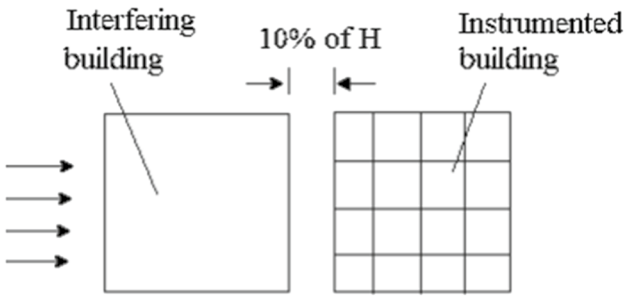


Fig. 6 Full Blockage (FBl) interference condition of Square-Plan Shape building model

The blockage in the present study is established by placing twin-building models in various orientations at a distance of 10% of the height of the model i.e., 60 mm as suggested by Cook [47], Houghton, and Carruthers [48]. The orientations of twin models for the interference study for the SPS, FS-1, and FS-2 models have been given in detail by Figs. 6, 7 and 8, respectively.

3.1 Maximum and minimum C_p at face distribution along building periphery

Suction pressure is produced as a result of flow separation at the faces. Windows panes break in suction regions, and the broken shards are dispersed outside the building. Average C_p values may suffice for structural design, but for cladding surface design, maximum and minimum C_p at the face magnitudes make a significant difference to actual conditions. The Indian Standard code [10] only deals with the average magnitudes of C_p on the face, with no information on maximum pressure or suction.

Table 2 shows maximum and minimum pressure and suction at SPS, FS-1, and FS-2 models. Maximum C_p is experienced by isolated SPS (0.97), whereas a decrement of about 28% to 23% in pressure is seen for isolated FS-1 and FS-2 models, respectively. Due to orientation of both Face-F and Face-P of FS-2 to the neighboring ones and thereby detachment of shear layer and hence, maximum suction is generated at isolated 180° wind condition. In the SPS model, a large elevated surface area of the face attracts the generation of large pressure and suction regions. Whereas, the smaller elevated surface area of the face and gradual change in the overall exterior shape of FS-1 and FS-2 thereby reduces the generation of maximum and minimum pressure and suction. In FBl condition due to the large shielding effect generated by IB, the PB is experiencing suction throughout the model. Maximum suction at FBl (−040) is reduced by about 54% as compared to isolated SPS (−0.87) due to the shielding effect of PB. Maximum suction at FF for FS-2 is reduced by 37% as compared to isolated FS-2 at 0° wind incidence. Similarly, maximum suction at BF for FS-1 is increased by 5% as compared to isolated FS-1 at 0° wind incidence. SPS, FS-1, and FS-2 models at all interference conditions do show an appreciable reduction in maximum suction from their respective isolated conditions due to the shielding effect of IB on PB. However, the reduction change in the

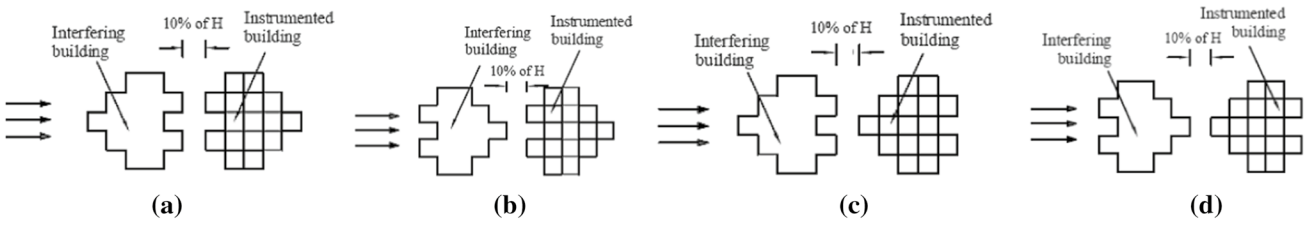


Fig. 7 Various Interference Conditions of Fish-plan shape-1 building model at **a** Back to Back (BF); **b** Front-to-Back (FB); **c** Back to Front (BF) and **d** Front-to-Front (FF)

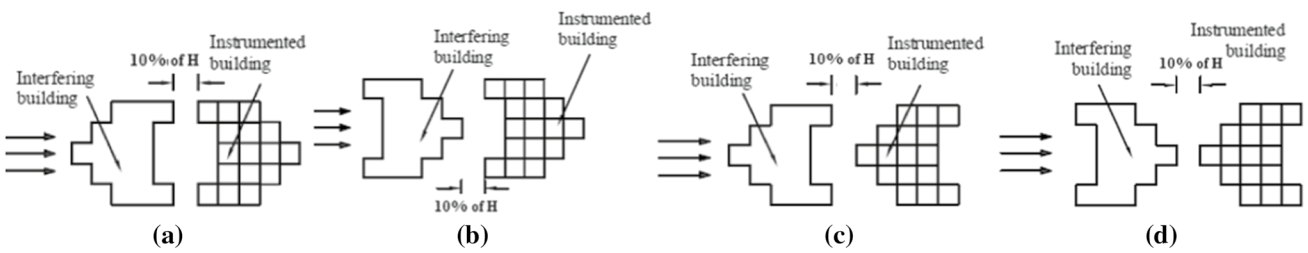


Fig. 8 Various Interference Conditions of Fish-plan shape-2 building model at **a** Back to Back (BF); **b** Front-to-Back (FB); **c** Back to Front (BF) and **d** Front-to-Front (FF)

Table 2 Maximum and minimum C_p at face distribution among various interference conditions

Interference condition	Maximum C_p	Minimum C_p
Square-Plan shape model		
Isolated (0°)	0.97	−0.87
Full blockage	−0.07	−0.40
Fish-plan shape-1 model		
Isolated (0°)	0.70	−0.60
Isolated (180°)	0.66	−0.77
Back-to-back	−0.04	−0.60
Front-to-back	0.06	−0.52
Back-to-front	0.11	−0.63
Front-to-front	0.25	−0.56
Fish-plan shape-2 model		
Isolated (0°)	0.68	−0.78
Isolated (180°)	0.75	−0.91
Back-to-back	0.06	−0.44
Front-to-back	−0.09	−0.48
Back-to-front	0.23	−0.41
Front-to-front	−0.01	−0.49

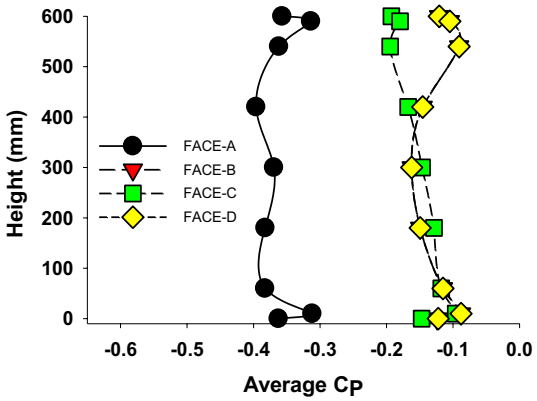


Fig. 9 Distribution of Average C_p along height for SPS for FBl interference condition

maximum reduction of suction mainly depends upon the exterior shape of the building model.

3.2 Distribution of average pressure coefficient along with the height of model for each face

Variation of Average C_p along (Avg C_p) height for all faces at various interfering wind incidence conditions is shown in Figs. 9, 10 and 11. All the vertical line plots are simplified images of complex contour plots for faces and give a full picture of variation of Average C_p along with the height of each face. All the models (SPS, FS-1, and FS-2) are symmetrical about one axis i.e., parallel to the incident wind in all the interference conditions. In Fig. 9 for FBl the opposite placed face pairs Face-B and Face-D will behave symmetrically due to the positioning of faces and hence the vertical Avg C_p distribution overlaps each other. Similarly, for FS-1(Fig. 10) face pairs such as Face-B and Face-T, Face-C and Face-S have similar Avg C_p distribution. In a likewise manner at FS-2 (Fig. 11) for all interference conditions face pairs such as Face-B and Face-P, Face-C and Face-O, etc. would have mirrored C_p distribution over the face. The comparison vertical line plots of all faces over varied interference conditions give a fine picture of the change in flow pattern along with faces among various conditions. The vortex formations and shedding mechanisms are also investigated in detail through Avg C_p distribution along with height for each face.

An attempt is been made to find out a generalized relation between α (constant), V (mean wind velocity), H_i/H (relative height), and Avg C_p along with height at each face for all interference conditions. Because of the wide variable ranges, the variations in Avg C_p along height cannot be linearly regressed, and thus the curve fitting process of all the interference parameters is complex. Based on the regression analysis, a basic equation (Eq. 3) is developed to consider average pressure coefficient distribution along with height at each face for FBl interference condition.

$$\mathrm{Avg}C_p = -a \times \left[\frac{\alpha}{\left(1 + e^{-0.30V \times \left(\frac{H_i}{H}\right)}\right)^{0.47}} \right] \tag{3}$$

where H is the height of the model and H_i varying as 0 mm, 10 mm, 60 mm, 180 mm, 300 mm, 420 mm, 540 mm, 590 mm and a is constant which can be seen varying with the faces due to change in the fluid flow pattern around building accordingly.

The current study exhibits the symmetrical flow pattern of Avg C_p distribution along with height at each observation level for all interference conditions at both models due to the symmetrical shape of the models along the direction of wind incidences. This is also evident in a study for an isolated '+' plan shape building model [11] in a similar working environment.

Due to shielding of PB and vortex generation at Face-A at FBl (Fig. 9) the vertical centerline shows large variation along with the height. Table 3 represents the magnitude of constant "a" that satisfies the distribution of Avg C_p for FBl model only. The distributions of Face-B and Face-D are the same (overlap in Fig. 9) due to the symmetrical plan shape of the model to the wind direction. It has been seen that due to interference of the twin building model the principal model experiences no pressure throughout all surfaces and maximum suction is reduced to about 54% to that at isolated condition. Fluctuation in Avg C_p is seen for all the faces

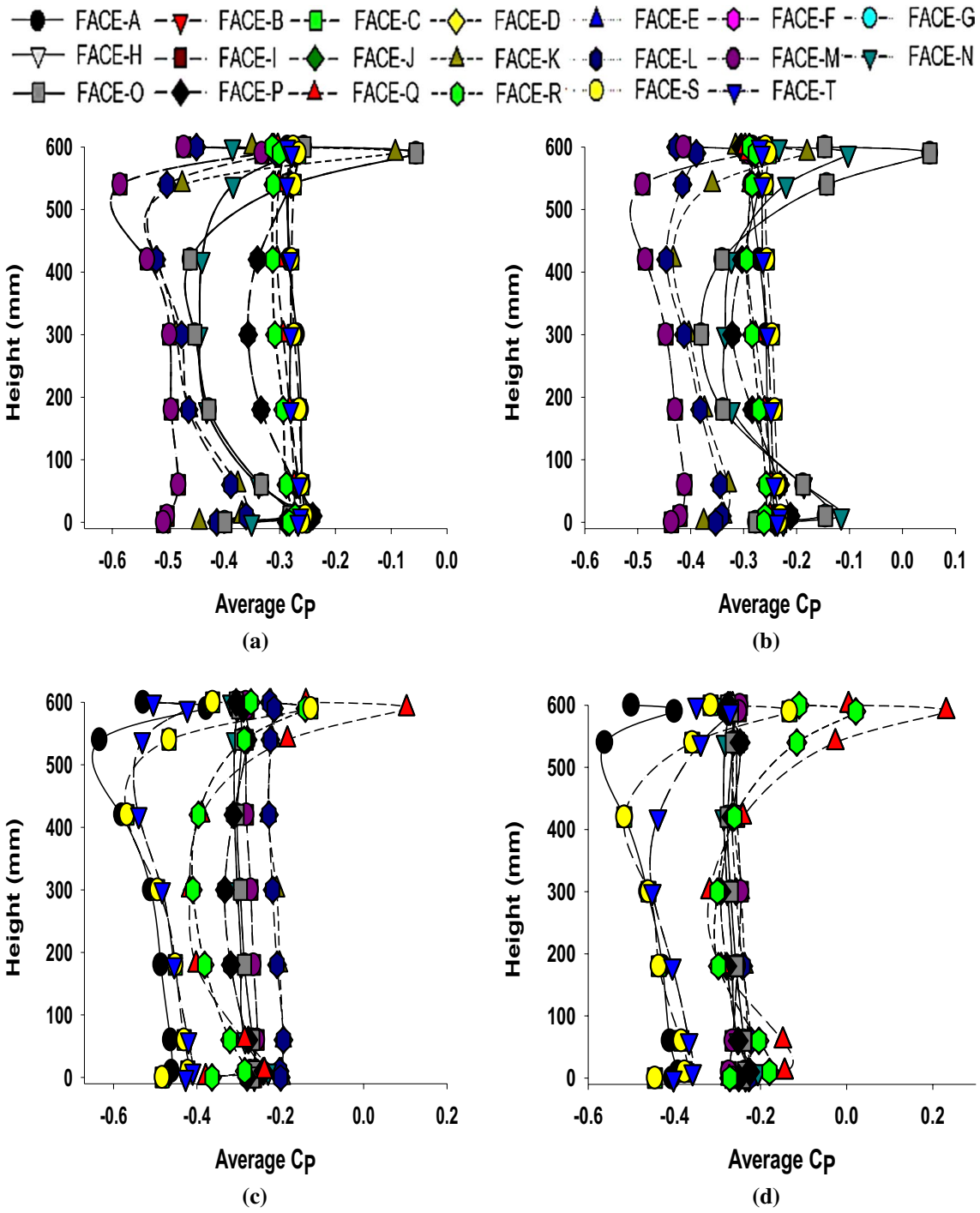


Fig. 10 Distribution of Average C_p along height for FS-1 model at **a** BB; **b** FB; **c** BF and **d** FF interference condition

at 10 mm and 590 mm. The variation at 10 mm is caused by trapped air between the ground and 10 mm. At 590 mm upwash is noticeable and hence a sudden change in Avg C_p distribution is noticeable. Face-A experiences maximum fluctuation of Avg C_p resulting because of the generation of vortices at the face due to the shielding effect of IB. Maximum fluctuation at Face-A is satisfied through Eq. 3 with "$a = 1.56$". The detachment of the shear layer is noticeable at side Face-B and Face-D, thus the least fluctuation of seen and is thus satisfied through "$a = 0.67$" at Eq. 3.

Further, through regression Eq. 4 is developed to consider average pressure coefficient distribution along with height at each face for all interference conditions at FS-1 model. All abbreviations used are the same as that used in Eq. 3.

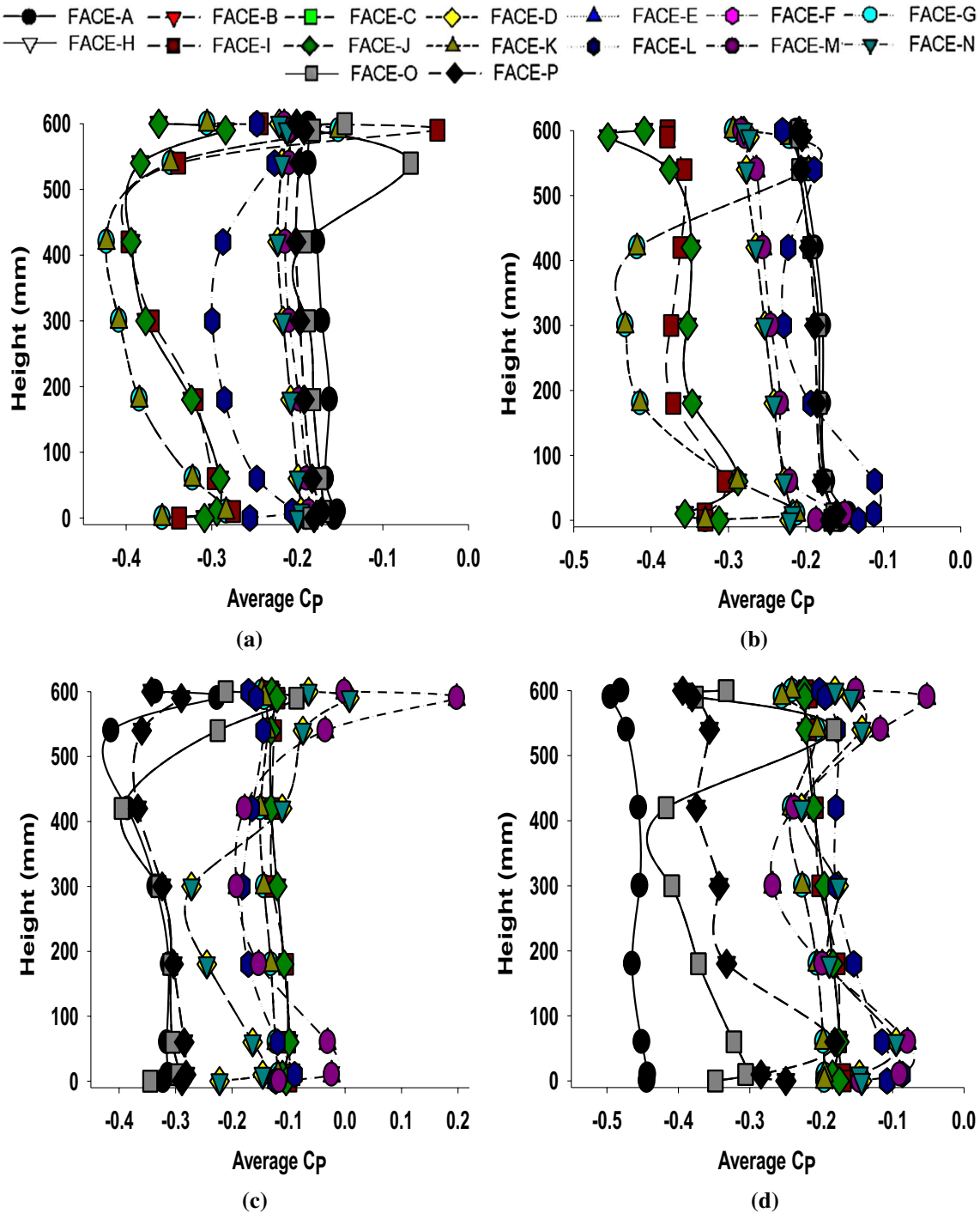


Fig. 11 Distribution of Average C_p along height for FS-2 model at **a** BB; **b** FB; **c** BF and **d** FF interference condition

Table 3 Magnitude of constant "a" for SPS model at FBl interference condition

Face	*a*
Constant "a" for SPS model at FBl interference condition:	
A	1.56
B	0.67
C	1.17
D	0.67

Table 4 Magnitude of "a" for FS-1 model at interfering conditions

Face	Back-to-Back	Front-to-Back	Back-to-Front	Front-to-Front
Constant "a" for FS-1 model at interfering conditions				
A	1.60	1.60	2.00	1.80
B	1.60	1.60	2.00	1.80
C	1.60	1.60	2.00	1.80
D	1.60	1.60	1.41	0.87
E	1.60	1.60	1.20	0.61
F	1.60	1.60	1.41	0.87
G	1.60	0.99	1.41	0.87
H	1.60	0.99	1.41	0.87
I	2.23	2.23	1.41	0.87
J	2.23	2.23	1.41	0.87
K	1.80	1.80	1.41	0.87
L	2.23	2.23	1.41	0.87
M	2.23	2.23	1.41	0.87
N	1.60	0.99	1.41	0.87
O	1.60	0.99	1.41	0.87
P	1.60	1.60	1.41	0.87
Q	1.60	1.60	1.20	0.61
R	1.60	1.60	1.41	0.87
S	1.60	1.60	2.00	1.80
T	1.60	1.60	2.00	1.80

$$\mathrm{Avg}C_{\mathrm{p}} = -a \times \left[\frac{\alpha}{\left(1 + \mathrm{e}^{-1.23V \times \left(\frac{H_i}{H}\right)}\right)^{0.20}} \right] \tag{4}$$

Although Eq. 4 satisfies all of the vertical centerline Avg C_{p} distributions of FS-1. However, due to the gradual increase and then decrease in the cross-sectional dimension of FS-1, there is a large variation in the Avg C_{p} distribution among neighboring faces, thus attempting to generalize all distributions into a single equation impossible.

At all FS-1 interference conditions, the interfering model acts as a shield for the downstream principal building model, and thus all other faces of the principal model are in the wake region and experience suction (except few pressure points). At (Fig. 10a) due to BB orientation of twin building models, the large cross-sectional size of grouped rear faces between Face-G and Face-O of PB acts as a shield to downstream faces (Face-A, Face-B to Face-H and Face-N to Face-T). The separation of wind takes place at wide angles from the edges of Face-G and Face-O and thus Avg C_{p} distribution along with height for rear faces Face-I, Face-J, Face-L and Face-M are satisfied through "$a = 1.60$" (Table 4). Due to the depressed position of Face-K the Avg C_{p} distribution is entirely different from other faces thus "$a = 1.80$" (Table 4) satisfies the distribution. The maximum distinction between experimental and regressed values is seen at Face-G and Face-O due to the fluid flow pattern around the building. FB interference (Fig. 10b) condition of Fish-plan shape twin building models distribution of Avg C_{p} along height for most faces acts similar to that at BB interference conditions and hence can be regressed with same values of "$a = 1.60$" for similar faces. However, due to the front orientation of interfering building model variation in Avg C_{p} distribution is detected at Face-G, Face-H, Face-N, and Face-O thus, "$a = 0.99$" (Table 4) is proposed. Again, the maximum variance between experimental and regressed values is seen at Face-G and Face-O due to the fluid flow pattern around the building.

At BF (Fig. 10c) and FF (Fig. 10d) interference condition of FS-1 due to orientation of PB to upstream IB model channeling effect is dominant at the interface between twin models. Due to this channeling effect, the neighboring faces are experiencing near about similar distribution. The distribution of upstream faces is governed by the point of separation of the wind stream as well as the model's cross-sectional shape, which in this condition gradually increases and then decreases, unlike regular square-plan shape buildings. Due to the large fluctuation of Avg C_{p} at all the surfaces generalizing the equation for vertical centreline profile is difficult and thus values of "a" in Table 4 are generated to satisfy the vertical Avg C_{p} distributions of said faces.

Further, through regression Eq. 5 is developed to consider Avg C_{p} distribution along with height at each face for all interference conditions at FS-2 model.

$$\mathrm{Avg}C_{\mathrm{p}} = -a \times \left[\frac{\alpha}{\left(1 + \mathrm{e}^{-0.31V \times \left(\frac{H_i}{H}\right)}\right)^{0.12}} \right] \tag{5}$$

Turbulence at the face is represented by the fluctuation of Avg C_{p} along height. Whereas, deviation to the negative side represents the formation of a vortex and deviation to the positive side represents the formation of a pressure region. The position and orientation of the faces, as well as the principal building model to incident wind, influence the

Table 5 Magnitude of "*a*" for FS-2 model at interfering conditions

Face	Back-to-Back	Front-to-Back	Back-to-Front	Front-to-Front
Constant "a" for FS-2 model at interfering conditions				
A	0.82	0.82	1.30	1.80
B	0.82	0.82	1.30	1.80
C	0.82	0.82	1.30	1.80
D	0.82	0.82	0.67	0.67
E	0.82	0.82	0.27	0.67
F	1.40	0.82	0.67	0.67
G	1.40	1.40	0.67	0.67
H	1.40	1.40	0.67	0.67
I	1.40	1.40	0.67	0.67
J	1.40	1.40	0.67	0.67
K	1.40	1.40	0.67	0.67
L	1.40	0.82	0.67	0.67
M	0.82	0.82	0.27	0.67
N	0.82	0.82	0.67	0.67
O	0.82	0.82	1.30	1.80
P	0.82	0.82	1.30	1.80

position and formation of turbulence, which is characterized by a high fluctuating Avg C_p distribution.

In Fig. 11a for BB interference condition at FS-2, generation of a vortex with the center of smaller suction of wind is noticeable between 550 and 500 mm height at Face-C and Face-O. All parallel faces to the direction of wind incidence experience lesser fluctuation of wind along with height due to detachment of shear layer at the faces. However, Face-J and Face-H are experiencing fluctuation along with height due to the depressed position of faces. Distribution of Avg C_p at Face-F and Face-L display side wash due to orientation of face but with lesser fluctuation along with the height. Detachment of shear layer is evident from Face-F to Face-L due to positioning of faces at the interface between twin models. Table 5 displays the magnitude of constant "*a*" for all interference conditions of FS-2, which satisfies the distribution of Avg C_p along with height through Eq. 5. Face-G and Face-K show the large fluctuation of Avg C_p along with height as these faces are directly exposed to the shielding effect of IB.

In FB interference condition at FS-2 (Fig. 11b) due to orientation of PB model Face-G and Face-K acts as a shield to downstream faces. Large fluctuations of Avg C_p for Face-G and Face-K are dominated by the position and orientation of the face. In BF interference condition at FS-2 (Fig. 11c) and FF interference condition at FS-2 (Fig. 11d) due to orientation of PB channeling effect is dominating front faces that are positioned at the interface between the models. Thus, Face-A to Face-C and Face-P to Face-O is experiencing large fluctuation of Avg C_p and nearby similar distributions between BF and FF interference conditions.

The distribution of Avg C_p at BF and FF interference conditions is also governed by the orientation of IB thus varied distribution is shown for the front face as described by constant "a" at Table 5.

3.3 Interference factor

The interference effect is an experience caused by the presence of one or more objects in the path fluid flow that obstructs it from reaching the principal or test object. The interference factor represents the severity of the presence of the interfering building on the principal building. In this paper, the average interference factor (Avg IF_p) for Avg C_p is calculated out for each height level and is represented by a distribution along the building's periphery. With the help of this factor, the face pressure variation on the instrumented tall building is analyzed. Equation 6 [34] gives interference factor for selected points:

$$\text{Pressure Interference Factor (IFp)} = \frac{C_{pi} \text{ with interfering building}}{C_{pi} \text{ without interfering building}} \quad (6)$$

At the current working conditions for SPS (Fig. 12), due to a 10% H difference between twin building models, IF lies between − 1 and 1 (Fig. 13), indicating improved performance of the principal model due to interference effect with both pressure and suction decreases as indicated by these magnitudes. At a gap of 25% of the height of principal model between tandem arranged twin Square shape models IF reaches up to − 2.6 which shows increased suction due to turbulence as a result of the presence of interfering twin building model [35]. The maximum Avg IFp experienced by side Face-B and Face-D at SPS model is 0.40 (Fig. 12), whereas the maximum IFp for side face at parallel arrangement of square models [35] at $y/b = -0.18$ is 1.8 and the maximum IFp for side face at Zu and Lam [34]

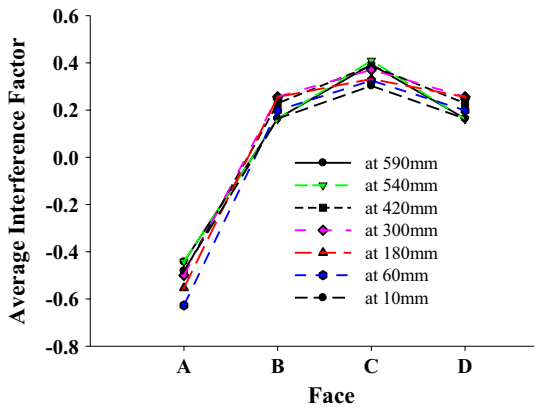


Fig. 12 Distribution of average IF_p along periphery of SPS model at FBl interference condition

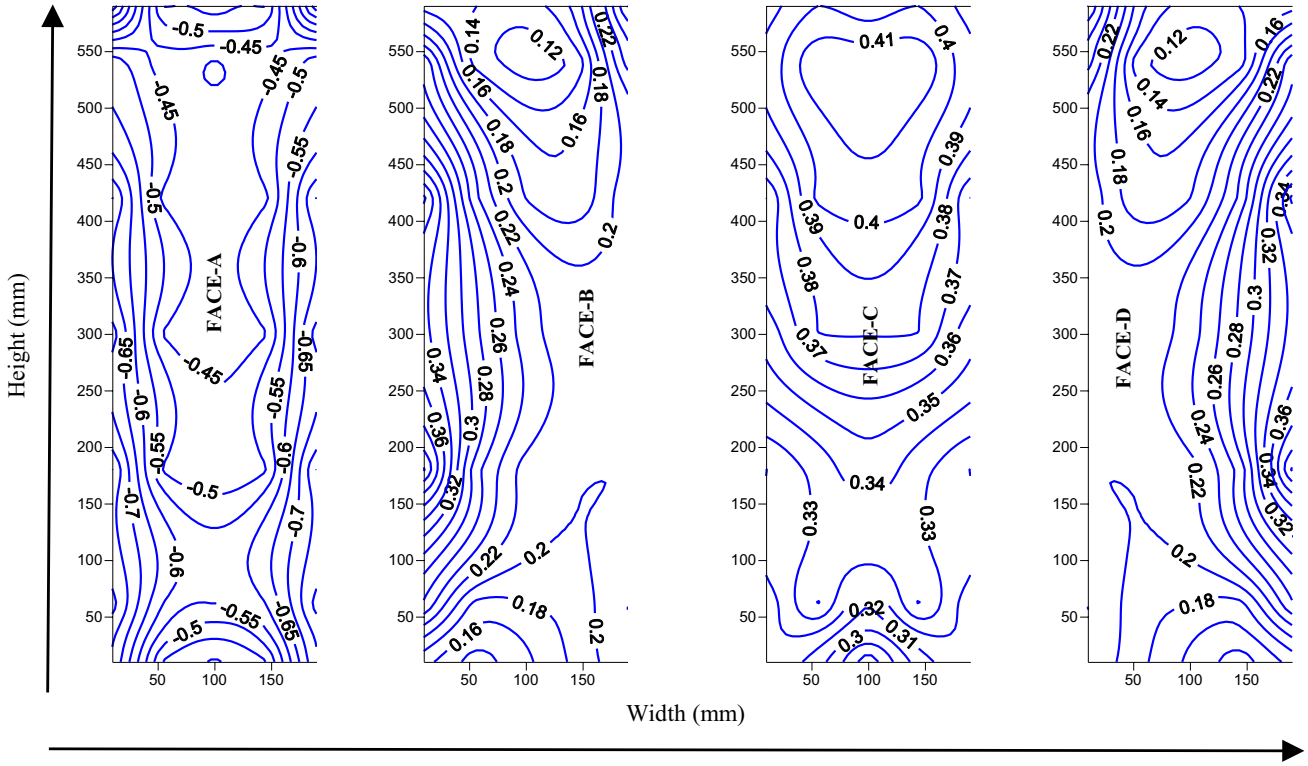


Fig. 13 Contours of IF_p for principal building model at SPS at FBl interference condition

investigations is 2.0. The difference in the Avg IFp magnitudes between the two studies is primarily due to the difference in the percentage gap between PB and IB, as well as the H/b ratio of PB.

Figure 14a–d shows very high turbulence at some faces of all interference conditions at FS-1 PB due to the model's unconventional cross-sectional plan shape with gradual increase and then decrease in cross-sectional plan magnitude, which is unlike the Square-plan shape building model and any other reported interference studies. Figure 15 shows IF_p contour plots with $-1 < IF$ and $IF > 1$ as these values show the increase in turbulence at the principal building model. Maximum IF at BB condition (Fig. 14a) lies at Face-G and Face-O at H/2 and H/3 of magnitude − 15 and 15, respectively. High variation in Avg IF_p of Face-G as is evident from Fig. 15a plot, caused due to depressed position of the face. Face-H to Face-N (Fig. 14a) is found to have a maximum average IF (− 5.88) at 0.9H, with $IF > 1$ is seen to be distributed from H/2 to H. Due to the depressed positioning of Face-J and Face-L, $IF > 1$ is experienced at the face at top H/10.

At FB interference (Fig. 14b) condition maximum influence of interfering building with Avg $IF_p > 1$ is found to be uniformly distributed from H/10 to 9H/10 at Face-H and Face-N (Fig. 15b) due to position of faces at the interface between twin interfering building model and thus influenced by channeling effect at the position. At BF Fig. 14c and FF Fig. 14d interference conditions the maximum influence of interfering building model is seen only at Face-B and Face-P at 590 mm. However, from the detailed contour distribution of Avg IF_p at Fig. 15c and d the influence of high Avg IF_p is visible only at the top corner.

Kim and Kanda through their interference study [49] found that the highest suctions increase with the increase in height of the interfering buildings. Hui et al. [50] through their interference study have cautioned for detailed study at the edges and corners as these parts are concentrated with huge pressure variations.

In Fig. 16a Avg $IF_p > -1$ is experienced at Face-G, Face-H, Face-J, and Face-K only. Back orientation of PB is causing the development of large vortices at these faces due to their positions. Maximum Avg IF_p of about − 2 is seen at Face-G and Face-K at 0.1H and 0.7H. At Face-H and Face-J $IF > -1$ is seen at 590 mm only which is distributed at a very small portion of the face (refer to Fig. 17a). Figure 16b for FB interference condition Avg $IF > -1$ is visible at rear Face-G, Face-I, and Face-K, as these faces lie at the junction between twin models. At all these faces, high Avg IF_p is distributed at the top 10 mm only (refer to Fig. 17b).

Due to the orientation of PB at BF (Fig. 16c) and FF (Fig. 16d) huge channeling effect is seen and thus front Face-A to Face-D and Face-P to Face-N experiences Avg

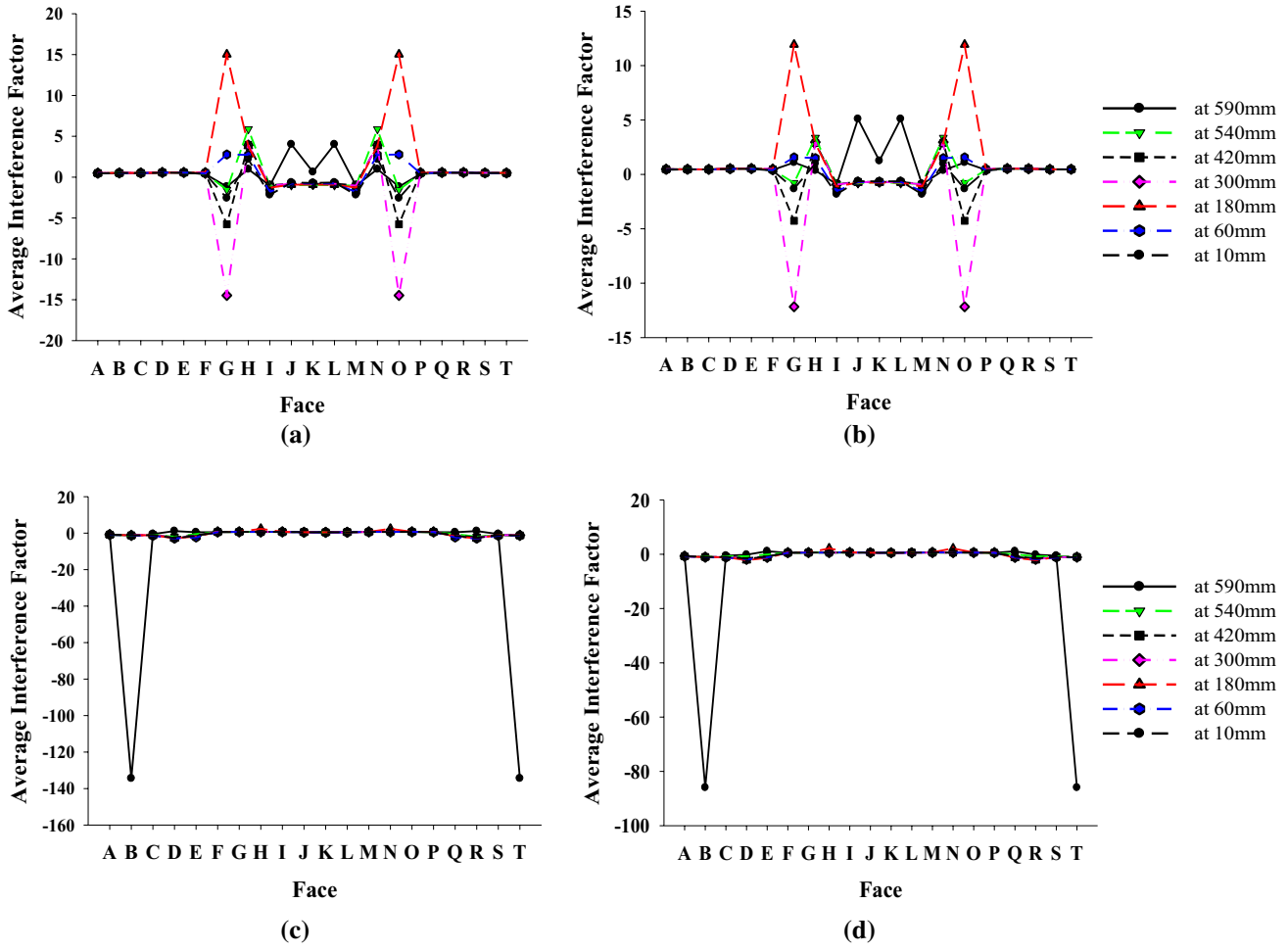


Fig. 14 Distribution of average IF_p along periphery of FS-1 at **a** BB; **b** FB; **c** BF; **d** FF interference condition

$IF_p > \pm 1$ at several height levels. For these faces $IF > \pm 1$ is distributed throughout the height (refer to Fig. 17c, d) and is concluded to be under the maximum influence of IB. Similar behavior is also seen for FS-1 for BF and FF interference conditions.

The high magnitude of IF_p as seen for that at FS-1 and FS-2 at present working condition is not reported in any other past interference studies. The difference found in the IF_p for SPS, FS-1, and FS-2 for similar blockage ratios is highly influenced by the exterior shape of the building. It concludes that the complex pan shape of FS-1 and FS-2 building model causes a high influence of interference at certain pressure points.

For complex plan-shaped structures IF_p likely not all give proper results when in some isolated conditions the C_p is near to zero. Detailed study of such critical pressure points is needed and hence, another factor is studied to calculate out interference-effect.

3.4 Interference Difference

The effect of aerodynamic interference is usually exhibited in the form of the interference factor as mentioned in the above section. The influence of IB on PB can also be studied in the form of interference difference, which is defined as [51]:

$$\text{Interference Difference (ID)} = C_{\text{pi}} \text{ with IB} - C_{\text{pi}} \text{ without IB} \quad (7)$$

ID assists in determining the position of a significant increase in pressure and suction due to IB on PB. The interference effect on PB is sometimes favorable such as the shielding effect and causes a reduction in pressure and suction. ID = 0, means there is no change in Avg C_p due to the interference effect. The average interference difference (Avg ID) for Avg C_p is calculated for each height level in

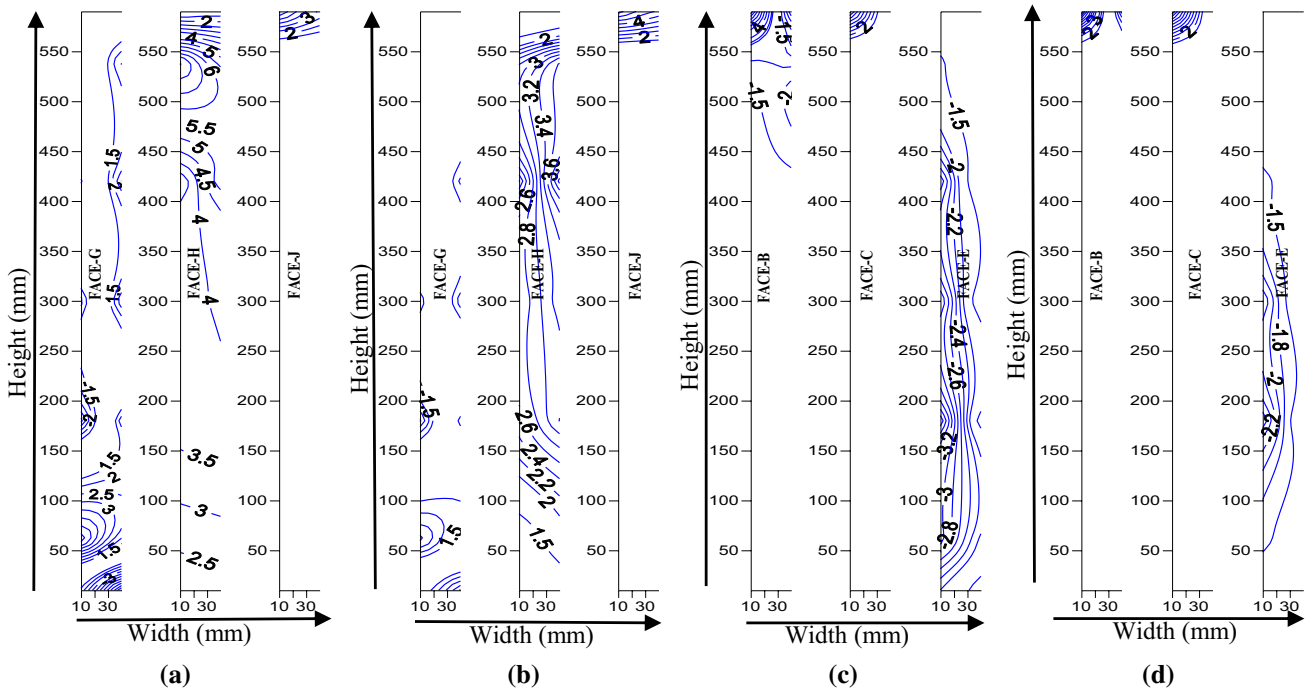


Fig. 15 Contours of IF_p at principal building's facades (for selective faces) at FS-1 for **a** BB; **b** FB; **c** BF; **d** FF interference condition

this paper and shown by the distribution along the building periphery.

Although IF_p is very high at several taping points under certain conditions, ID does not show a significant difference in Avg C_p and thus an appreciable interference effect of IB can be ignored.

The Avg C_p distribution along with height for most of the interfering conditions are negative which is visible from Figs. 9, 10 and 11. For Avg $C_p < -1$ at any interference condition and ID > 0 means Avg C_p condition at respective isolated condition is greater than that at interference condition and ID < 0 explains vice versa results.

For positive values of Avg C_p at interference condition, ID > 0 means the positive wind pressure in interference condition is higher than that at isolated condition or wind action at isolated condition is already under the influence of suction. In this analysis, the second situation is seen for some faces at several interference conditions. For positive values of Avg C_p at interference condition, ID < 0 can be concluded that the positive wind pressure is reduced in interference condition.

For negative values of Avg C_p at interference condition, ID > 0 means that the negative wind pressure in interference condition is smaller than that in isolated condition and can be concluded with reduced suction at PB due to interference condition. For negative values of Avg C_p at interference condition, for ID < 0 one can conclude that a negative wind pressure at interference condition is a higher value than in isolated condition and also wind action has changed from positive C_p at the isolated condition to negative pressure at interference condition.

As seen in Fig. 18 the minimum Avg ID for Face-A lies between − 1 and − 1.2 for height levels of 0.3H, 0.5H, and 0.7H, although the respective Avg IF_p is shown a magnitude less than − 1. From these two factors, it can be concluded that the nature of pressure has changed between isolated and interference studies and due to which suction at FBl is increased to significant magnitude.

At BB and FB interference conditions, the appreciable magnitude of Avg ID i.e., more than 1 is exhibited at Face-I to Face-M (Fig. 19a, b) for 0.5H and 0.9H, however at these taping points Avg IF_p does not show large magnitudes. At certain conditions at several taping points although Avg IF_p lies between − 1 and 1, whereas Avg ID shows a significant difference in Avg C_p and thus appreciable interference effect of IB should be studied in detail, as the magnitude of only IF_p or ID might be misleading for some conditions. Face-A exhibits Avg ID > − 1 under BF and FF interference conditions. However, under both of these conditions, Avg IF_p is less than − 1. Figure 14c and d shows that the highest influence of interference, as represented by Avg IF_p magnitudes, is found at Face-B of BF and FF interference conditions. Very high magnitudes of about − 135 and − 85 at 590 mm are seen for IF_p at Face-B. However, for these respective pressure points, Avg ID shows a very less magnitude of about − 0.42 and − 0.28,

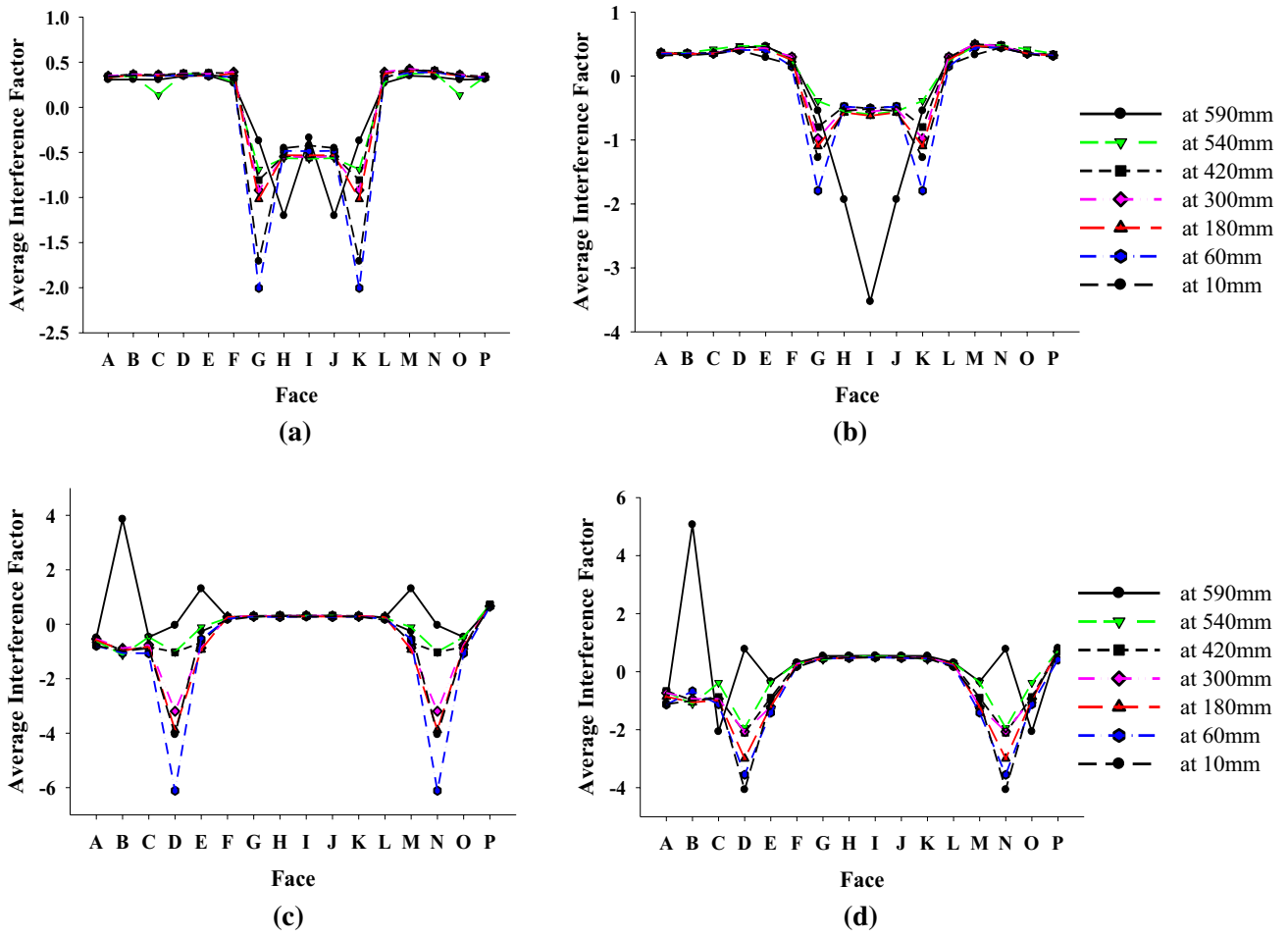


Fig. 16 Distribution of average IF_p along periphery of FS-2 at **a** BB; **b** FB interference condition, **c** BF; **d** FF interference condition

respectively. This high discrepancy between both the interference studies is due to the method of calculation.

In BB and FB interference conditions Avg ID > − 1 is exhibited by Face-H, Face-I, and Face-J (Fig. 20a, b). Due to the shielding effect of IB the faces at the interface between twin models are experiencing a maximum change in Avg C_p magnitudes as exhibited by these faces. At BF and FF interference conditions Face-A lies at the adjoining region between twin models. Due to the orientation of PB model Face-A experiences a channeling effect and hence large turbulence leading to maximum difference in Avg C_p between interference and respective isolated condition. Avg IF_p at Face-D and Face-N at BF and FF (Fig. 16a, c) is about − 6 for 0.1H, however, at this pressure point, Avg ID is − 0.19, hence the severity of interference can be neglected.

From past studies, it is clear that more priority for interference studies has been given to IF_p as compared to ID. But, from the present study, it can be concluded that in several conditions, merely the study of IF_p magnitudes can be misleading, and hence the study of both IF_p and ID is suggested when interference study of complex plan shape building models is concerned.

3.5 Base Shear and Base moment

3.5.1 Force Coefficients

Force coefficients are the representation of forces and calculated with the help of Eqs. 8 and 9 as incorporated in the previous studies [52, 53]:

$$C_D = \frac{F_D}{0.6BHV^2} \tag{8}$$

$$C_L = \frac{F_L}{0.6BHV^2} \tag{9}$$

It has been widely recognized that external shapes of tall buildings play an important role in the generation of wind loads on high-rise structures [47, 54, 55]. Also, it was relayed by Sakamoto and Haniu [56] and Song et al. [57]

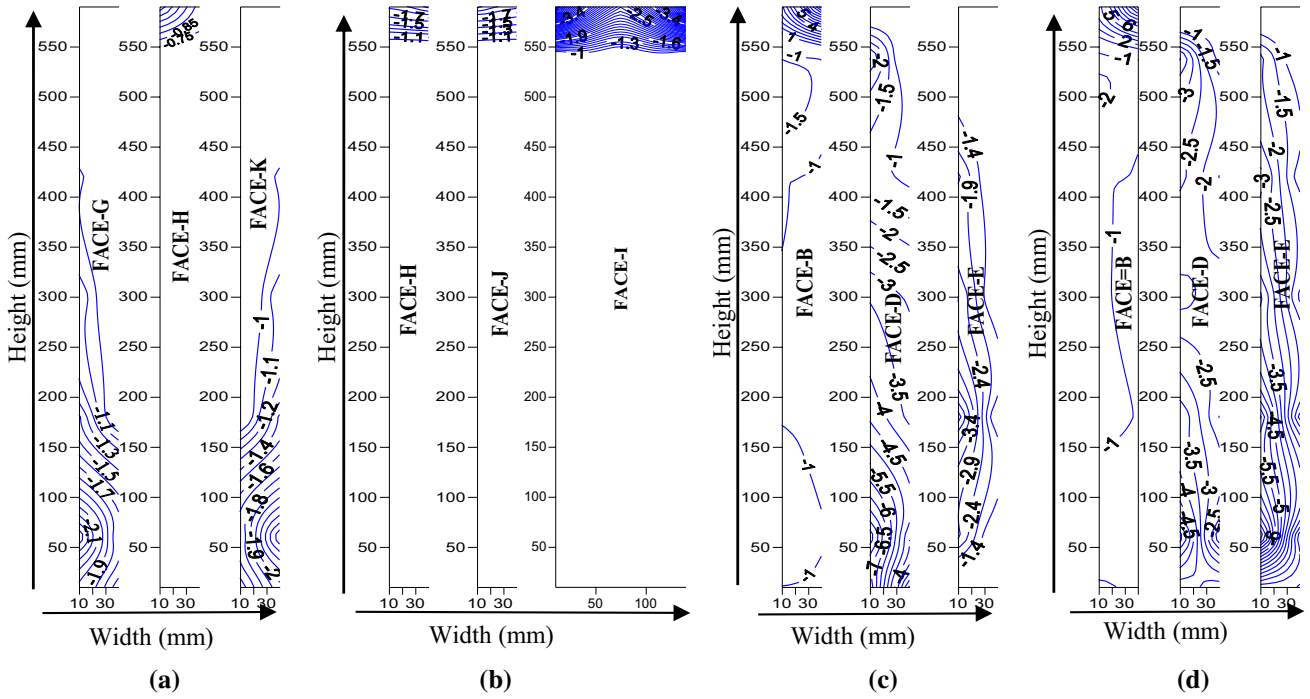


Fig. 17 Contours of IF_p at principal building's facades (for selective faces) at FS-2 for **a** BB; **b** FB; **c** BF; **d** FF interference condition

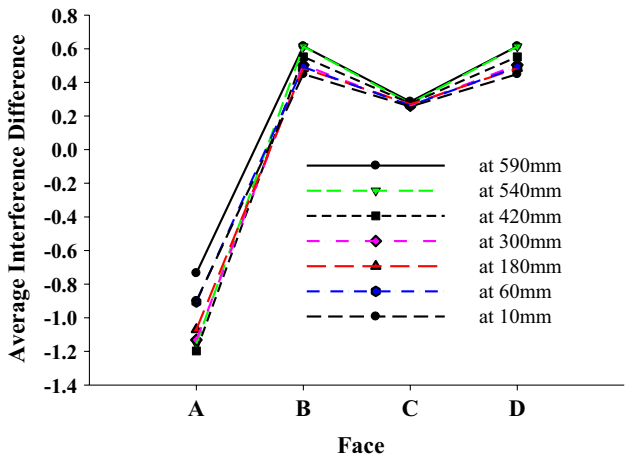


Fig. 18 Distribution of average ID along the periphery of SPS model at FBl interference condition

that the fluctuating components of drag force coefficients C_D and lift force coefficients C_L of the principal building, the model is not significantly affected by the gap distance of the interfering model.

The current study shows that when the magnitudes of Drag force coefficients are compared to magnitudes of Lift force coefficients, there is a large variation due to the normal wind directions in the model's direction of symmetry. The present study's dominant wind direction is demonstrated by along-isolated wind conditions. Where the C_D at FBl condition shows decrement, the C_L value shows near about the same magnitudes. Drag and Lift Force coefficients for interference cases largely depend upon the orientation of the interfering model in the present study (Fig. 21a, b). The shielding effect refers to the situation where, before reaching the principal structure wind has to move through any structure(s) located on the upstream wind side [10]. Due to the shielding effect, it is clear from the Avg C_p distribution of all the interference conditions that the principal structure is under suction, i.e., negative pressure. The faces lying at the interface between the twin models experience maximum suction in all interference conditions. Due to its high magnitude of coefficients, the nature of force and pressure at the interfaces dominates the resulting direction of force and pressure. Thus the resultant net pressure coefficients and net Drag force coefficients are negative in nature and are acting against the direction of wind incidence.

In Fig. 22 for FS-1 model maximum C_D of 0.80 is found at 180° isolated wind incidence followed by 0° (0.765) wind incidence with 4% decrement. Maximum C_L is manifested by 180° wind incidence is 0.075 closely followed by 180° (0.065) wind incidence with only 13% decrease. Overall minimum C_D and C_L are exhibited by FF interference condition which concludes the highest enhanced overall performance at this interference condition. An enormous decrease in C_D and C_L is shown for all interfering conditions when observed against isolated wind

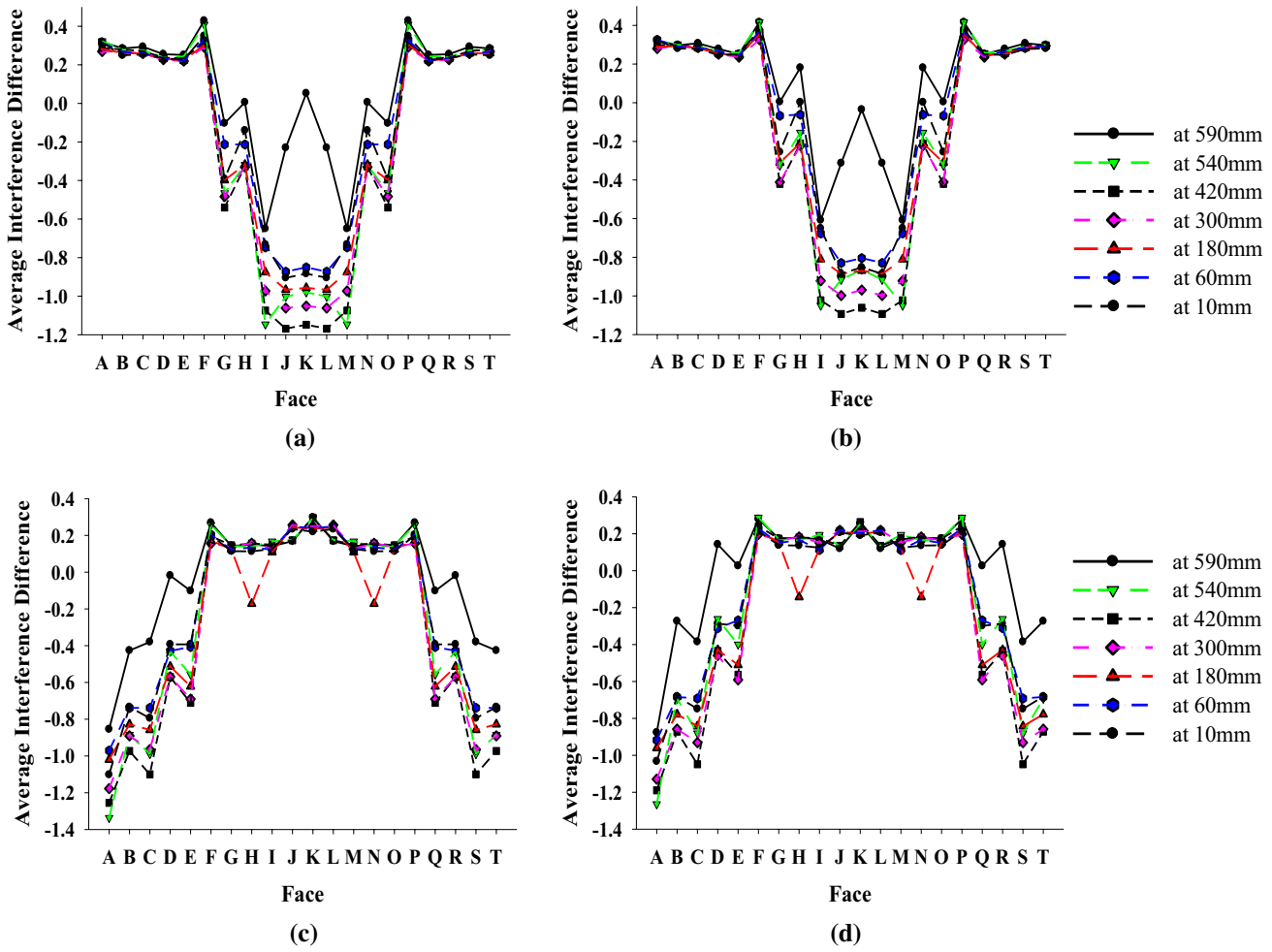


Fig. 19 Distribution of average ID along periphery of FS-1 at **a** BB; **b** FB; **c** BF; **d** FF interference condition

conditions. From the results, it is visible that the overall efficiency of the principal building is enhanced due to the interference effect.

In Fig. 23 for FS-2 model, a maximum C_D of 0.90 is found at 180° isolated wind incidence which about 12% higher than that at FS-1 similar wind incidence condition. The maximum magnitude of C_L is manifested by 180° wind incidence is −0.30 which is very high as compared to all isolated and interference conditions among SPS and FS-1 models. All the interference conditions show positive lift force coefficients, which is unlike all other IB models at interference conditions. However, for all the interference conditions at FS-2 model the magnitude of C_L ranges between 0.01 and 0.07. Unlike FS-1 model due to change in the exterior shape of the model C_L magnitudes show the appreciable difference between all interference conditions with maximum $C_L = 0.07$ exhibited by FF condition. From the comparative study, it is concluded that the overall efficiency of PB is enhanced due to the interference effect, as the magnitudes of force coefficients are very less.

3.5.2 Moment Interference Factor

A Moment Interference Factor has been defined as the ratio of the base overturning moment when there are upstream building structures present to the base overturning moment at isolated building [58–61]. Thus buffeting factor is calculated with the help of Eq. 10.

$$\text{Buffeting Factor} = \frac{\text{An overturning moment with interfering building}}{\text{An overturning moment without interfering building}} \quad (10)$$

Mara et al. [36] discovered that the overturning IF (Along-wind) at $X/D = 3$ is 0.3 for urban areas and 0.2 for open country, and that IF increases with increasing distance between twin models. For FBl increased Buffeting Factor is seen for the reduced clear gap between twin models. Even with a smaller gap between models and a change in the exterior shape of the model, all interference conditions for FS-1 model, magnitudes of Buffeting Factor show good agreement with test results by Mara et al. [36]. The Buffeting Factor (Table 6) of BF and FF at FS-2, on the other

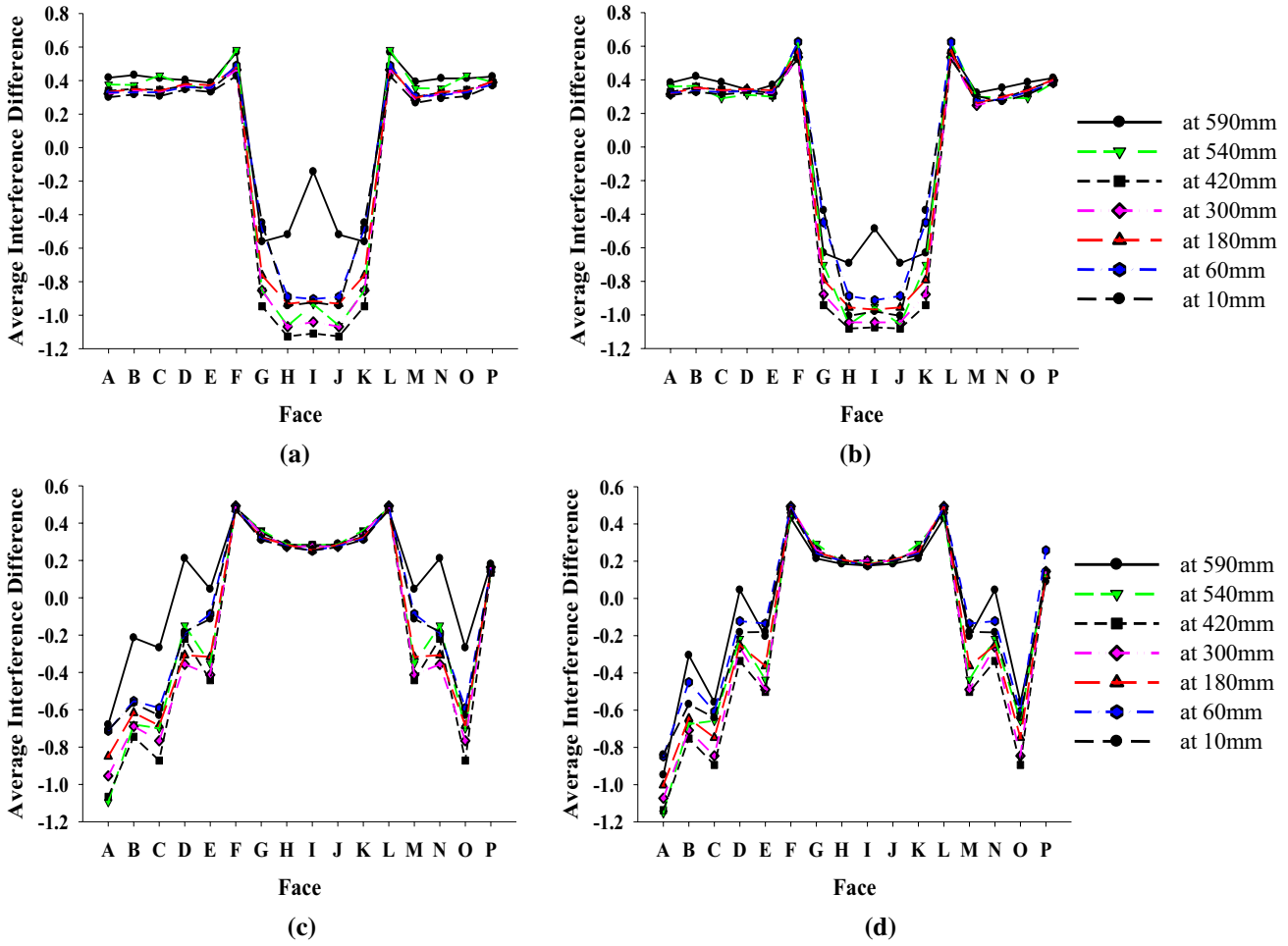


Fig. 20 Distribution of average ID along periphery of FS-2 at **a** BB; **b** FB; **c** BF; **d** FF interference condition

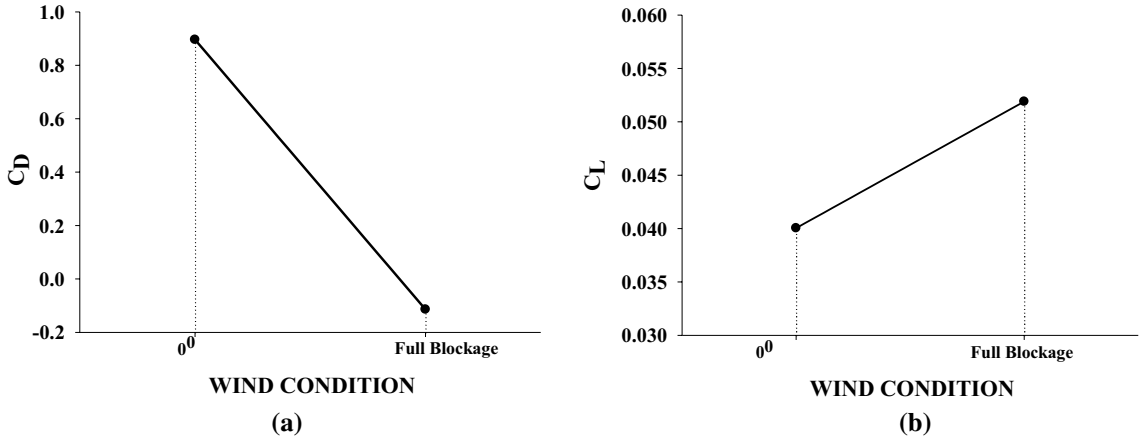


Fig. 21 Base force coefficients of SPS building model at all wind conditions **a** Drag Force Coefficients; **b** Lift Force Coefficients

hand, indicates increased along-wind overturning moment, as reported by Kim et al. [61] at variable Hr ratio. With such a small gap between twin models and a complex plan shape, the FS-1 model shows improved overall performance.

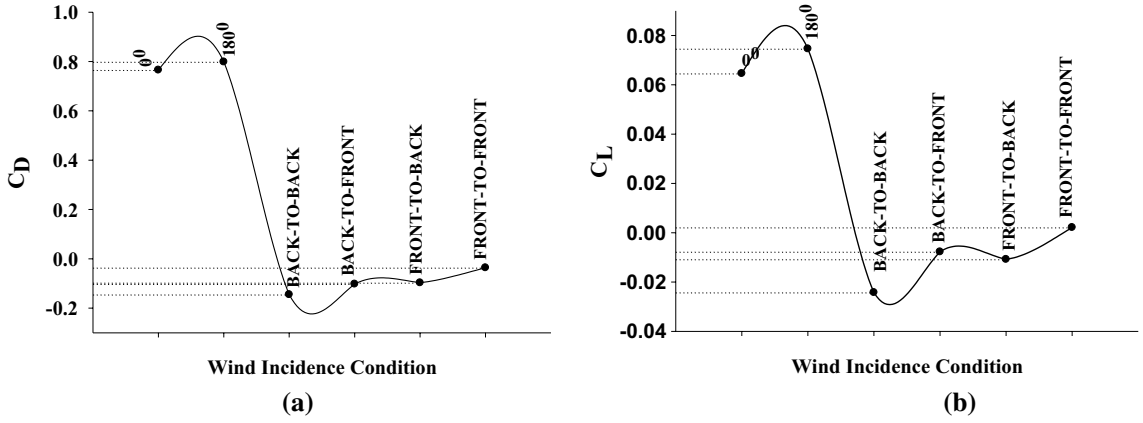


Fig. 22 Base force coefficients for FS-1 model at all wind conditions **a** Drag Force Coefficients; **b** Lift Force Coefficients

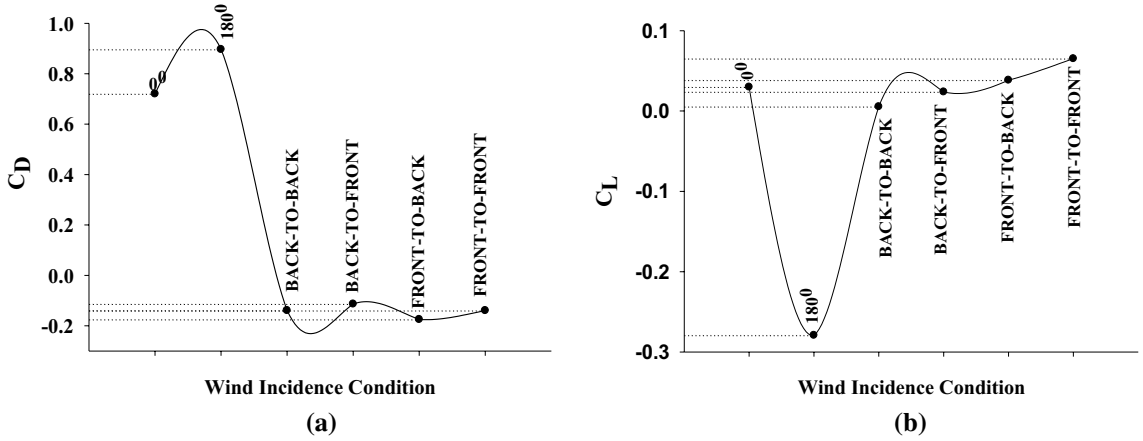


Fig. 23 Base force coefficients for FS-2 model at all wind conditions **a** Drag force coefficients; **b** Lift Force coefficients

Table 6 Buffeting Factors for Square and Fish-Plan Shape building models

Model	SPS	FS-1				FS-2			
Interference condition	FBl	BB	FB	BF	FF	BB	FB	BF	FF
Buffeting factor	0.88	−0.36	−0.05	−0.17	0.00	0.02	0.12	1.57	1.82

4 Conclusions

The purpose of this paper is to investigate the effects of twin Square and Fish-plan shape-1 and Fish-plan shape-2 building model at various orientations between twin building models under 100% blockage conditions. The experiments were carried out in conditions with a power-law exponent of 0.22 and a constant flow of wind at a velocity of 10 m/s. The plan area of all the models is 40,000mm^2. The important outcomes of the present study are summarized as below:

- The Avg C_p values of the Fish-plan shape building model differ from those of the Square and Rectangular-plan shape building models, so results from regular plan shape buildings at similar working conditions will not suffice for structural and cladding design.
- Maximum suction of principal model at Full Blockage condition is reduced to about 54% to that at the isolated condition of Square-plan shape model at same working conditions.
- Among all interference conditions, maximum suction (−0.63) is experienced by the Back-to-Front interference condition at the Fish-plan shape-1 model due to the formation channel at the interface between twin models, which is 57% higher than that at Square-plan shape model at Full Blockage condition (−0.40).

- Because of the model's unconventional cross-sectional plan shape, with gradual increase and then decrease in cross-sectional plan of Fish-plan shape building model, as opposed to Square-plan shape building model and any other interference studies, very high turbulence is seen at some faces of all interference conditions.
- $-1 >$ Avg $IF_p > 1$ as experienced by Fish-plan shape-1 and Fish-plan shape-2 models exhibits the reduced performance of the principal model in terms of generation of high suction regions due to complex exterior shape of twin building models, unlike Square-plan shape model at present working condition.
- Avg ID magnitudes of all the model's interference conditions with $-1 >$ Avg $IF_p > 1$ suggest misleading results if only the Avg IF_p factor is studied. As in some isolated conditions with near about zero Avg C_p magnitudes gives very high Avg IF_p magnitudes.
- Isolated wind condition at 180° for Fish-plan shape-1 and Fish-plan shape-2 shows the maximum drag and lift force coefficients, thus 180° orientation of isolated building model should be avoided for both models.
- All the interfering building models except the Back-to-Front and Front-to-Front interfering conditions of Fish-plan shape-2 show decreased C_L magnitudes as compared to respective isolated conditions. Back-to-Front and Front-to-Front orientation of interfering twin-building models of Fish-plan shape-2 should be avoided.
- All the interfering building models except the Back-to-Front and Front-to-Front interfering conditions of Fish-plan shape-2 show enhanced overturning moment (Along-wind) performance as compared to their respective isolated conditions. As a result, the Back-to-Front and Front-to-Front interfering conditions of Fish-plan shape-2 are identified as unfavorable orientation conditions because of increased overturning moments.
- The study of interference effects on principal building with twin building model in various other configurations, such as varying angle of wind incidence on 100% blockage in various tandem and staggered arrangements between twin building models, is to be addressed in future research studies.

Acknowledgements Grateful acknowledgment is owed to Prof. A. K. Ahuja, Visiting Faculty, Civil Engineering Department, IIT Jammu (India). Without his support and invaluable suggestions, this experiment would not have been feasible.

References

1. El-Heweity, M.M.; Abdelnaby, M.H.; Eshra, E.M.: Numerical simulation of buffeting longitudinal wind forces on buildings. Alex. Eng. J. **58**, 225–236 (2019). https://doi.org/10.1016/j.aej.2018.08.001
2. Xu, A.; Xie, Z.N.; Fu, J.Y.; Wu, J.R.; Tuan, A.: Evaluation of wind loads on super-tall buildings from field-measured wind-induced acceleration response. Struct. Des. Tall Spec. Build. **23**, 641–663 (2014). https://doi.org/10.1002/tal.1065
3. Zheng, C.; Xie, Y.; Khan, M.; Wu, Y.; Liu, J.: Wind-induced responses of tall buildings under combined aerodynamic control. Eng. Struct. **175**, 86–100 (2018)
4. Aly, A.M.: Pressure integration technique for predicting wind-induced response in high-rise buildings. Alex. Eng. J. **52**, 717–773 (2013). https://doi.org/10.1016/j.aej.2013.08.006
5. Farouk, M.I.: Check the comfort of occupants in high rise building using CFD. Ain Shams Eng. J. **7**, 953–958 (2016)
6. AS/NZS: 1170.2:2011. Structural Design Actions, Part 2: Wind Actions. Australian/ New Zealand Standard (2011)
7. ASCE: 7-02: Minimum Design Loads for Buildings and Other Structures 2002 (2002)
8. BS 6399-2:1997 British standard: loading for buildings part 2. Code of practice for wind loads. British Standard Institution. London 1997
9. EN 1991-1-4:2005/AC: 2010(E). European Standard Eurocode 1: actions on structures-part 1–4: General actions -wind actions. European Committee for Standardization (CEN). Europe 2010
10. IS: 875- Part-3 (2015). Code of Practice for Design Loads (other than earthquake loads) for Buildings and Structures- Wind Loads. India 2015
11. Chakraborty, S.; Dalui, S.K.; Ahuja, A.K.: Wind load on irregular plan shaped tall building—a case study. Wind Struct. **19**(1), 59–73 (2014)
12. Chakraborty, S.; Dalui, S.K.; Ahuja, A.K.: Experimental and numerical study of surface pressure on '+' plan shape tall building. J. Civ. Eng. **8**, 251–262 (2013)
13. Cheng, L.; Kit, M.L.; Wong, S.Y.: POD analysis of crosswind forces on a tall building with square and H-shaped cross sections. Wind Struct. Int. J. **21**(1), 63–84 (2015)
14. Nagar, S.K.; Raj, R.; Dev, N.: Experimental study of wind-induced pressures on tall buildings of different shapes. Wind Struct. **31**(5), 441–453 (2020). https://doi.org/10.12989/was.2020.31.5.441
15. Paul, R.; Dalui, S.K.: Wind effects on 'Z' plan-shaped tall building: a case study. Int. J. Adv. Struct. Eng. **8**, 319–335 (2016)
16. Gomes, M.G.; Rodrigues, A.M.; Mendes, P.: Experimental and numerical study of wind pressures on irregular-plan shapes. J. Wind Eng. Ind. Aerodyn. **93**, 741–756 (2005)
17. Amin, J.A.; Ahuja, A.K.: Experimental study of wind-induced pressures on buildings of various geometries. Int. J. Eng. Sci. Technol. **3**(5), 1–19 (2011)
18. Mallick, M.; Kumar, A.; Patra, K.C.: Experimental investigation on the wind-induced pressures on C-shaped buildings. KSCE J. Civ. Eng. **23**(8), 3535–3546 (2019)
19. Li, Y.; Li, S.Q.; Chen, F.: Wind tunnel study of wind-induced torques on L-shaped tall buildings. J. Wind Eng. Ind. Aerodyn. **167**, 41–50 (2017)
20. Bhattacharyya, B.; Dalui, S.K.; Ahuja, A.K.: Wind induced pressure on 'E' plan shaped tall buildings. Jordan J. Civ. Eng. (2014). https://doi.org/10.1016/j.engstruct.2018.08.031
21. Bhattacharyya, B.; Dalui, S.K.: Experimental and numerical study of wind-pressure distribution on irregular-plan-shaped building. J. Struct. Eng. **146**(7), 04020137 (2020). https://doi.org/10.1061/(ASCE)ST.1943-541X.0002686
22. Bandi, E.K.; Tamura, Y.; Yoshidaa, A.; Kim, Y.C.; Yang, Q.: Experimental investigation on aerodynamic characteristics of various triangular-section high-rise buildings. J. Wind Eng. Ind. Aerodyn. **122**, 60–68 (2013)

23. Bandi, E.K.; Tanaka, H.; Kim, Y.; Ohtake, K.; Tamura, Y.: Peak pressures acting on tall buildings with various configurations. Int. J. High. Rise Build. **2**, 229–244 (2013)
24. Tamura, Y.; Tanaka, H.; Ohtake, K.; Kim, Y.C.; Yoshida, A.; Bandi, E.K.; Xu, X.; Yang, Q.: Aerodynamic control of wind-induced vibrations and flow around super-tall buildings. In: Proceedings of the 6th International Conference Adv. Exp. Strutural Eng. International Work. Adv. Smart Mater, Smart Strutures Technol. Urbana-Champaign
25. Ming, G.: Wind-resistant studies on tall buildings and structures. Sci. China Technol. Sci. Oct. **53**(10), 2630–2646 (2010). https://doi.org/10.1007/s11431-010-4016-2
26. Hayashida, H.; Iwasa, Y.: Aerodynamic shape effects of tall buildings for vortex induced vibration. J. Wind Eng. Ind. Aerodyn. **33**, 237–242 (1990)
27. Ahmed, E.; Bitsuamlak, G.; Damatty, A.E.: Enhancing wind performance of tall buildings using corner aerodynamic optimization. Eng. Struct. **136**, 133–148 (2017)
28. Stathopoulos, T.; Zhou, Y.: Numerical simulation of wind induced pressures on buildings of various geometries. J. Wind Eng. Ind. Aerod. **46–47**, 419–430 (1993)
29. Sharma, A.; Mittal, H.; Gairola, A.: Mitigation of wind load on tall buildings through aerodynamic modifications: review. J. Build. Eng. **18**, 180–194 (2018)
30. Xie, J.: Aerodynamic optimization in super-tall building designs. In: The Seventh International Colloquium on Bluff Body Aerodynamics and its Applications (BBAA7) Shanghai, China, pp. 2–6 (2012)
31. Lam, K.M.; Zhao, J.G.; Leung, M.Y.H.: Wind-induced loading and dynamic responses of a row of tall buildings under strong interference. J. Wind Eng. Ind. Aerodyn. **99**, 573–583 (2011)
32. Hui, Y.; Tamura, Y.; Yoshida, A.: Mutual interference effects between two high-rise building models with different shapes on local peak pressure coefficients. J. Wind Eng. Ind. Aerodyn. **104–106**, 98–108 (2012)
33. Zu, G.B.; Lam, K.M.: Across-wind excitation mechanism for interference of twin tall buildings in staggered arrangement. J. Wind Eng. Ind. Aerodyn. **177**, 167–185 (2018)
34. Zu, G.B.; Lam, K.M.: Across-wind excitation mechanism for interference of twin tall buildings in tandem arrangement. Wind Struct. **26**(6), 397–413 (2018)
35. Yu, X.F.; Xie, Z.N.; Zhu, J.B.; Gu, M.: Interference effects on wind pressure distribution between two high-rise buildings. J. Wind Eng. Ind. Aerodyn. **142**, 188–197 (2015)
36. Mara, T.G.; Terry, B.K.; Ho, T.C.E.; Isyumov, N.: Aerodynamic and peak response interference factors for an upstream square building of identical height. J. Wind Eng. Ind. Aerodyn. **133**, 200–210 (2014)
37. Cho, K.; Hong, S.; Hwang, K-S.: Effects of neighbouring building on wind loads. In: CTBUH, Seoul Conference (2004)
38. Bairagi, A.K.; Dalui, S.K.: Optimization of interference effects on high rise buildings for different wind angles using CFD simulation. Electron. J. Struct. Eng. **14**, 39–49 (2014)
39. Gurjar, S.; Amin, J.A.: Numerical simulation of wind induced mean interference between two tall buildings. J. Mater. Eng. Struct. **4**, 181–192 (2017)
40. Amin, J.A.; Ahuja, A.K.: Wind-induced mean interference effects between two closed spaced buildings. KSCE J. Civ. Eng. **16**(1), 119–131 (2012)
41. Kar, R.; Dalui, S.K.: Wind interference effect on an octagonal plan shaped tall building due to square plan shaped tall buildings. Int. J. Adv. Struct. Eng. **8**, 73–86 (2016). https://doi.org/10.1007/s40091-016-0115-z
42. Xie, Z.N.; Gu, M.: Mean interference effects among tall buildings. Eng. Struct. **26**, 1173–1183 (2004)
43. Mukherjee, S.; Bairagi, A.K.: Interference effect on principal building due to setback tall building under wind excitation. In: SEC18: Proceedings of the 11th Structural Engineering Convention, Paper No. 20180306. (2018)
44. Kheyari, P.; Dalui, S.K.: Estimation of wind load on a tall building under interference effects: a case study. Jordan J. Civ. Eng. **9**(1), 2015–2084 (2015)
45. Daemei, A.B.; Khotbehsara, E.M.; Nobarani, E.M.; Bahrami, P.: Study on wind aerodynamic and flow characteristics of triangular shaped tall buildings and CFD simulation in order to assess drag coefficient. Ain Shams Eng. J. **10**, 541–548 (2019)
46. Khanduri, A.C.; Stathopoulos, T.; Bedard, C.: Wind-induced interference effects on buildings—a review of the state-of-the-art. Eng. Struct. **20**(7), 617–630 (1998)
47. Cook, N.J.: The Designers Guide to Wind Loading of Building Structures, Part 2: Static Structures, Building Research Establishment Report, Butterworths, London, UK (1985)
48. Houghton, E.L.; Carruthers, N.B.: Wind Forces on Buildings and Structures: An Introduction. Wiley, New Jersey (1976)
49. Kim, C.Y.; Kanda, J.: Wind pressures on tapered and set-back tall buildings. J. Fluids Struct. **39**, 306–321 (2013)
50. Hui, Y.; Yoshida, A.; Tamura, Y.: Interference effects between two rectangular-section high-rise buildings on local peak pressure coefficients. J. Fluids Struct. **37**, 120–133 (2013)
51. Flaga, A.; Kocon, A.; Kłaput, R.; Bosak, G.: The environmental effects of aerodynamic interference between two closely positioned irregular high buildings. J. Wind Eng. Ind. Aerodyn. **180**, 276–287 (2018)
52. Lam, K.M.; Leung, M.Y.H.; Zhao, J.G.: Interference effects on wind loading of a row of closely spaced tall buildings. J. Wind Eng. Ind. Aerodyn. **96**, 562–583 (2008)
53. Yu, X.; Xie, Z.; Gu, M.: Interference effects between two tall buildings with different section sizes on wind-induced acceleration. J. Wind Eng. Ind. Aerodyn. **182**, 16–26 (2018)
54. Ding, F.; Kareem, A.: Tall buildings with dynamic facade under winds. Engineering (2020). https://doi.org/10.1016/j.eng.2020.07.020
55. Elshaer, A.; Bitsuamlak, G.: Multiobjective aerodynamic optimization of tall building openings for wind-induced load reduction. J. Struct. Eng. **144**(10), 04018198 (2018). https://doi.org/10.1061/(ASCE)ST.1943-541X.0002199
56. Sakamoto, H.; Haniu, H.: Aerodynamic forces acting on two square prisms placed vertically in a turbulent boundary layer. Wind Eng. Ind. Aerodyn. **31**, 41–66 (1988)
57. Song, J.; Tse, K.T.; Tamura, Y.; Kareem, A.: Aerodynamics of closely spaced buildings: with application to linked buildings. J. Wind Eng. Ind. Aerodyn. **149**, 1–16 (2016)
58. Blessman, J.: Buffeting effects on neighbouring tall buildings. J. Wind Eng. Ind. Aerodyn. **18**, 105–110 (1985)
59. Bailey, P.A.: Interference excitation of twin tall buildings. J. Wind Eng. Ind. Aerodyn. **21**, 323–338 (1985)
60. Saunders, J.W.; Melbourne, W.H.: Buffeting effects of upstream buildings. In: Proceedings of the 5th International Conference on Wind Engineering, Fort Collins, CO. 1979, Pergamon Press, Oxford, pp. 593–606 (1980)
61. Kim, W.; Tamura, Y.; Yoshida, A.: Interference effects on aerodynamic wind forces between two buildings. J. Wind Eng. Ind. Aerodyn. **147**, 186–201 (2015)

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 2939334, 13 pages
https://doi.org/10.1155/2021/2939334

Research Article

Explainable Artificial Intelligence for Sarcasm Detection in Dialogues

Akshi Kumar[1,2] **Shubham Dikshit**[3] **and Victor Hugo C. Albuquerque**[1,4]

[1]*Graduate Program on Telecommunication Engineering, Federal Institute of Education, Science and Technology of Ceará, Fortaleza, CE, Brazil*
[2]*Department of Computer Science and Engineering, Delhi Technological University, Delhi, India*
[3]*Department of Computer Science and Engineering, IMS Engineering College, Ghaziabad, India*
[4]*Graduate Program on Teleinformatics Engineering, Federal University of Ceará, Fortaleza, CE, Brazil*

Correspondence should be addressed to Akshi Kumar; akshikumar@dce.ac.in

Received 17 May 2021; Revised 11 June 2021; Accepted 21 June 2021; Published 3 July 2021

Academic Editor: Mohammad R. Khosravi

Copyright © 2021 Akshi Kumar 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.

Sarcasm detection in dialogues has been gaining popularity among natural language processing (NLP) researchers with the increased use of conversational threads on social media. Capturing the knowledge of the domain of discourse, context propagation during the course of dialogue, and situational context and tone of the speaker are some important features to train the machine learning models for detecting sarcasm in real time. As situational comedies vibrantly represent human mannerism and behaviour in everyday real-life situations, this research demonstrates the use of an ensemble supervised learning algorithm to detect sarcasm in the benchmark dialogue dataset, MUStARD. The punch-line utterance and its associated context are taken as features to train the eXtreme Gradient Boosting (XGBoost) method. The primary goal is to predict sarcasm in each utterance of the speaker using the chronological nature of a scene. Further, it is vital to prevent model bias and help decision makers understand how to use the models in the right way. Therefore, as a twin goal of this research, we make the learning model used for conversational sarcasm detection interpretable. This is done using two post hoc interpretability approaches, Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive exPlanations (SHAP), to generate explanations for the output of a trained classifier. The classification results clearly depict the importance of capturing the intersentence context to detect sarcasm in conversational threads. The interpretability methods show the words (features) that influence the decision of the model the most and help the user understand how the model is making the decision for detecting sarcasm in dialogues.

1. Introduction

Natural language is a vital information source of human sentiments. Automated sarcasm detection is often described as a natural language processing (NLP) problem as it primarily requires understanding the human expressions, language, and/or emotions articulated via textual or nontextual content. Sarcasm detection has attracted growing interest over the past decade as it facilitates accurate analytics in online comments and reviews [1, 2]. As a figurative literary device, sarcasm makes use of words in a way that deviates from the conventional order and meaning thereby misleading polarity classification results. For example, in a statement "*Staying up till 2:30am was a brilliant idea to miss my office meeting*," the positive word "*brilliant*" along with the adverse situation "*miss my office meeting*" conveys the sarcasm, because sarcasm has an implied sentiment (negative) that is different from surface sentiment (positive due to presence of "brilliant"). Various rule-based, statistical, machine learning, and deep learning-based approaches have been reported in pertinent literature on automatic sarcasm detection in single sentences that often rely on the content of utterances in isolation. These include a range of techniques such as sense disambiguation [3] to polarity flip detection in text [4] and multimodal (text +image) content [5, 6]. Furthermore, its use on social media platforms like Twitter and Reddit is primarily to convey user's frivolous intent, and therefore, the dialect is more casual and includes the use of microtext like

wordplay, neologism, emojis, and slangs. Few recent works have taken into account the additional contextual information along with the utterance to deal with these challenges. Different researchers have considered varied operational cues to typify context. In 2019, Kumar and Garg [7] defined five broad categories of context, namely, social-graph, temporal, content, modality, and user-profile based which can be used for improving the classification accuracy. Evidently, it is essential to capture the operational concern, that is, the pragmatic meaning defined by "context" as sarcasm. But the use of sarcasm in dialogues and conversational threads has further added to the challenges making it vital to capture the knowledge of the domain of discourse, context propagation during the course of dialogue, and situational context and tone of the speaker. For example, recently, several Indian airlines took to Twitter to engage users in a long thread meant to elicit laughs and sarcastic comebacks amid the coronavirus lockdown that has kept passengers and airlines firmly on the ground. IndiGo playfully teased its rivals by engaging in a Twitter banter resulting in comic wordplays on airlines' advertising slogans. IndiGo began by asking Air Vistara "*not flying higher?*" in reply to which the airlines tagged peer GoAir, punning on its tagline "fly smart" and what followed was other key airlines like AirAsia and SpiceJet joining the thread exchange equipped with witty responses using each other's trademark business taglines (https://www.deccanherald.com/business/coronavirus-indigo-vistara-spicejet-engage-in-banter-keep-twitterati-in-splits-amid-lockdown-blues-823677.html).

As seen in Figure 1, it is not only important to capture the intrasentence context but the intersentence context too to detect sarcasm in conversational threads. Moreover, the sarcastic intent of the thread is difficult to comprehend without the situational context as in this case is the unprecedented travel restrictions, including the grounding of all domestic and international passenger flights, to break the chain of the coronavirus disease (COVID-19) transmission.

But as sarcasm is a convoluted form of expression which can cheat and mislead analytic systems, it is equally important to achieve high prediction accuracy with decision understanding and traceability of actions taken. As models cannot account for all the factors that will affect the decision, explainability can account for context and help understand the included factors that will affect decision making so that one can adjust prediction on additional factors. Explainable artificial intelligence (XAI) [8, 9] is the new buzzword in the domain of machine learning which intends to justify the actions and understand the model behaviour. It enables building robust models with better decision-making capabilities.

Thus, in this paper, we firstly demonstrate the role of context in conversational threads to detect sarcasm in the MUStARD dataset [5], which is a multimodal video corpus for research in automated sarcasm discovery compiled using dialogues from famous sitcoms, namely, "Friends" by Bright, Kauffman, Crane Productions, and Warner Bros. Entertainment Inc., "The Big Bang Theory" by Chuck Lorre, Bill Prady, CBS, "The Golden Girls" by Susan Harris, NBC, and "Sarcasmaholics Anonymous." The data is labelled with true

Figure 1: Online sarcastic conversational thread.

and false for the sarcastic and nonsarcastic dialogues using the sequential nature of scenes in the episodes, and we use eXtreme Gradient Boosting (XGBoost) method [10] to primarily investigate how conversational context can facilitate automatic prediction of sarcasm. As a twin goal of this research, we aim to make the supervised learning models used for conversational sarcasm detection interpretable with the help of XAI. The goal is to show the words (features) that influence the decision of the model the most.

Using dialogue dataset from sitcoms can invariably relate to any real-life utterance making this work relevant for various sentiment analysis-based market and business intelligence applications for assessing insights from conversational threads on social media. Most situational comedies or sitcoms are led by the comedy of manners, vaudeville, and our tacit perceptions of everyday life. These are the story of our psychodynamics and sociodynamics on situations that could arise in everyday life and unfold the unexpected and ironic comedy of human behaviour in real-life situations. For example, in Friends, season 10, episode 3, Ross walks in with a clearly overdone tan to the point that his skin color is very dark and looks truly ridiculous. He tells Chandler that he went to the tanning place his wife (Monica) suggested. And Chandler came up with a sarcastic statement "*Was that place the sun?*" as it looked like the only tanning place that could make someone's skin look like that would be sitting directly beneath the scorching sun! The sarcasm in Chandler's dialogue could only be understood considering the entire conversation and not taking his dialogue in isolation (Figure 2).

XAI in a typical NLP task setting can offer twofold advantages, namely, transferability, as machine learning models are trained in a controlled setting, deployment in real time should also ensure that the model has truly learned to detect underlying phenomenon, and secondly, it can help determining the contextual factors that affect the decision. The terms interpretability and explainability are often used

Figure 2: Friends: season 10, episode 3.

interchangeably as both play a complementary role in understanding predictive models [11]. The term interpretability tells us what is going on in the algorithm, i.e., it enables us to predict what will happen if there are some changes in the parameters or input, and explainability tells the extent to which the internal working of any machine learning or deep learning model can be explained in human terms. Characteristically, interpretable machine learning systems provide explanations for their outputs. According to Miller [12], interpretability is defined as the capability to understand the decision and means that the cause and effect can be determined. Interpretable machine learning (ML) describes the process of revealing causes of predictions and explaining a derived decision in a way that is understandable to humans. The ability to understand the causes that lead to a certain prediction enables data scientists to ensure that the model is consistent with the domain knowledge of an expert. Furthermore, interpretability is critical to obtain trust in a model and to be able to tackle problems like unfair biases or discrimination. One way to apply interpretable ML is by using models that are intrinsically interpretable and known to be easy for humans to understand such as linear/logistic regression, decision trees, and K-nearest neighbors [13]. Alternatively, we can train a black-box model and apply post hoc interpretability techniques [14] (Figure 3) to provide explanations.

In this paper, we use two post hoc model agnostic explainability techniques called Local Interpretable Model-agnostic Explanations (LIME) [15, 16] and Shapley Additive exPlanations (SHAP) [17, 18] to analyze the models on the dataset by checking the evaluation metrics and select the model where explanation can be separated from the models. The intent is to evaluate the black-box model much easily on how each word plays an important role in the prediction of the sarcastic dialogues by the speaker using the sequential nature of a scene in the TV series. Thus, the key contributions of this research are as follows:

	Global	Local
Model-specific	Model internals; Intrinsic feature importance	Rule sets (Tree structure)
Model-agnostic	Partial dependence plots; Feature importance; Global surrogate models	Individual conditional expectation; Local surrogate models

Figure 3: Post hoc interpretability techniques.

(i) Using sequence of utterances to detect sarcasm in real-time dialogues

(ii) Using post hoc model-agnostic local surrogate machine learning interpretability methods to comprehend which words within a dialogue are the most important for predicting sarcasm

The scope of the research can be extended to real-time AI-driven sentiment analysis for improving customer experience where these explanations would help the service desk to detect sarcasm and word importance while predicting sentiment. The organization of the paper is as follows: the next section briefs about the taxonomy of machine learning interpretability methods followed by related work within the domain of sarcasm detection specifically in conversational data in Section 3. Section 4 discusses the key techniques used in this research followed by the results and conclusion in Section 5 and Section 6, respectively.

2. Taxonomy of Machine Interpretability Methods

Artificial intelligence (AI) is gradually participating in day-to-day experiences. Its entrusted adoption and encouraging acceptance in various real-time domains are highly contingent upon the transparency, interpretability, and explainability of models built. Particularly in customer-centric environments, trust and fairness can help customers achieve better outcomes. Introduced in the early 1980s, XAI is a framework and tool which helps humans to understand the model behaviour and enables building robust models with better decision-making capabilities. It is used for understanding the logic behind the predictions made by the model and justifies its results to the user.

A trade-off between the model interpretability and predictive power is commonly observed as shown in Figure 4. As the model gets more advanced, it becomes harder to explain how it works. High interpretability models include traditional regression algorithms (linear models, for example), decision trees, and rule-based learning. Basically, these are approximate monotonic linear functions. On the other hand, low interpretability models include ensemble methods and deep learning where the black-box feature extraction offers poor explainability.

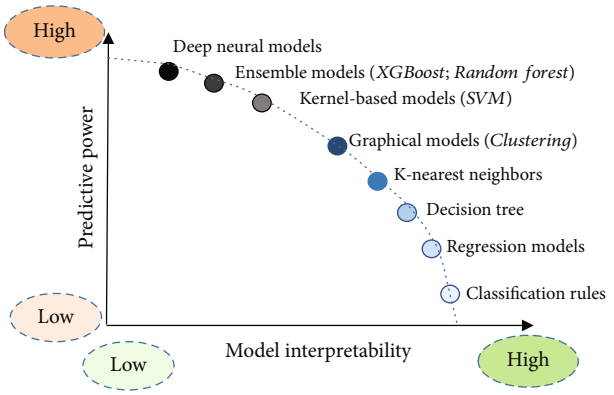


Figure 4: Predictive power vs. interpretability trade-off.

Machine interpretability methods are often categorized along three main criteria [19, 20]. The first discriminates based on the coverage of explanation as local or global for explanation for at instance-level (individual predictions) or model-level (entire model), respectively. Global interpretability methods explain the entire ML model at once from input to prediction, for example, decision trees and linear regression. Local interpretability methods explain how predictions change for when input changes and are applicable for a single prediction or a group of predictions. The second criteria differentiate between the explanations based on the interpretable design capabilities as intrinsically interpretable models and post hoc models (Figure 5). Intrinsically interpretable models are models that are interpretable by design, and no postprocessing steps are needed to achieve interpretability. These are self-explaining, and explainability is often achieved as a by-product of model training. On the other hand, in post hoc methods, explainability is often achieved after the model is trained and it requires postprocessing using external methods to achieve interpretability.

The third criterion to categorize interpretability methods is the applicability limitation to specific models or any ML model. Based on these criteria, the methods are divided into model-specific and model-agnostic methods. Model-specific techniques can be used for a specific architecture and require training the model using a dataset. Intrinsic methods are by definition model-specific. On the contrary, model-agnostic methods can be used across many black-box models without considering their inner processing or internal representations and do not require training the model. Post hoc methods are usually model-agnostic.

Post hoc interpretability methods consider interpretability of predictions made by black-box models after they have been built. These can further be categorized into four categories as surrogate models, feature contribution, visualisations, and case-based methods [19, 21]. Figure 6 shows the key model-agnostic methods available in literature [14].

In this work, we use two popular Python libraries, SHAP and LIME, to interpret the output and leverage model explanations.

3. Related Work

There is notable literary evidence apropos the versatile use of machine learning and deep learning algorithms for automated sarcasm detection. In the past, rule-based algorithms were employed initially to detect sarcasm [22]. Later, many researchers [23–29] used ML algorithms to detect sarcasm in textual content. Naive Bayes and fuzzy clustering models were employed by Mukherjee et al. [30] for sarcasm detection in microblogs. The researchers concluded that Naive Bayes models are more effective and relevant than the fuzzy clustering models. Prasad et al. [31] analyzed and compared various ML and DL algorithms to conclude that gradient boost outperforms the other models in terms of accuracy. In 2018, Ren et al. [32] employed contextual information for sarcasm detection on Twitter dataset by utilizing two different context-augmented neural models. They demonstrated that the proposed model performs better than the other SOTA models. In 2019, Kumar and Garg [33] compared various ML techniques like SVM, DT, LR, RF, KNN, and NN for sarcasm detection on Twitter and Reddit datasets. A hybrid deep learning model of soft attention-based bi-LSTM and convolution neural network with GloVe for word embeddings was proposed by Kumar et al. [34]. The results demonstrated that the proposed hybrid outperforms CNN, LSTM, and bi-LSTM. Kumar and Garg [4] reported a study on context-based sarcasm detection on Twitter and Reddit datasets using a variety of ML techniques trained using tf-idf and DL techniques using GloVe embedding.

Recent studies have also been reported on multimodal sarcasm detection. In 2019, Cai et al. [35] used bi-LSTM for detection of sarcasm in multimodal Twitter data. In the same year, Kumar and Garg [6] employed various supervised ML techniques to study context in sarcasm detection in typographic memes and demonstrated that multilayer perceptron is best among all the models. In 2020, a study by Kumar et al. [36] built a feature-rich support vector machine and proposed a multihead attention-based bi-LSTM model for sarcasm detection in Reddit comments. Few studies on sarcasm detection in online multilingual content have also been reported. In 2020, Jain et al. [2] had put forward a hybrid of bi-LSTM with softmax attention and CNN for sarcasm detection in multilingual tweets. In 2021, Farha et al. [37] compared many transformer-based language models like BERT and GPA on Arabic data for sarcasm detection. Faraj et al. [38] proposed a model based on ensemble techniques with an AraBERT pretrained model for sarcasm detection in Arabic text with an accuracy of 78%.

Sarcasm detection in conversations and dialogues has created a great interest with NLP researchers. Ghosh et al. [39] used conditional LSTM and LSTM with sentence-level attention to understand the role of context in social media discussions. Hazarika et al. [40] proposed a CASCADE (a ContextuAl SarCasm DEtector) model which extracted contextual information from online social media discussions on Reddit to detect sarcasm by taking into consideration

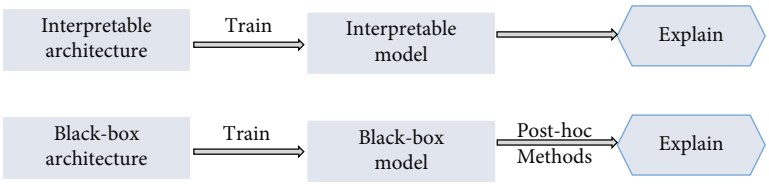


Figure 5: Intrinsic vs. post hoc interpretability.

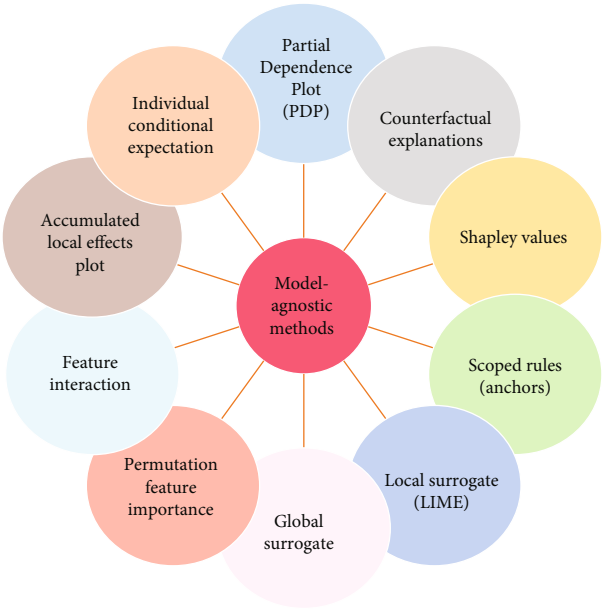


Figure 6: Model-agnostic methods.

stylometric and personality features of the users and trained a CNN for content-based feature extraction. Castro et al. [5] proposed the MUStARD dataset which contains audio-visual data from popular sitcoms and showed how multimodal cues enhance the primary sarcasm classification task. In 2020, Baruah et al. [41] implemented BERT, bi-LSTM, and SVM classifiers for sarcasm detection utilizing the context of conversations. Jena et al. [42] performed the task of sarcasm detection in conversations using a C-Net model which comprised BERT models. Recently, Zhang et al. [43] proposed a model based on quantum theory and fuzzy logic to detect sarcasm in conversations in MUStARD and Reddit datasets.

The use of explainable AI for interpretability of the underlying ML techniques for sarcasm detection has been studied by few researchers. In 2018, researchers Tay et al. [44] improved the interpretability of the algorithms by employing multidimensional intra-attention mechanisms in their proposed attention-based neural model. The proposed model was validated on various benchmark datasets of Twitter and Reddit and compared with other baseline models. Akula et al. [45] focused on detecting sarcasm in texts from online discussion forums of Twitter, dialogues, and Reddit datasets by employing BERT for multihead self-attention and gated recurrent units, to develop an interpretable DL model as self-attention is inherently interpretable.

4. XAI for Sarcasm Detection in Dialogue Dataset

Black-box ML models have observable input-output relationships but lack transparency around inner workings. This is typical of deep-learning and boosted/random forest models which model very complex problems with high nonlinearity and interactions between inputs. It is important to decompose the model into interpretable components and simplify the model's decision making for humans. In this research, we use XAI to provide insights into the decision points and feature importance used to make a prediction about sarcastic disposition of conversations. The architectural flow of the research undertaken in this paper is shown in Figure 7.

The MUStARD dataset used for this research consists of 690 dialogues by the speakers from four famous television shows. It is publicly available and manually annotated for sarcasm. The dataset consists of details about the speaker, utterance, context, context speakers, and sarcasm. For example, the dataset entry for a conversational scene as given in Figure 8 from Friends, season 2, episode 20, is shown in Table 1.

It is noted that most of the dialogues in this dataset are from two most popular shows, namely, the Big Bang Theory and Friends. The data is balanced with an equal number of sarcastic and nonsarcastic dialogues. Figure 9 shows the dataset distribution for the respective TV shows.

The following subsections discuss the details.

4.1. Supervised Machine Learning for Sarcasm Detection in Dialogues. The data was cleaned as the dialogues obtained had some errors in spelling, emoticons, and unnecessary brackets and names of the subtitle providers; any column which had any missing values or wrong data was removed from the dataset. The evaluation of an utterance relies strongly on its context. The contextual interaction between associated chronological dialogues is based on conversational common ground and thereby raising it to prominence in the current context as shown in Figure 10.

Therefore, we use the punch-line utterance, its accompanied context, and the sarcastic/nonsarcastic label to train our model. tf-idf vectorization [46] is done to transform the textual features into representation of numbers. The data is trained using an ensemble learning approach, eXtreme Gradient Boosting (XGBoost). As a popular implementation of gradient tree boosting, XGBoost provides superior classification performance in many ML challenges. In gradient

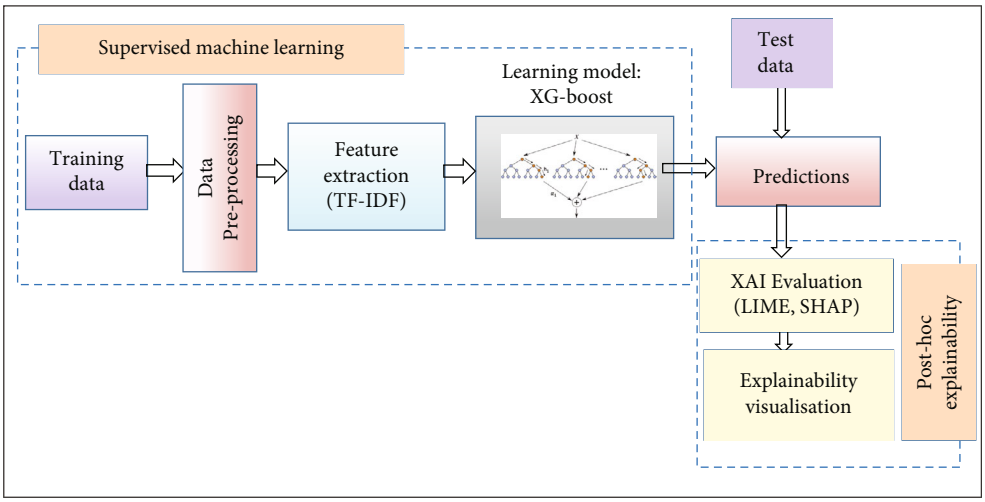


Figure 7: The architectural flow of the research undertaken.

Figure 8: Friends, season 2, episode 20.

boosting, a shallow and weak tree is first trained and then the next tree is trained based on the errors of the first tree. The process continues with a new tree being sequentially added to the ensemble, and the new successive tree improves on the errors of the ensemble of preceding trees. The key advantages of using XGBoost are that it is highly flexible, leverages the power of parallel processing, supports regularization, handles missing values, allows tree pruning, and has built-in cross-validation and high computational speed. On the flip side, explaining the XGBoost predictions seems hard and powerful tools are required for confidently interpreting tree models such as XGBoost. Subsequently, we discuss the two model-agnostic methods selected for seeking explanations that justify and rationalize the black-box model of XGBoost for sarcasm detection in dialogues.

4.2. Post Hoc Explainability Models for Sarcasm Detection in Dialogues. Post hoc interpretability approaches propose to generate explanations for the output of a trained classifier in a step distinct from the prediction step. These approximate the behaviour of a black box by extracting relationships between feature values and the predictions. Two widely accepted categories of post hoc approaches are surrogate models and counterfactual explanations [14]. Surrogate

Table 1: Dataset entry for the given conversational scene from Friends.

Utterance	An utterance is a unit of speech bound by breaths or pauses. The dialogues are spoken by the speaker with respect to the context in the scene.	But younger than some buildings!
Speaker	The character of the series who is giving the dialogue delivery.	Chandler
Context speakers	The side characters to whom the dialogue is being uttered by the main character of that scene.	Chandler
Context	The reason or the scene on the series which led to the dialogue utterance by the speaker.	I know Richard's really nice and everything, it's just that we do not know him really well you know, plus he is old *(Monica glares)* -er than some people.
Show	Name of the show	Friends
Sarcasm	This is the feature in the data to show whether the dialogue utterance by the speaker is sarcastic or nonsarcastic utterance is given as true and nonsarcastic comment is given as false in the dataset.	True

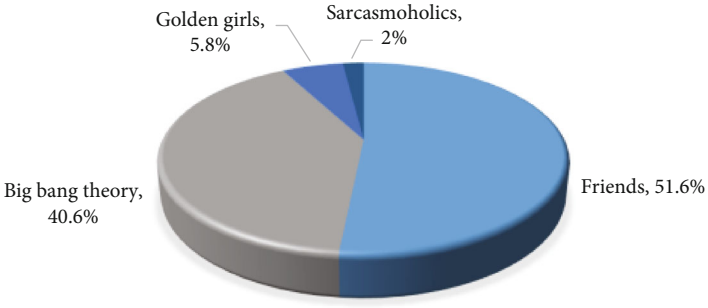


Figure 9: Dataset distribution for TV shows.

Figure 10: Utterance and associated context from a scene in The Golden Girls.

model approaches are aimed at fitting a surrogate model to imitate the behaviour of the classifier while facilitating the extraction of explanations. Often, the surrogate model is a simpler version of the original classifier. Global surrogates are aimed at replicating the behaviour of the classifier in its entirety. On the other hand, local surrogate models are

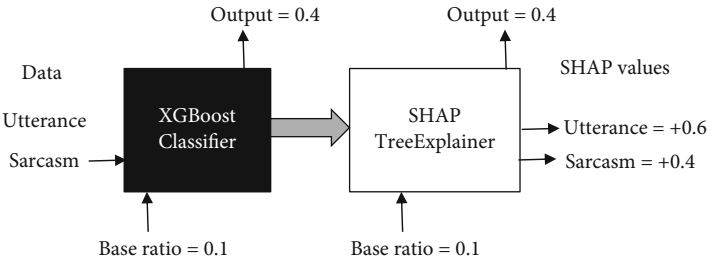


Figure 11: SHAP model architecture.

Table 2: Performance results using utterance + context.

Learning models	Accuracy	Precision	Recall	F-1 score
XGBoost	0. 931	0.965	0.887	0.924
Random forest	0.586	0.402	0.637	0.492
SVM [5]	—	0.579	0.545	0.541

Table 3: Performance results using only utterance.

Learning models	Accuracy	Precision	Recall	F-1 score
XGBoost	0.879	0.852	0.918	0.883
Random forest	0.547	0.369	0.579	0.405
SVM [5]	—	0.609	0.596	0.598

trained to focus on a specific part of the rationale of the trained classifier. In this research, we use two different post hoc local surrogate explainability methods, namely, Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive exPlanations (SHAP). The methods fundamentally differ in terms of the interpretability technique used to explain the working of the black-box model. Typically, LIME creates a new dataset or text from the original text by randomly removing words from the original test and gives the probability to each word to eventually predict based on the calculated probability. SHAP, on the other hand, does not create a separate dataset but uses Shapley values to explain the prediction of any input by computing the contribution of each feature for prediction.

4.2.1. LIME. LIME is available as an open-source Python package. It is a local surrogate approach that specifies the importance of each feature to an individual prediction. LIME does not work on the training data; in fact, it gives the prediction by testing it with variations of the data. It trains a linear model to approximate the local decision boundary for that instance, which then generates a new dataset consisting of all the permutation samples along with their corresponding predictions. New data is created by randomly removing words from the original data. The dataset is represented with binary features for each word. A feature is set to 1 if the corresponding word is included and 0 if it is not included. The new dataset of the LIME then trains the interpretable model, i.e., the RF model which is then weighted by the proximity of the sampled instances to the instance of interest. The learned model should be able to give the general idea of the machine learning model prediction locally, but it may not be a good global approximation. The generic steps of LIME include sampling of instances followed by training the surrogate model using these instances to finally generate the final explanation given to the user through a visual interface provided with the package. Mathematically, LIME explanations are determined using

$$\text{explanation}(x) = \text{argarg}\ \min\ g \in G\, L(f, g, \pi_x) + \Omega_g. \quad (1)$$

According to the mathematical formula, the explanation model for instance x is the ML model (random forest, in our case) which then minimises the loss L, such as mean square error (MSE). This L measures the closeness of the explanation to the prediction of the original model f, while keeping the model complexity $\Omega(g)$ low. G is the pool of possible explanation, and π_x is the proximity measure of how large the neighborhood is around the instance x. LIME optimizes only the loss part of the data.

The idea for training the LIME model is simple:

(i) Select the instance which the user wants to have explanation of the black-box prediction

(ii) Add a small noisy shift to the dataset and get the black-box prediction of these new points

(iii) Weight the new point samples according to the proximity of the instance x

(iv) Weighted, interpretable models are trained on the dataset with the variations

(v) With the interpretable local model, the prediction is explained

4.2.2. SHAP. SHAP is aimed at explaining individual explanations based on the cooperative game theory Shapley values. Shapley values are used for the prediction to be explained by the assumption of each feature value of the instance as a "player." These values tell the user how fairly the distribution is among the "players" in game. The Shapley value is the average marginal contribution of a feature value across all possible coalitions. The reason to choose SHAP as our second explainable model was because SHAP computes the contribution of each feature of the prediction. These features act as "players" which will then be used to see if the payoff of the distribution is fair or not. It needs to satisfy the local accuracy, missingness, and consistency properties making predictions [17].

Utterance +context

	Predicted 0	Predicted 1
Actual 0	TN = 309	FP = 36
Actual 1	FN = 11	TP = 334

Only utterance

	Predicted 0	Predicted 1
Actual 0	TN = 290	FP = 28
Actual 1	FN = 55	TP = 317

Figure 12: Confusion matrix of XGBoost on MUStARD dataset.

SHAP explains the output of the black-box model by showing the working of the model to explain the prediction of an instance computing each feature's contribution to the prediction. As given in (2), mathematically, SHAP specifies explanation of each prediction as it gives out the local accuracy of the represented features

$$g\left(z'\right) = \varnothing_0 + \sum_{j=1}^{M} \phi_j z'_j, \tag{2}$$

where g is the explanation model and $z' \epsilon \{0,1\}^M$ is the coalition vector in the dataset. M denotes the maximum size of the coalition in SHAP where entry 1 represents that the feature is present and 0 represents that the feature is absent.

SHAP basically follows three properties for the result, and those properties are as follows:

(i) *Local Accuracy*. Local accuracy means that the explanation model should match the original model as given in

$$f(x) = g\left(x'\right) = \phi_0 + \sum_{j=1}^{M} \phi_j x'_j \tag{3}$$

(ii) *Missingness*. Missing feature gets the attribution score of 0 where 0 represents the absence of the feature. It means that the simplified input feature and the original input feature should be the same so that it does not have any impact. It is given as shown in

$$x'_j = 0 \Longrightarrow \phi_j = 0 \tag{4}$$

(iii) *Consistency*. Consistency means that the values increase or remain the same according to the marginal contribution of the feature values of the model. It is given by

$$f'_x\left(z'\right) - f'_x\left(z'_j\right) \geq f_x\left(z'\right) - f_x\left(z'_j\right) \tag{5}$$

In the paper, the features which are used for the target prediction and the SHAP value for the contribution of that feature are the difference between the actual prediction and the mean prediction. SHAP provides both local and global interpretability by calculating SHAP values on the local level for feature importance and then providing a global feature importance by summing the absolute SHAP values for each of the individual predictions. The SHAP model architecture is shown in Figure 11.

We use KernelSHAP (https://docs.seldon.io/projects/alibi/en/stable/methods/KernelSHAP.html) in this work for the estimation of the instance x of each feature contribution. KernelSHAP uses weighted local linear regression to estimate the Shapley values for any model.

5. Results and Discussion

We implemented the model using scikit-learn, a framework in Python. The classification performance of XGBoost was evaluated using accuracy, F1 score, precision, and recall as metrics. The training : test split was 70 : 30. The model is trained with default parameters using the Python XGBoost package. The performance of XGBoost was compared with another ensemble learning method—random forest and superior results were observed using XGBoost. Also, the primary goal of this research was to investigate the role and importance of context we trained and tested the model with and without context. A comparison with the existing work [5] that uses support vector machines (SVM) as the primary baseline for sarcasm classification in speaker-independent textual modality is also done. The results obtained using the punch-line utterance and its associated context are shown in Table 2 whereas the results obtained using only the punch-line utterance that is without using context as a feature are shown in Table 3.

It is evident from the results that sarcastic intent of the thread is more efficiently captured using context, improving the accuracy by nearly 5%. The confusion matrix for the XGBoost classifier with and without context is shown in Figure 12. To compute the confusion matrices, we take a count of four values as follows:

(i) *True Positives* (TP): number of sarcastic utterance correctly identified

(ii) *False Positives* (FP): number of nonsarcastic utterance that was incorrectly identified as sarcastic utterance

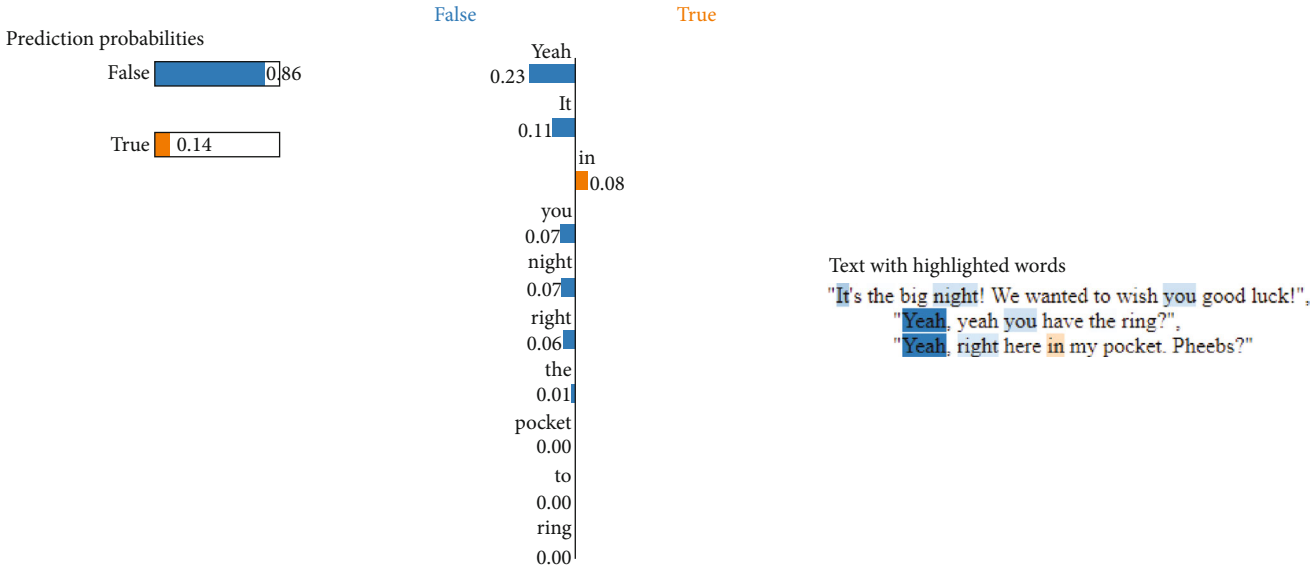


Figure 13: LIME visualisation.

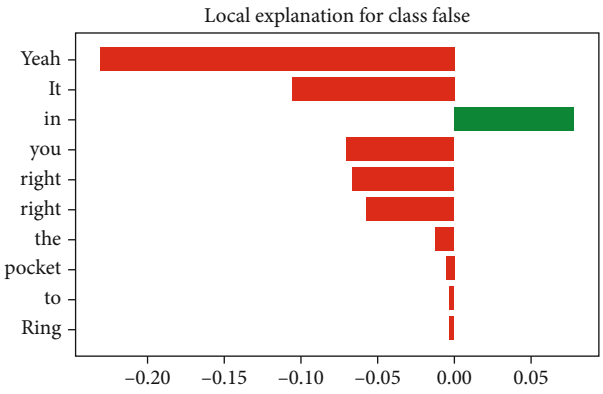


Figure 14: Local explanation for false class.

(iii) *False Negatives* (FN): number of sarcastic utterance that was incorrectly identified as nonsarcastic utterance

(iv) *True Negatives* (TN): number of nonsarcastic utterance correctly identified

The objective was not only to produce higher results but also to produce a better analysis. Therefore, after the evaluation of the learning algorithm, explainable models of LIME and SHAP were used for prediction interpretability. LIME text classifier and LIME text explainer were used to obtain the explanation model for LIME. The class names were set to true and false according to the label, for the LIME text explainer with random state of 42. For SHAP, it was trained and tested on the training and testing vectors generated by tf-idf vectors with 200 background samples to generate the force plot and summary plot of the XGBoost using utterance and context as features.

The explanation model for LIME and SHAP shows which words in the dialogues of the characters influence the model

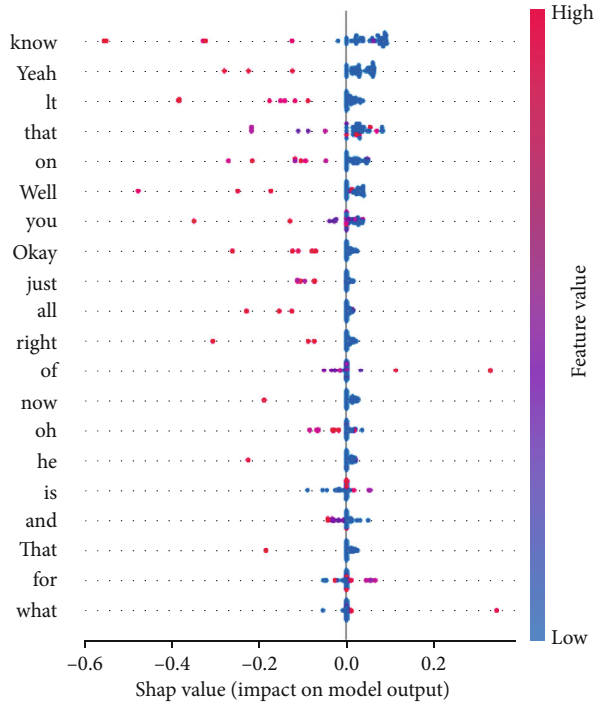


Figure 15: SHAP summary plot.

to label the utterance as sarcastic or not. The explainability scores from each of the methods are generated for every feature in the dataset. Evidently, for an utterance with sarcasm, certain words receive more importance than others. Figure 13 shows the LIME visualisation, where it can be observed that only some parts of the dialogue (taken arbitrarily) are being used to determine the probability of the sarcasm of the utterance by the speaker. As we randomly select an utterance in the test set, it happens to be an utterance that is labelled as nonsarcastic, and our model predicts it as

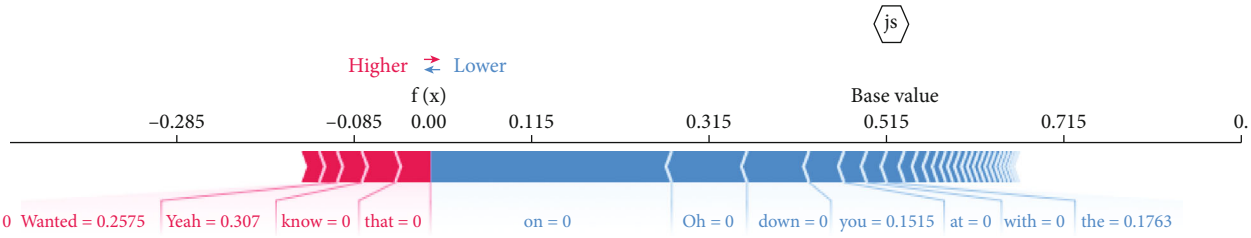


Figure 16: SHAP force plot.

nonsarcastic as well. Using this utterance, we generate explanations.

Noticeably, for this conversation, word "*Yeah*" has the highest negative score for class sarcasm and our model predicts this conversation should be labelled as nonsarcastic with the probability of 86%.

Figure 14 shows how the weights are trained, and the weights of each word given in the utterance are used to determine the sarcasm of the utterance by the speaker.

The same goes for the visualisation of the SHAP model as given in Figure 15, which helps the user understand how the model is making the decision for detecting sarcasm in dialogues. It is using each and every word as a "player" and giving the coalition of whether the model can equally pay off or not. This is a very helpful view that shows at a global level in which direction each feature contributes as compared to the average model prediction. The y-axis in the right side indicates the respective feature value being low vs. high. Each dot represents 1 instance in the data, and the cluster of dots indicates there are many instances in the data with that particular SHAP value.

Thus, the SHAP summary plot combines the feature effect with the importance of the feature. In the SHAP summary plot, each point is the Shapley value for a feature and an instance. The y-axis and x-axis in the summary plot show the feature and the Shapley values, respectively. The colors in the summary plot indicate the impact of the feature from high to low, and the overlapping points in the plot show the distribution of the Shapley values per feature.

Another way to understand the explainability of the utterance using SHAP can be done using the force plot of the data. A force plot helps visualising Shapley values for the features. Feature values in pink cause to increase the prediction. The size of the bar shows the magnitude of the feature's effect. Feature values in blue cause to decrease the prediction. Sum of all feature SHAP values explains why model prediction was different from the baseline. Figure 16 gives the multiprediction force plot used in the given instance with utterance and context for the analysis of the prediction path. Again, the word "*Yeah*" has higher feature importance.

The results support the hypothesis that how each word in the utterance with respect to the context of the dialogues is important for sarcasm detection.

6. Conclusion

With the accelerated use of sentiment technologies in online data streams, companies have integrated it as an enterprise solution for social listening. Sarcasm is one of the key NLP challenges to sentiment analysis accuracy. Context incongruity can be used to detect sarcasm in conversational threads and dialogues where the chronological statements formulate the context of the target utterance. We used an ensemble learning method to detect sarcasm in benchmark sitcom dialogue dataset. Results clearly establish the influence of using context with the punch-line utterance as features to train XGBoost. Further, the predictions given by the black-box XGBoost are explained using LIME and SHAP for local interpretations. These post hoc interpretability methods demonstrate that how few words unambiguously contribute to the decision and word importance is the key to accurate prediction of the sarcastic dialogues. As a future work, we would like to evaluate other XAI methods such as PDP for the detection of sarcasm. Also, temporal context and span analysis for context incongruity are another promising line of work. Gauging other rhetorical literary devices in online data streams is also an open domain of research. Auditory cues such as tone of the speaker and other acoustic markers such as voice pitch, frequency, empathetic stress and pauses, and visual cues for facial expressions that can assist sarcasm detection in audio-visual modalities need further investigation.

Data Availability

Publicly accessible data has been used by the authors.

Additional Points

Code Availability. Can be made available on request.

Ethical Approval

The work conducted is not plagiarized. No one has been harmed in this work.

Conflicts of Interest

The authors certify that there is no conflict of interest in the subject matter discussed in this manuscript.

Authors' Contributions

All the authors have equally contributed in the manuscript preparation. All the authors have given consent to submit the manuscript. The authors provide their consent for the publication.

References

[1] N. Majumder, S. Poria, H. Peng et al., "Sentiment and sarcasm classification with multitask learning," *IEEE Intelligent Systems*, vol. 34, no. 3, pp. 38–43, 2019.

[2] D. Jain, A. Kumar, and G. Garg, "Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature-rich CNN," *Applied Soft Computing*, vol. 91, article 106198, 2020.

[3] D. Ghosh, W. Guo, and S. Muresan, "Sarcastic or not: word embeddings to predict the literal or sarcastic meaning of words," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1003–1012, Lisbon, Portugal, 2015.

[4] A. Kumar and G. Garg, "Empirical study of shallow and deep learning models for sarcasm detection using context in benchmark datasets," *Journal of Ambient Intelligence and Humanized Computing*, 2019.

[5] S. Castro, D. Hazarika, V. Pérez-Rosas, R. Zimmermann, R. Mihalcea, and S. Poria, "Towards multimodal sarcasm detection (an _obviously_ perfect paper)," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 4619–4629, Florence, Italy, 2019.

[6] A. Kumar and G. Garg, "Sarc-m: sarcasm detection in typographic memes," in *International Conference on Advances in Engineering Science Management & Technology (ICAESMT)-2019*, Uttaranchal University, Dehradun, India, 2019.

[7] A. Kumar and G. Garg, "The multifaceted concept of context in sentiment analysis," in *Cognitive Informatics and Soft Computing*, P. Mallick, V. Balas, A. Bhoi, and G. S. Chae, Eds., vol. 1040 of Advances in Intelligent Systems and Computing, pp. 413–421, Springer, Singapore, 2020.

[8] D. Gunning, *Explainable artificial intelligence (XAI)*, Defense Advanced Research Projects Agency (DARPA), 2017, https://www.cc.gatech.edu/~alanwags/DLAI2016/(Gunning)%20IJCAI-16%20DLAI%20WS.pdf.

[9] B. Mahbooba, M. Timilsina, R. Sahal, and M. Serrano, "Explainable artificial intelligence (XAI) to enhance trust management in intrusion detection systems using decision tree model," *Complexity*, vol. 2021, Article ID 6634811, 11 pages, 2021.

[10] C. Li, X. Zheng, Z. Yang, and L. Kuang, "Predicting short-term electricity demand by combining the advantages of ARMA and XGboost in fog computing environment," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 5018053, 18 pages, 2018.

[11] F. K. Došilović, M. Brčić, and N. Hlupić, "Explainable artificial intelligence: a survey," in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 210–215, Opatija, Croatia, 2018.

[12] T. Miller, "Explanation in artificial intelligence: insights from the social sciences," *Artificial Intelligence*, vol. 267, pp. 1–38, 2019.

[13] G. Stiglic, P. Kocbek, N. Fijacko, M. Zitnik, K. Verbert, and L. Cilar, "Interpretability of machine learning-based prediction models in healthcare," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 5, article e1379, 2020.

[14] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine learning interpretability: a survey on methods and metrics," *Electronics*, vol. 8, no. 8, p. 832, 2019.

[15] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why should I trust you?": explaining the predictions of any classifier," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pp. 1135–1144, San Diego, California, 2016.

[16] C. Meske and E. Bunde, "Transparency and trust in human-AI-interaction: the role of model-agnostic explanations in computer vision-based decision support," in *International Conference on Human-Computer Interaction, HCII 2020*, H. Degen and L. Reinerman-Jones, Eds., vol. 12217 of Lecture Notes in Computer Science, pp. 54–69, Springer, Cham, 2020.

[17] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," 2017, https://arxiv.org/abs/1705.07874.

[18] A. Messalas, Y. Kanellopoulos, and C. Makris, "Model-agnostic interpretability with shapley values," in *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pp. 1–7, Patras, Greece, 2019.

[19] A. Adadi and M. Berrada, "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)," *IEEE access*, vol. 6, pp. 52138–52160, 2018.

[20] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal, "Explaining explanations: an overview of interpretability of machine learning," in *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 80–89, Turin, Italy, 2018.

[21] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM Computing Surveys*, vol. 51, no. 5, pp. 1–42, 2019.

[22] E. Riloff, A. Qadir, P. Surve, L. De Silva, N. Gilbert, and R. Huang, "Sarcasm as contrast between a positive sentiment and negative situation," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 704–714, Seattle, Washington, USA, 2013.

[23] A. Joshi, V. Sharma, and P. Bhattacharyya, "Harnessing context incongruity for sarcasm detection," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pp. 757–762, Beijing, China, 2015.

[24] J. Tepperman, D. Traum, and S. Narayanan, ""Yeah right": sarcasm recognition for spoken dialogue systems," in *Ninth international conference on spoken language processing*, Pittsburgh, Pennsylvania, 2006https://www.isca-speech.org/archive/archive_papers/interspeech_2006/i06_1821.pdf.

[25] R. Kreuz and G. Caucci, "Lexical influences on the perception of sarcasm," in *Proceedings of the Workshop on Computational Approaches to Figurative Language - FigLanguages '07*, pp. 1–4, Rochester, New York, 2007.

[26] O. Tsur, D. Davidov, and A. Rappoport, "ICWSM—a great catchy name: semi-supervised recognition of sarcastic sentences in online product reviews," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 4, no. 1, 2010.

[27] D. Davidov, O. Tsur, and A. Rappoport, "Semi-supervised recognition of sarcasm in Twitter and Amazon," in *Proceedings of the fourteenth conference on computational natural language learning*, pp. 107–116, Uppsala, Sweden, 2010.

[28] A. Reyes, P. Rosso, and T. Veale, "A multidimensional approach for detecting irony in Twitter," *Language Resources and Evaluation*, vol. 47, no. 1, pp. 239–268, 2013.

[29] D. Bamman and N. Smith, "Contextualized sarcasm detection on Twitter," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 9, Oxford, UK, 2015.

[30] S. Mukherjee and P. K. Bala, "Sarcasm detection in microblogs using Naïve Bayes and fuzzy clustering," *Technology in Society*, vol. 48, pp. 19–27, 2017.

[31] A. G. Prasad, S. Sanjana, S. M. Bhat, and B. S. Harish, "Sentiment analysis for sarcasm detection on streaming short text data," in *2017 2nd International Conference on Knowledge Engineering and Applications (ICKEA)*, pp. 1–5, London, UK, 2017.

[32] Y. Ren, D. Ji, and H. Ren, "Context-augmented convolutional neural networks for Twitter sarcasm detection," *Neurocomputing*, vol. 308, pp. 1–7, 2018.

[33] A. Kumar and G. Garg, "Sarcasm detection using feature-variant learning models," *Proceedings of ICETIT 2019*, P. Singh, B. Panigrahi, N. Suryadevara, S. Sharma, and A. Singh, Eds., , pp. 683–693, Springer, Champ, 2020.

[34] A. Kumar, S. R. Sangwan, A. Arora, A. Nayyar, and M. Abdel-Basset, "Sarcasm detection using soft attention-based bidirectional long short-term memory model with convolution network," *IEEE Access*, vol. 7, pp. 23319–23328, 2019.

[35] 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*, pp. 2506–2515, Florence, Italy, 2019.

[36] A. Kumar, V. T. Narapareddy, V. Aditya Srikanth, A. Malapati, and L. B. M. Neti, "Sarcasm detection using multi-head attention based bidirectional LSTM," *IEEE Access*, vol. 8, pp. 6388–6397, 2020.

[37] I. A. Farha and W. Magdy, "Benchmarking transformer-based language models for Arabic sentiment and sarcasm detection," in *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pp. 21–31, Kyiv, Ukraine (Virtual), 2021.

[38] D. Faraj and M. Abdullah, "Sarcasmdet at sarcasm detection task 2021 in arabic using arabert pretrained model," in *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pp. 345–350, Kyiv, Ukraine (Virtual), 2021.

[39] D. Ghosh, A. R. Fabbri, and S. Muresan, "The role of conversation context for sarcasm detection in online interactions," in *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 186–196, Saarbrücken, Germany, 2017.

[40] D. Hazarika, S. Poria, S. Gorantla, E. Cambria, R. Zimmermann, and R. Mihalcea, "Cascade: contextual sarcasm detection in online discussion forums," in *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 1837–1848, Santa Fe, New Mexico, USA, 2018.

[41] A. Baruah, K. Das, F. Barbhuiya, and K. Dey, "Context-aware sarcasm detection using BERT," in *Proceedings of the Second Workshop on Figurative Language Processing*, pp. 83–87, Online, 2020.

[42] A. K. Jena, A. Sinha, and R. Agarwal, "C-net: contextual network for sarcasm detection," in *Proceedings of the Second Workshop on Figurative Language Processing*, pp. 61–66, Online, 2020.

[43] Y. Zhang, Y. Liu, Q. Li et al., "CFN: a complex-valued fuzzy network for sarcasm detection in conversations," *IEEE Transactions on Fuzzy Systems*, p. 1, 2021.

[44] Y. Tay, L. A. Tuan, S. C. Hui, and J. Su, "Reasoning with sarcasm by reading in-between," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1010–1020, Melbourne, Australia, July 2018.

[45] R. Akula and I. Garibay, "Explainable detection of sarcasm in social media," in *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 34–39, Washington D.C., 2021.

[46] J. Ramos, "Using tf-idf to determine word relevance in document queries," *Proceedings of the First Instructional Conference on Machine Learning*, vol. 242, no. 1, pp. 29–48, 2003.

Biomass Conversion and Biorefinery
https://doi.org/10.1007/s13399-021-01701-9

ORIGINAL ARTICLE

Fabrication of activated carbon electrodes derived from peanut shell for high-performance supercapacitors

Lokesh Pandey[1] · Subhajit Sarkar[2] · Anil Arya[3] · A. L. Sharma[3] · Amrish Panwar[4] · R. K. Kotnala[5] · Anurag Gaur[2,6]

Received: 18 March 2021 / Revised: 16 June 2021 / Accepted: 26 June 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract
In this work, the activated carbon (AC) is derived from the waste material of peanut shells. The obtained activated carbon has been synthesized using the chemical activation process via impregnation with NaOH at various proportions (1:2 and 1:3). The structural, morphological, and electrochemical properties of obtained AC are examined by X-ray diffraction (XRD), BET surface analysis, scanning electron microscopy (SEM), and electrochemical techniques. The resulting AC shows the enhancement in specific surface area (SSA) from 584 to 735 m^2/g at 1:2 ratio and 826 m^2/g at 1:3 ratio after the activation process. The electrodes are fabricated using derived activated carbon for electrochemical characterization. It has been observed through cyclic voltammetry and galvanostatic charge–discharge measurements that 1:3 NaOH-activated samples reveal the maximum value of specific capacitance (263 F/g at 10 mV/s and 290 F/gat 0.2A/g). The sample also conveys the pre-dominant ~ 98% of efficiency from 400 to 1000th cycle with fast and almost constant capacitance. The electrochemical impedance spectroscopy (EIS) result suggests the pure capacitive nature of the sample and indicates the dominating electric double-layer capacitive (EDLC) behavior. The obtained results demonstrate that the peanut shell–derived activated carbon is a potential candidate as an electrode material for supercapacitors.

Keywords Peanut shell · Activated carbon · EDLC · Electrochemical characterization

1 Introduction

Energy is essential for the life of human beings. At present, the energy demand is being met by traditional (non-renewable) sources of energy such as fossil fuels, coal. But they are on the verge of depletion and also producing harmful gases that result in global warming. Hence, to counter the energy demand, alternate sources of energy need to be developed for the protection of nature and natural systems. The utmost feasible alternates are renewable sources of energy like solar energy and wind energy. The energy harvested from these sources needs to be stored in any energy storage device (battery and supercapacitor) and that device can provide this energy as per need or shortage of supply. The battery has high energy density, and the supercapacitor (SC) has high power density. The SC has grabbed the attention of researchers owing to its high power density, long cyclic stability, and broad operating range. The only drawback is the low energy density, which is the locus of research nowadays [1, 2].

The supercapacitor has three components: electrode, electrolyte, and separator. The electrode is a crucial component that impacts the electrochemical properties. The important electrode materials that have been explored are carbon materials, metal oxide, and their composites. Out of the carbon material, the activated carbon is a suitable candidate due to its high surface area, porous nature, ease of availability, better stability, a high degree of surface activity, and low cost and abundance in nature.

✉ Anurag Gaur
anuragdph@gmail.com

[1] Uttrakhand Technical University, Dehradun, Uttarakhand, India

[2] Department of Physics, National Institute of Technology, Kurukshetra, India

[3] Department of Physics, Central University of Punjab, Bathinda, Punjab, India

[4] Department of Applied Physics, Delhi Technological University, New Delhi, India

[5] CSIR, National Physical Laboratory, Dr. K.S. Krishnan Road, New Delhi, India

[6] Department of Physics, J.C Bose University of Science & Technology, Faridabad, India

Published online: 06 July 2021

The biomass derived from agricultural sources has been found as the source of energy, and a lot of research work has been performed to derive the carbon from agricultural waste material in recent time. Various efforts have been made to obtain the activated carbon from biomass such as sunflower seed shells, sugarcane-bagasse, wheat straw/husks, and peanut shells [3]. Peanut is a legume crop and is widely grown in tropics and subtropics. The peanut shell is an industrial waste product, and the global annual volume of peanut shells is about 7.44 million tons [4]. Therefore, it needs to be used efficiently to alleviate the environmental burden. Activated carbon derived from bio-waste has the potential to store the charge, and peanut shells can be used to derive activated carbon for use as an electrode material [5–10]. Moreover, it has been reported that peanut shells are having a different chemical composition, material morphologies, surface functionalities, and carbon properties, which attracted great attention to utilizing it as the precursor to produce the activated carbon [11–22]. Because of various characteristic features of activated carbon such as low cost, high surface area, and porosity, a lot of efforts have been made by the research community to use this material as an electrode for supercapacitors.

To further enhance the charge storage characteristics of derived carbon, it is activated through different chemicals as the activating agent impacts the electrochemical performances. The important chemical activating agents are $ZnCl_2$, KOH, K_2CO_3, H_2SO_4, Na_2Co_3, NaOH, etc. [23–25]. Natalia et al. [23] prepared the activated carbon using KOH as an agent, and it exhibits a surface area close to 250 m^2/g with a pore size of 2.3 nm. The synthesized electrodes demonstrate a specific capacitance of 64 F/g (at 5 mV/s). The effect of temperature and impregnation ratio on activated carbon was examined by Yahya et al. [25]. Activated carbon was derived from peanut shells using KOH as an activating agent by Wu et al. [26] and exhibits an energy density of 8.38 Wh/kg and good stability for 1000 cycles. Activated carbon derived from peanut shells and rice husk was examined by Xiaojun et al. [27]. Kim et al. [28] synthesized the activated carbon from biomass using KOH as an activating agent. In the previous work, our group also derived activated carbon from sugarcane bagasse using K_2CO_3, KOH, and NaOH as activating agents. The results showed that the carbon produced through impregnation with K_2CO_3 has a specific surface area of 725 m^2/g and has a specific capacitance of 265 F/g [29].

Herein, we report the extraction of activated carbon from the most abundant peanut shell waste material via the chemical activation method. The effect of the activating agent (NaOH) at different impregnation ratios (1:2 and 1:3) has been investigated on the structural and morphological properties. Finally, the charge storage performance of activated carbon has been explored to be used as an electrode material in supercapacitors.

2 Experimental details

2.1 Preparation of activated carbon

Firstly, the peanut shell waste material was washed with deionized water to remove any dust particles followed by drying in an oven at 100 °C for 12 h. The dried materials are crushed and ground using a mortar-pestle. The ground peanut shell material was carbonized at 450 °C for 1 h in a muffle furnace (heating rate = 10 °C/min) to obtain a fine powder. The obtained black carbon powder was chemically activated by NaOH as an activating agent at 1:3 and 1:2 impregnation ratios, and these samples are referred to as AC1 and AC2, respectively. The non-activated sample is referred to as NAC. The impregnated samples were stirred properly and kept for 24 h at room temperature for better mixing. Then, the samples were dried at 120 °C for 12 h in a hot air oven. Finally, the prepared samples were kept at 650 °C for 2 h to complete the activation process. The obtained powder material was characterized using a BET surface area analyzer, an X-ray diffractometer, a scanning electron microscope, and a potentiostat/galvanostat electrochemical analyzer.

2.2 Fabrication of carbon electrode

The working electrode was prepared by 80 wt% of derived AC sample mixed with 10 wt% of acetylene black (AB) and 10 wt% of PVDF (polyvinylidene fluoride) using a pestle-mortar until the homogenous powder was obtained. The mixture was added with 1 ml of ethanol, and sonication was done for 1 h to disperse the fine particles. Then the homogenous slurry was prepared along with constant stirring, which was deposited onto a nickel foil using the drop-casting method. The electrodes were dried at 100 °C in a hot air oven overnight to evaporate the solution, which as a result assembles strong mechanical strength over the foil. The Ni foil with activated carbon coat (working electrode), platinum electrode (counter electrode), and saturated calomel electrode (reference electrode) were chosen as electrodes. The working electrode contains 1 mg of measured activated material with a geometric surface area of about (1×1) cm^2 and thickness of 10 μm as shown in Fig. 1.

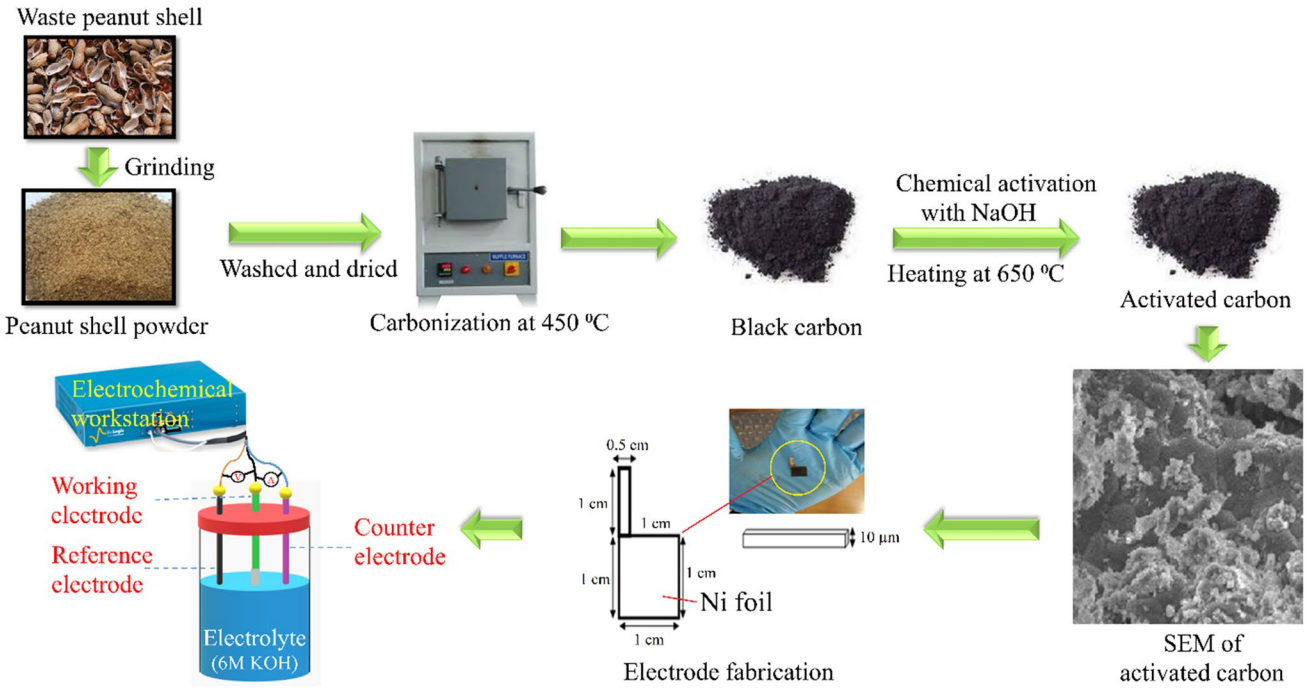


Fig. 1 Activated carbon preparation and fabrication of a working electrode

2.3 Electrochemical measurements

The electrochemical properties of the prepared electrodes (NAC, AC1, AC2) were examined in a three-electrode cell configuration. The cyclic voltammetry (CV), complex impedance spectroscopy (CIS), and galvanostatic charge–discharge (GCD) measurements were performed with a potentiostat/galvanostat electrochemical analyzer (Biologic SP-240). The electrochemical performance was examined in 6 M KOH electrolyte at room temperature. The CV measurement was performed in a voltage window of − 0.2 to + 0.8 V at different scan rates (10 to 150 mV/s). The specific capacitance (C_{sp}) was calculated using Eq. (1):

$$C_{sp}(F/g) = \frac{\int idV}{2 \times m \times s \times \Delta V} \tag{1}$$

where $\int$ idV is the integrated area of the CV curve, m (g) is the mass of AC, s is the scan rate (mV/s), and ΔV(V) is the voltage window.

The EIS measurements were performed from 10 mHz to 1 MHz frequency range. GCD measurements were performed at different current densities (0.2 to 4 A/g), and specific capacitance was calculated using the below equation:

$$C_{sp}(F/g) = \frac{i \times \Delta t}{m \times \Delta V} \tag{2}$$

where i is the discharge current (A), Δt is the discharge time (s), m is the mass of active material (g), and ΔV is the voltage window (V).

The Coulombic efficiency is calculated using the following equation:

$$\eta(\%) = \frac{t_d}{t_c} \times 100\% \tag{3}$$

where t_d and t_c are the discharging and charging times, respectively.

3 Results and discussion

3.1 Structural and morphological analysis

The surface area, pore diameter, and pore size distribution have been examined from the nitrogen adsorption–desorption process. The isotherm plots of AC1, AC2, and NAC samples are shown in Fig. 2a. According to the IUPAC classification, NAC exhibits type-I isotherm curve and the horizontal adsorption plateau, and no hysteresis region suggests that micropores are dominating in the structure. The presence of a small hysteresis region and little divergence from the horizontal plane is well-observed in AC1 and AC2 samples, which further suggests the presence of mesopores.

Table 1 outlines the values of average pore diameter and surface area of AC1, AC2, and NAC samples. The

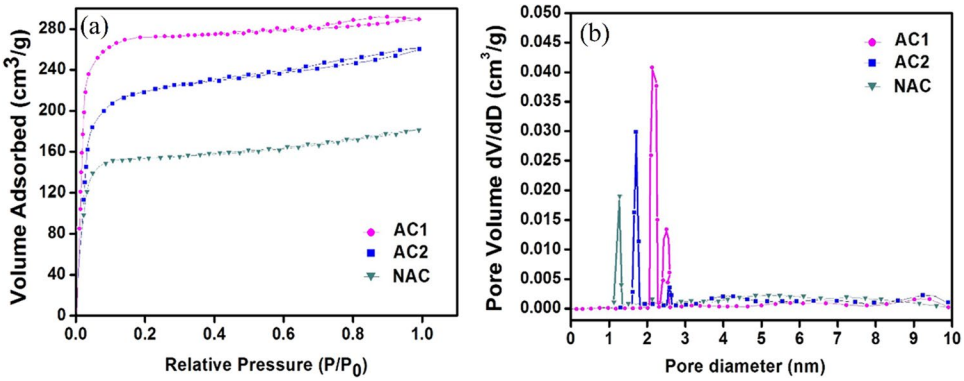


Fig. 2 **a** N_2 adsorption–desorption isotherm, and (**b**) pore size distribution

Table1 Pore diameter (nm) and BET surface area (m^2/g) for samples

Carbon samples	Pore diameter (nm)	BET surface area (m^2/g)
1: 3 NaOH (AC1)	2.15	826
1:2 NaOH (AC2)	1.85	735
Non-activated (NAC)	1.32	584

enhancement in surface area as shown in Table 1 suggests that the activation process is an effective strategy to improve electrode charge storage capability.

The BET surface results show that the surface area of non-activated carbon (NAC) is 584 m^2/g and has increased to 735 and 826 m^2/g for the activated carbons AC2 and AC1, respectively. The chemical treatment of NaOH at 1:2 impregnation ratio results in a surface area of about 735 m^2/g with a pore size of 1.8 nm (Fig. 2b). In addition to this, when the impregnation ratio further increased, the NaOH etching results in the enlargement of pore size and has created a majority pore size of 2.1 nm and minor pore size of 2.4 nm. This porous nature is due to the effect of dehydration during the carbonization process, which has contributed to the pore drilling and pore widening. Hence, the chemical treatment of NaOH at 1:3 impregnation demonstrates the highest surface area and pore diameter.

Figure 3a–c displays the SEM micrographs of NAC, AC1, and AC2 samples (at 10 µm). The micrographs evidence the significant changes in morphology after activation at different impregnation ratios. Figure 3a shows the NAC sample in which the carbon particles are scattered, whereas in Fig. 3b–c, particles have disintegrated into larger cavities leading to an enhanced surface area for activated AC1 and AC2 samples.

The intercalation of ions into the carbon layers resulted in disruption of lamellae structural arrangement during the activation process, and it results in large pore formation. The change in impregnation ratio from 1:2 to 1:3 has increased the specific surface area (SSA) of AC1 to 826 m^2/g from 735 m^2/g of AC2 and resulted in new pore formation, which is supported by the BET surface area analysis. The high surface area results in the contact area between electrode and electrolyte within a cell assembly.

Figure 4 shows the XRD patterns of non-activated carbon (NAC), 1:3 NaOH (AC1), and 1:2 NaOH (AC2) samples. The XRD data show that a peak around 26.50° (002) for the NAC sample corresponds to graphitic nature as confirmed through JCPDS card no: 02–0456. However, this peak shifts towards a higher angle at 27.34° for AC2 and 27.36° for AC1 activated samples and depicts the formation of more

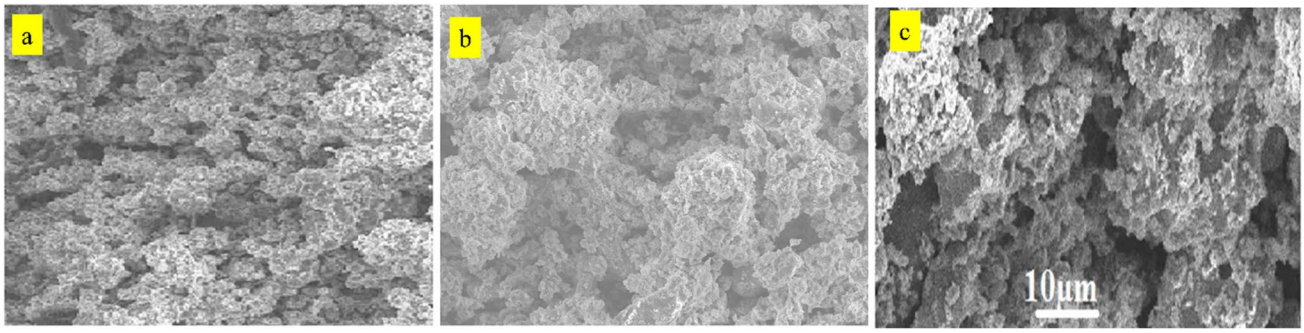


Fig. 3 SEM micrographs for (**a**) non activated carbon (NAC), (**b**) AC1, and (**c**) AC2

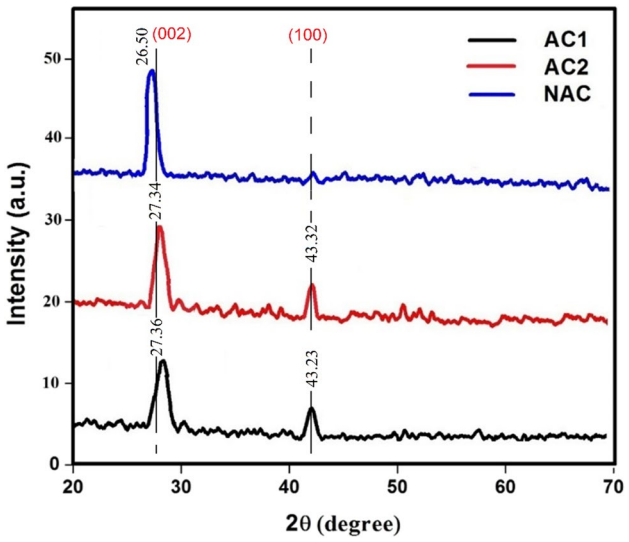


Fig. 4 XRD diffractogram for different samples (NAC, AC1, AC2)

disordered graphitic carbon. A small broad peak corresponding to (100) plane that was located near 43.23° for NaOH-activated samples indicates the presence of graphitic carbon with the amorphous character [30]. The graphitic form of carbon favors the fast electrochemical process by lowering the resistance. Moreover, few small peaks were observed in AC1 and AC2 samples due to the insertion of metallic elements into carbon layers after the mixing of NaOH during the impregnation process. The presence of small peaks indicates the formation of other compounds of lower intensities, which leads to the decrease of carbon content in the sample.

Thus, impregnation ratios have a huge impact on the degree of porosity level that determines how well the ions of electrolytes could get adsorbed on the membranous structure due to the release of volatile matters from the surface [31]. This fact is even supported by the SEM images and N_2 adsorption analysis.

3.2 Electrochemical analysis

A comparative CV for the different samples (AC1, AC2, and NAC) at a scan rate of 10 mV/s is shown in Fig. 5a. It may be observed that the surface area is maximum for the 1:3 NaOH (AC1) sample. The curve shows a clear indication of the change in the loop area as well as specific capacitance. The specific capacitance increases from 76 F/g for the NAC sample to 212 F/g (for AC2) and 263 F/g (for AC1). The high specific capacitance is being exhibited by the AC1 sample and is about 263 F/g. The enhancement in capacitance after activation is attributed to the large pore formation and enhanced surface area for impregnation ratio (1:3 NaOH).

The CV performance for the 1:3 NaOH (AC1) activated carbon electrode is shown in Fig. 5b. With the increase in

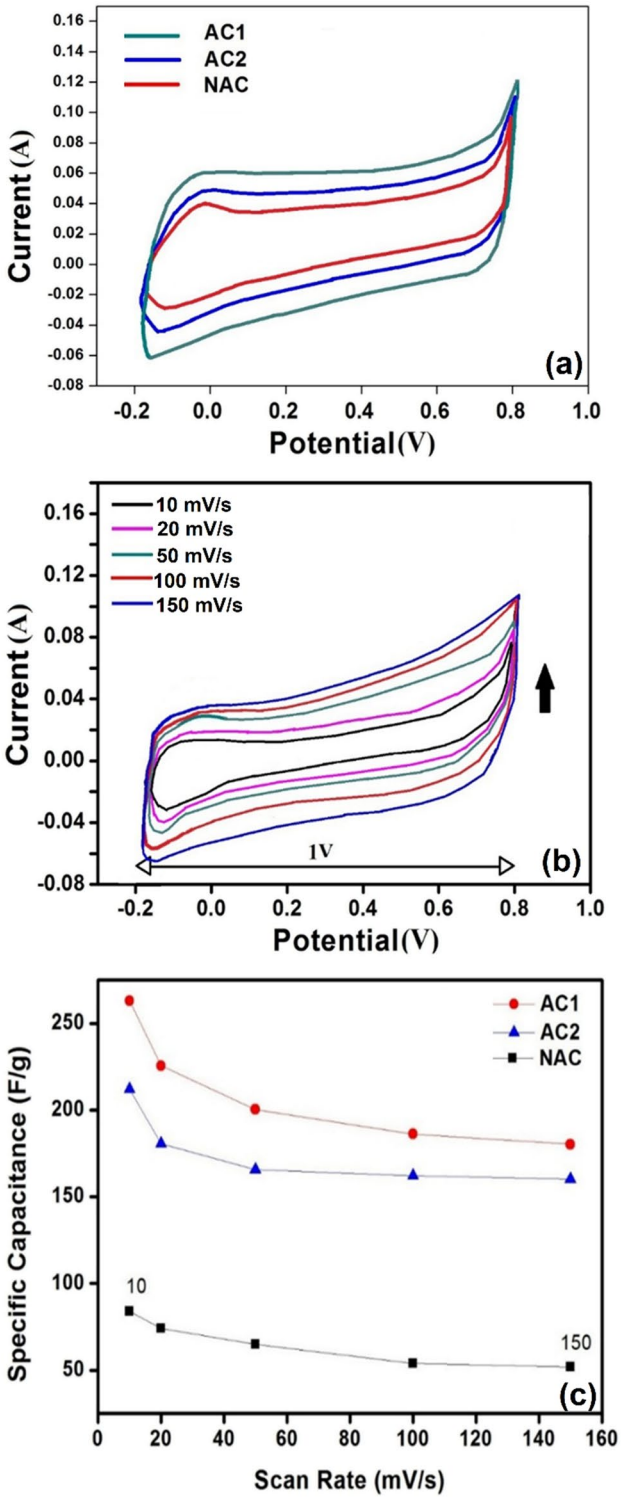


Fig. 5 **a** CV curve for non-activated carbon (NAC, AC2, and AC1 at a scan rate of 10 mV/s), (**b**) cyclic voltammetry for the AC1 sample at different scan rates, and (**c**) plot of specific capacitance against scan rate for NAC, AC1, and AC2 (saturated calomel used as a reference electrode)

Table 2 Specific capacitance, efficiency, and capacity retention for different samples

Carbon samples	Chemical agent	CV (F/g) at 10 mV/s	GCD (F/g) at 0.2 A/g	Efficiency (%)	Capacity retention (after 1000th cycle)
AC1	1:3 NaOH	263	290	98	90
AC2	1:2 NaOH	212	228	97	91
NAC	-	76	84	95	86

scan rate, the increase in area under the curve is observed. The consistent symmetric rectangular shape over a particular voltage range indicates the preservation of ions during adsorption and desorption, which shows better capacitive behavior and stability [32]. Figure 5c depicts the decrease of specific capacitance with the increase of scan rate for all the samples, but still maintains a high value of capacitance, indicating the stability of the electrode material. This is possibly due to the fast diffusion of ions on the surface of the electrode with the increase of the scan rate. Now, the mobility of the ions increases, and all pores may not be equally accessed which relates to the decrease of capacitance from 263 to 168 at 10 mV/s and 150 mV/s respectively for the AC1 sample.

Figure 6a shows the galvanostatic charge–discharge (GCD) plots for NAC, AC2, and AC1 samples. This Fig. 6a depicts a better duration of discharge time with almost good

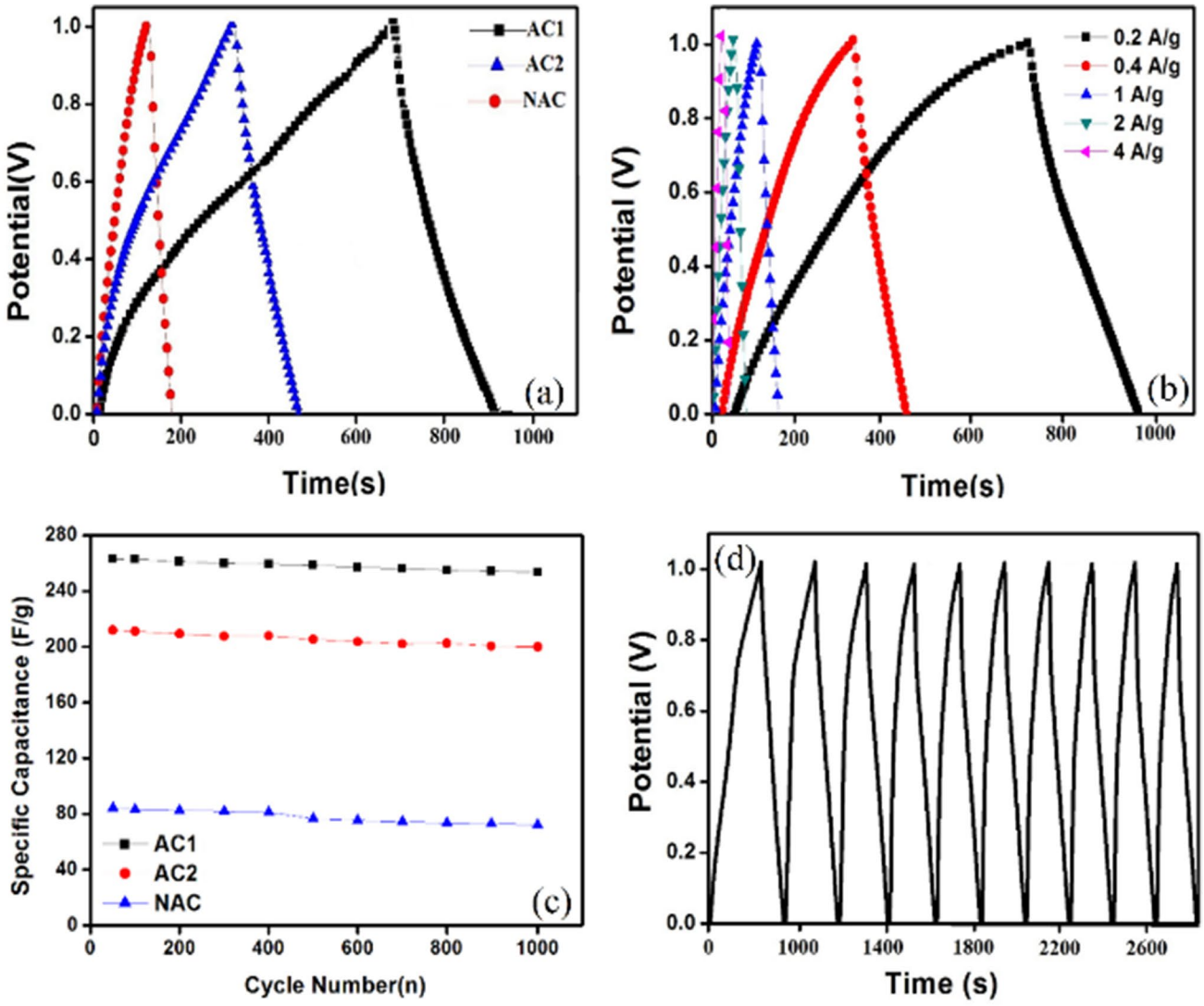


Fig. 6 (**a**) Comparative GCD plot for NAC, AC2, and AC1 at 0.2 A/g; (**b**) GCD plot for AC1 for 0.2 to 4 A/g current densities; (**c**) cyclic stability for 1000 cycles at 0.2 A/g current density for AC1, AC2, and NAC samples; (**d**) durability test at 0.2 A/g for AC1 sample

linearity and symmetry of the curve for activated AC1 and AC2 samples compared to that for NAC. The values of specific capacitance are found to be 290, 228, and 84 F/g at 0.2 A/g current density for AC1, AC2, and NAC samples, respectively. It is also found from this graph that the AC1 has high specific capacitance, better charge–discharge efficiency, higher discharge time, and high capacity retention as a result of better double-layer capacitive behavior compared to AC2 and NAC samples as demonstrated in Table 2.

Figure 6b shows the GCD measurement for AC1 sample at different current densities (0.2 to 4 A/g). The specific capacitance for AC1 sample decreases from 290 to 180 F/g with an increase in current density from 0.2 to 4 A/g. This decrease is due to a small discharge time interval because of the faster accumulation of ions in the carbon matrix at higher current density, which results in a decrease of specific capacitance. Furthermore, it is also suggested that ions were unable to accumulate uniformly during the double-layer formation process because of the non-uniform porous structures and pore distribution that was analyzed in Fig. 2b, due to the decaying of available electrolyte ions after a repeated number of cycles [33].

As shown in Fig. 6c, the stability test was performed for 1000 cycles at the current density of 0.2 A/g for AC1, AC2, and NAC samples. The durability and cycling stability imply that the number of times the electrolyte ions could migrate into the available pores for the repeated number of cycles. Figure 6c depicts excellent stability up to 1000th cycle; however, a slight degradation appeared after the 800th cycle which may be due to the increase of internal resistance. The specific capacitance of 263F/g was recorded at 100th cycle and was well-maintained until the end of the 800th cycle with a decrement of 261 F/g up to the 1000th cycle. However, a slight decrement was also observed from 212 to 207 F/g for the AC2 sample and 76 to 72 F/g for the NAC sample from 100th cycle to 1000th cycles. Figure 6d shows the durability test using GCD analysis of the AC1 sample at 0.2 A/g current density for the first 10 cycles.

Figure 7 shows the Coulomb efficiency with cycle number of AC1, AC2, and NAC samples. It is identified from the figure that the efficiency increases sharply and remains constant up to 800 cycles then degraded until 1000 cycles for all the samples. The values of Coulombic efficiencies are 95, 97, and 98% for NAC, AC2, and AC1 samples, respectively. The high and constant efficiency value of AC1 sample may be due to the retention of ions inside the mesopores for a longer period as affirmed from the longer discharge time of GCD measurement with comparatively less overall impedance as confirmed from EIS curves from Fig. 8.

The Nyquist plot for AC1, AC2, and NAC samples is shown in Fig. 8. Figure 8 suggests that the activation process has enhanced the capacitive behavior due to the increase in surface area as surveyed in Table 1. Hence, the hydrated ions

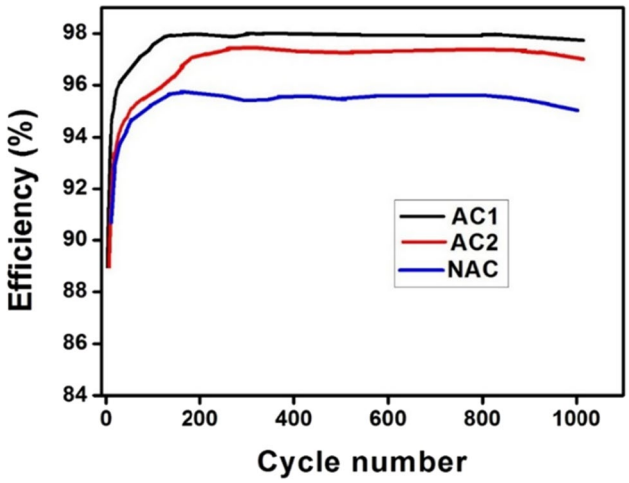


Fig. 7 Efficiency plot for AC1, AC2, and NAC samples

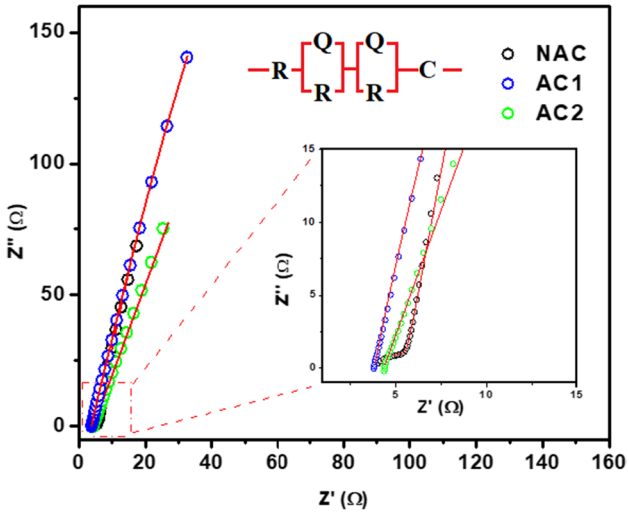


Fig. 8 Nyquist plot for AC, AC2, and NAC samples

could now more easily diffuse inside the bigger pore size with less pore resistance [34, 35]. The straight line appeared almost vertically close to 90° at a lower frequency region for activated samples (AC1 and AC2), which further suggests the pure capacitive nature of AC1 and AC2 samples and indicating the dominating EDLC behavior [36, 37]. The inset in Fig. 8 shows the magnified view for all samples. The AC1 sample shows the lowest overall resistance compared to AC2 and NAC sample; however, there is an enlargement of pore diameter of 2.1 nm from 1.8 nm and 1.3 nm with the increase in impregnation ratio. A smaller value of resistance suggests the faster ion diffusion owing to a high electronic conductivity [38–40].

For better analysis, the Nyquist plot has been fitted with an equivalent electrical circuit, and the fitted plot is in perfect agreement with experimental data in the whole

Table 3 Comparison of electrochemical characteristics of carbons derived from biomass

Material	Specific capacitance (F/g)	Scan rate/current density	Electrolyte	Cell (3E/2E)	Ref
Auricularia fungi	196	5 mV/s	6 M KOH	3E	[41]
Waste coffee grounds	121	0.5 A/g	BMIM BF_4/AN	2E	[42]
Bamboo	146	0.2 A/g	EMIM TFSI	2E	[43]
Chitosan	291	2 mV/s	6 M KOH	3E	[44]
	135	2 mV/s	$TEABF_4$	2E	
Cigar filter	154	1 A/g	6 M KOH	3E	[45]
Corn starch	162	0.625 A/g	6 M KOH	3E	[46]
Rice husk	112	1 A/g	1 M Na_2SO_4	2E	[47]
Willow catkin	249	0.5 A/g	6 M KOH	3E	[48]
Onion peel	127	0.75 A/g	1 M H_2SO_4	3E	[49]
Sugarcane bagasse	241	10 mV/s	6 M KOH	3E	[29]
	265	0.2 A/g		3E	
Palm tree decorated with α-MnO2	259	10 mV/s	1 M NaCl	3E	[50]
Silver nanospheres decorated pomegranate peel–derived AC	180	50 mV/s	-	3E	[51]
Peanut shell	**263**	10 mV/s	6 M KOH	3E	**This work**
	290	0.2 A/g		3E	**This work**

frequency. The solid red line shows the fitted plot, and the fitted circuit is shown in the inset. The low value of resistance suggests the enhanced capacitive behavior due to the mesoporous size and more inclination towards the imaginary axis compared to AC2 and NAC samples. For better analysis, the electrochemical performance of the present work is compared in Table 3. It may be concluded that the activated carbon derived from peanut shells (shown in bold) demonstrates superior performance compared to other bio-waste-derived carbon electrodes.

4 Conclusions

In summary, we have successfully derived activated carbon from peanut shell waste material via a chemical activation method using NaOH as an activating agent at 1:2 and 1:3 impregnation ratios. The BET and SEM analyses suggest that there is an enhancement of surface area from 584 to 826 m^2/g with an increase of impregnation ratio. This study shows that the impregnation ratio has a huge impact on the surface area of activated carbon through the chemical activation process. Furthermore, electrochemical measurements indicate an increase of specific capacitance with an excellent efficiency of 98% up to 1000 cycles. The obtained specific capacitance of 263 F/g at 10 mV/s and 290 F/g at 0.2 A/g for 1:3 NaOH-activated sample is significant. Results indicate that the activated carbon derived from peanut shell waste material could mitigate the energy demand as an electrode material for supercapacitors.

Funding Anurag Gaur received funding from the Science and Engineering Research Board (SERB), Govt. of India, through research project No: CRG/2018/001694.

References

1. Abioye AM, Ani FN (2015) Recent development in the production of activated carbon electrodes from agricultural waste biomass for supercapacitors: a review. Renew Sust Energ Rev 52:1282–1293
2. Tanwar S, Arya A, Gaur A, Sharma AL (2021). Transition Metal Dichalcogenide (TMDs) electrodes for Supercapacitors: A Comprehensive Review. J Phys Condens Matter. https://doi.org/10.1088/1361-648X/abfb3c
3. Duan F, Zhang JP, Chyang CS, Wang YJ, Tso J (2014) Combustion of crushed and pelletized peanut shells in a pilot-scale fluidized-bed combustor with flue gas recirculation. Fuel Process Technol 128:28–35
4. Guo PZ, Ji QQ, Zhang LL, Zhao SY, Zhao XS (2011) Preparation and characterization of peanut shell-based microporous carbons as electrode materials for supercapacitors. Acta Phys -Chim Sin 27(12):2836–2840
5. Subramanian V, Luo C, Stephan AM, Nahm KS, Thomas S, Wei B (2007) Supercapacitors from activated carbon derived from banana fibers. J Phys Chem C 111:7527–7531
6. Hassan AF, Youssef AM (2014) Preparation and characterization of microporous NaOH-activatedcarbonsfrom hydrofluoric acid leached rice husk and its application for lead (II) adsorption. Carbon Letters 15(1):57–66
7. Yahya MA, Qodah ZA, Ngah CWZ (2015) Agricultural bio-waste materials as potential sustainable precursors used for activated carbon production: A review. Renew Sust Energ Rev 46:218–235
8. Zhang Y, Yu S, Lou G, Shen Y, Chen H, Shen Z, Zhao S, Zhang J, Chai S, Zou Q (2017) Review of macroporous materials as electrochemical supercapacitor electrodes. J Mater Sci 52:11201–11228

9. Enock TK, King'ondu CK, Pogrebnoi A, Jande YAC (2017) Status of biomass derived carbon materials for supercapacitor application. Int J Electrochem Sci 20:6453420
10. He X, Ling P, Qiu J, Yu M, Zhang X, Yu C, Zheng M (2013) Efficient preparation of biomass-based mesoporous carbons for supercapacitors with both high energy density and high power density. J Power Sources 240:109–113
11. Yahya MA, Baharuddin M, Ngah CWZ, Hashim MA (2015) Activated carbon electrode from desiccated coconut residue for high performance of supercapacitors. J Mater Sci Res 4:45
12. Deng S, Hu B, Chen T, Wang B, Huang J, Wang Y, Yu G (2015) Activated carbons prepared from peanut shell and sunflower seed shell for high CO_2 adsorption. Adsorption 21:125–133
13. Iro ZS, Subramani C, Dash SS (2016) A brief review on electrode materials for supercapacitor. Int J Electrochem Sci 11:10628–10643
14. Simon P, Gogotsi Y (2008) Materials for electrochemical capacitors. Nat Mater 7:845
15. Stoller MD, Ruoff RS (2010) Best practice methods for determining an electrode material's performance for ultracapacitors. Energy Environ Sci 3:1294–1301
16. Zhang S, Pan N (2015) Supercapacitors performance evaluation. Adv Energy Mater 5:1401401
17. Gao Z, Zhang Y, Song N, Li X (2016) Biomass-derived renewable carbon materials for electrochemical energy storage. Mater Res Lett 5(2):69–88
18. Jayalakshmi M, Balasubramanian K (2008) Review: Simple capacitors to supercapacitors. Int J Electrochem Sci 3:1196–1217
19. Ajina A, Isa D (2010) Symmetrical supercapacitor using coconut shell-based activated carbon. Pertanika J Sci & Technol 18(2):351–363
20. Foo KY, Hameed BH (2011) Utilization of rice husks as a feedstock for preparation of activated carbon by microwave induced KOH and K2CO3 activation. Bioresour Technol 102:9814–9817
21. Volperts A, Dobele G, Ozolins J, Ulmane NM (2015) Synthesis and application of nanoporous activated carbon in supercapacitors. Mater Sci Appl Chem 31:3–6
22. He T, Ren X, Nie J, Ying J, Cai K (2015) Investigation of imbalanced activated carbon electrode supercapacitors. Int J Electrochem Sci 2015:801217
23. Natalia M, Sudhakar YN, Selvakuma M (2013) Activated carbon derived from natural sources and electrochemical capacitance of double layer capacitor. Indian J Chem Technol 20:392–399
24. Kamariya S, Pandya J (2016) Preparation and characterization of activated carbon from agricultural waste, peanut shell by chemical activation. Int J Trendres Dev 3(3):2394–9333
25. Yahya MA, Ngah CWZCW, Hashim MA, Al-Qodah Z (2016) Preparation of activated carbon from desiccated coconut residue by chemical activation with NaOH. J Mater Sci Res 5(1):24
26. Wu MB, Li RC, He XJ, Zhang HB, Sui WB, Tan MH (2015) Microwave-assisted preparation of peanut shell-based activated carbons and their use in electrochemical capacitors. New Carbon Mater 30:86–91
27. He X, Ling P, Qiu J, Yu M, Zhang X, Yu C, Zheng M (2013) Efficient preparation of biomass-based mesoporous carbons for supercapacitors with both high energy density and high power density. J Power Sources 240:109–113
28. Kim YJ, Lee BJ, Suezaki H, Chino T, Abe Y, Yanagiura T, Park KC, Endo M (2006) Preparation and characterization of bamboo-based activated carbons as electrode materials for electric double layer capacitors. Carbon 44:1592–1594
29. Sarkar S, Arya A, Gaur UK, Gaur A (2020)Investigations on porous carbon derived from sugarcane bagasse as an electrode material for supercapacitors. Biomass Bioenergy 142:105730
30. Peng C, Yan XB, Wang RT, Lang JW, Ou YJ, Xue QJ (2013) Promising activated carbons derived from waste tea-leaves and their application in high performance supercapacitors electrodes. Electrochim Acta 87:401–408
31. Kwaghger A, Ibrahim JS (2013) Optimization of conditions for the preparation of activated carbon from mango nuts using HCl. Am J Eng Res 2(7):74–85
32. Wang Y, Song Y, Xia Y (2016) Electrochemical capacitors: mechanism, materials, systems, characterization and applications. Chem Soc Rev 45:5925–5950
33. Chen X, Zhang J, Zhang B, Dong S, Guo X, Mu X, Fei B (2017) A novel hierarchical porous nitrogen-doped carbon derived from bamboo shoot for high performance supercapacitor. Sci Rep 7:7362
34. Kant K, Priest C, Shapter JG, Losic D (2014) The influence of nanopore dimensions on the electrochemical properties of nanopore arrays studied by impedance spectroscopy. Sensors 14:21316–21328
35. Song Z, Xu Z (2015) Ultimate osmosis engineered by the pore geometry and functionalization of carbon nanostructures. Sci Rep 5:10597
36. Ribeiro DV, Souza CAC, Abrantes JCC (2015) Use of Electrochemical Impedance Spectroscopy (EIS) to monitoring the corrosion of reinforced concrete. Ibracon Struct Mater J 8:529–546
37. Lei C, Markoulidis F, Ashitaka Z, Lekakou C (2013)Reduction of porous carbon/Al contact resistance for an electric double-layer capacitor (EDLC). Electrochim Acta 92:183–187
38. Basri NH, Dolah BNM (2013) Physical and electrochemical properties of supercapacitor electrodes derived from carbon nanotube and biomass carbon. Int J Electrochem Sci 8:57–273
39. Iqbal M, Saykar NG, Arya A, Banerjee I, Alegaonkar PS and Mahapatra SK (2021) High-performance supercapacitor based on MoS2@ TiO2 composite for wide range temperature application. J Alloy Compd 883:160705
40. Kamboj V, Arya A, Tanwar S, Kumar V, Sharma AL (2021) Nanofiller assisted Na+ conducting polymer nanocomposite for ultracapacitor: structural, dielectric, and electrochemical properties. J Mater Sci 51:6167–6187
41. Zhu H, Wang X, Yang F, Yang X (2011) Promising carbons for supercapacitors derived from fungi. Adv Mater 23:2745–2748
42. Yun YS, Park MH, Hong SJ, Lee ME, Park YW, Jin HJ, Appl ACS (2015) Hierarchically porous carbon nanosheets from waste coffee grounds for supercapacitors. Mater Interfaces 7:3684–3690
43. Tian W, Gao Q, Tan Y, Yang K, Zhu L, Yang C, Zhang H (2015) Bio-inspired beehive-like hierarchical nanoporous carbon derived from bamboo-based industrial by-product as a high performance supercapacitor electrode material. J Mater Chem A 3:5656–5664
44. Hao P, Zhao Z, Leng Y, Tian J, Sang Y, Boughton RI, Wong CP, Liu H, Yang B (2015) Graphene-based nitrogen self-doped hierarchical porous carbon aerogels derived from chitosan for high performance supercapacitors. Nano Energy 15:9–23
45. Lee M, Kim GP, Don Song ,Park S, Yi J (2014) Preparation of energy storage material derived from a used cigarette filter for a supercapacitor electrode. Nanotechnology 25:345601
46. Pang L, Zou B, Han X, Cao L, Wang W, Guo Y (2016) One-step synthesis of high-performance porous carbon from corn starch for supercapacitor. Mater Lett 184:88–91
47. Ganesan A, Mukherjee R, Raj J, Shaijumon MM (2014) Nanoporous rice husk derived carbon for gas storage and high performance electrochemical energy storage. J Porous Mater 21:839–847
48. Qu WH, Xu YY, Lu AH, Zhang XQ, Li WC (2015)Converting biowaste corncob residue into high value added porous carbon for supercapacitor electrodes. Bioresour Technol 189:285–291

49. Mehare MD, Deshmukh AD, Dhoble SJ (2020) Preparation of porous agro-waste-derived carbon from onion peel for supercapacitor application. J Mater Sci 55(10):4213–4224
50. Govindan B, Alhseinat E, Darawsheh IF, Ismail I, Polychronopoulou K, Jaoude MA, ... Banat F (2020) Activated carbon derived from phoenix dactylifera (Palm Tree) and decorated with mno2 nanoparticles for enhanced hybrid capacitive deionization electrodes. Chemistry Select 5(11):3248–3256
51. Bharath G, Hai A, Rambabu K, Ahmed F, Haidyrah AS, Ahmad N, ... Banat F (2021) Hybrid capacitive deionization of NaCl and toxic heavy metal ions using faradic electrodes of silver nanospheres decorated pomegranate peel-derived activated carbon. F Environ Res 197:111110

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

Fishing out the Phishing Websites

Peya Mowar
Information Technology Department
Delhi Technological University
Delhi, India
peyajm29@gmail.com

Mini Jain
Information Technology Department
Delhi Technological University
Delhi, India
mini.j2799@gmail.com

2021 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA) | 978-1-6654-2529-2/20/$31.00 ©2021 IEEE | DOI: 10.1109/CyberSA52016.2021.9478237

***Abstract*— Phishing is a cybercrime in which deceitful websites lure naive users and trick them into disclosing confidential information, such as social media passwords or financial data. Phishing websites are crafted such that they superficially appear similar to popular legitimate websites. This paper aims to detect such phishing websites by proposing a novel classifier that takes lexical-based, script-based, rule-based, and address-based features extracted from a website into account. A large-scale balanced dataset of 38,800 active phishing and legitimate websites is created, on which tree-based ensemble classifiers are trained, out of which the XGBoost (eXtreme Gradient Boosting) model performs the best with a testing accuracy of 99.6%. The classifier can detect zero-day phishing attacks without requiring any third- party features such as page rank. Several other benefits of using this model over the state-of-the-art techniques are discussed.**

Keywords—phishing, cyber security, classification, ensemble learning, cyber-attacks, cybercrime

I. INTRODUCTION

Phishing is among the most prevalent cyber-attack techniques used by malicious entities to steal sensitive data such as passwords, banking details, addresses, and other private information from credulous users. Phishing websites often mimic popular legitimate websites to confuse and mislead the user into visiting them. Phishers use deceptive techniques such as shortening their Uniform Resource Locator (URL) to make it less suspicious or redirecting the user to a different webpage.

While manifold research has been conducted in this domain, the number of phishing websites being detected per day is on the rise, and attackers are adopting diverse styles to pass undetected through phishing detection methods. According to Anti-Phishing Working Group's (APWG) Phishing Activity Trends Report for Q3 2020, September alone saw detection of 199,133 unique phishing websites and the number has kept on increasing since the beginning of the year [1]. Fig. 1 displays instances from our dataset of a popular legitimate website and a phishing website designed to copy it. A few features, such as the presence of a company's copyright, that are indicative of a website being benign are highlighted in green. Certain observations, such as an outdated logo being used, that might suggest that a website is phishing are highlighted in red. In this particular figure, such a phishing website might aim to steal user credentials for a popular social media website, Instagram. Numerous methodologies have been proposed to detect phishing websites by extracting features from the website.

These include a heuristics-based approach[2], content based approach that studies webpage content[3], blacklist based approach that checks the URL against existing phishing black-lists[4] and a machine learning based approach that uses algorithms for classification or a hybrid combination of two or more approaches[5].

(a) Legitimate website (b) Phishing website

1. Example of (a) a legitimate and (b) a phishing website from our dataset

In this paper, we create a large dataset consisting of active phishing and benign website URLs, for which feature engineering is carried out. We propose a hybrid approach that considers features under a diverse umbrella of methods. Finally, we suggest the use of tree-based ensemble models for classification. Following are the advantages of the proposed scheme:

1. Enhanced applicability for real-time detection
2. Fast computation and classification
3. Independent of third-party services
4. Malware-safe: website is not visited or downloaded
5. Up to date, robust and adaptive detection model

The paper is arranged as follows. Section II reviews the related work, research gaps, and future scope in this domain. Section III provides an overview of the curated dataset and our proposed methodology baked on a study of tree-based ensemble learning techniques. Section IV discusses the results of our experimentation with the proposed scheme. Finally, Section V marks this paper's conclusion and explains the future scope for the underlying research.

II. LITERATURE REVIEW

A lot of research has already been done in the phishing detection domain, which can be primarily categorised into three divisions: Blacklist and Heuristic-based approaches, Content and Visual Similarity-based approaches and finally, URL and Fuzzy Rule-based approaches. The work in these categories has been discussed below.

A. Early Work

Prior research and studies have focused on blacklist-based approaches, wherein all websites that belong to a blacklist of phishing webpages are blocked from getting accessed. Often, web browsers also operate on list-based techniques; however, most make use of whitelists instead of blacklists, where they only allow access to legitimate web

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:11:41 UTC from IEEE Xplore. Restrictions apply.

pages that are contained in the whitelist. This is done because the primary approach results in a higher number of false negatives, which, when accessed, might be a danger to the user. PhishTank, Google Safe Browsing, and OpenPhish are some of the popular blacklists available to the public. This approach is employed by [4], [6], [7] and [8]. Another common technique used since early on is the Heuristic-based approach to detect phishing websites and emails. In this procedure, various features are extracted from various components of a webpage and they are used in classification. Important examples of such heuristic-based features include the lexical features of a URL and third-party website page ranks. This method has been used in [9], [2] and [10].

B. Content and Similarity Based Approaches

In a content-based approach, features are extracted from the source HTML code of websites to classify the website. These features include text content-based features, logo image features, tag-based features, etc. CANTINA, a content-based classifier proposed by [3] operates on the textual features extracted from a webpage. However, a significant disadvantage of this classifier is that it falters when an image displaces the text. Such content-based approaches have also been employed by [11], [12] and [13]. Often phishers try to replicate and bring visual and semantic similarity from legitimate, commonly browsed websites. This deceives a user into thinking that he is browsing a legitimate website. Research has been done to create visual similarity-based classifiers, to determine whether two web pages are suspiciously similar. [14], [15], [16], [17] and [18] are a few significant research papers that are based on this approach.

C. URL and Fuzzy Rules

Some approaches make use of extracting features from URLs by deriving from a URL's components and computing a metric for each of them. Common examples of these features include lexical or count-based features, or bag of words contained in the URL. Association rule mining is a common method for classification with such features. Some significant papers that have made use of such features are [19], [20], [21] and [22]. URL-based features, though important, if used singularly might not generate the best predictions, simply because phishers keep up with the trends and format their URL so that it is difficult to detect phishing websites. A significant amount of research has been done in employing a fuzzy rule-based approach to label a website as phishing or benign. [23], [24] and [25] have done important work in this regard.

D. Research Gap and Future Scope

Although various techniques for classification and feature extraction have been applied for phishing website detection, each scheme brings certain limitations or gaps in knowledge. List-based and URL-based practices are outdated and require frequent updates in the learning (lists or features) to keep up with the new phishing website trends. Content-based methods require lots of computational power and time. They also need access to the website's source code, and hence are impractical when it comes to real-time classification. Similarity-based approaches require a broad array of legitimate websites that are being mimicked and also face the same issue of impractical real-time implementation. Hence, there is scope to create a classification model that works on these approaches in conjunction, reduces the individual limitations for each of the above-mentioned categories, and gives robust and accurate results. This is precisely what the classifier that we have proposed in this paper does.

III. EXPERIMENTATION

For our research, we created a phishing dataset with hand-picked features extracted from the latest URLs to build a tree-based classification model that can accurately classify unseen URLs as phishing or benign. Fig. 2 shows the pipeline of the experimentation that has been carried out in this paper.

A. Data Extraction

For a large-scale study, we extracted 38,800 URLs, comprising 19,400 phishing and 19,400 benign URLs. The data extraction has been performed in 2020 so that recent and active websites are used for up-to-date training of the model. We used different methods to extract the phishing and benign URLs, which are explained below.

1) ***Extraction of Phishing URLs:*** The collection of phishing URLs was performed using two open source services, namely, PhishTank and OpenPhish. PhishTank is a public phish repository made available online that allows developers to access their anti-phishing data free of charge. This data is updated hourly. The labelled information for each website is verified by multiple users and trusted by organisations such as Mc Afee and Yahoo Mail. OpenPhish is another publicly-available blacklist that is updated hourly. We used these two blacklists to extract 19,400 phishing URLs on 25th October, 2020. The extraction process was carried out in the following three steps:

a) Registration for an application key was carried out.

b) A request to the PhishTank server was sent using this key.

c) A CSV file containing phishing URLs was then downloaded.

2) ***Extraction of Benign URLs:*** To balance the dataset, 19,400 legitimate URLs were collected from [26]. The dataset originally contained an assortment of malicious and benign webpages, collected by crawling the internet using MalCrawler and verifying the labels by Google Safe Browsing API. We randomly picked a batch of benign websites from this source.

The phishing and legitimate URLs were then combined and assigned the respective labels to create our dataset to train our phishing detection model. Next, we extracted various hand-picked features for each URL in this dataset.

B. Feature Engineering

The collected URLs were processed to extract different features to get diverse information about the websites for training using mathematical formulae and python scripts. As previous work has indicated, the variety of attributes used for classification played a critical role in prediction techniques and results. We were meticulous in hand-picking the features discussed in literature for extraction and chose those that limited their individual category's shortcomings while also giving robust and accurate results. A total of 107 features were extracted, which fell under these four umbrella categories:

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:11:41 UTC from IEEE Xplore. Restrictions apply.

2.

*1) **Address-based features:*** Address-based features include all the information about the address-bar, i.e. the URL and it's domain. Examples include the domain name, time elapsed since activation, time left till expiration, whether any URL-shortening service has been used and the number of name servers associated with the URL. These are a subset of **URL-based features** that were discussed in literature review. A phisher has access and control over subdomain, path and file components of a URL and may alter them at will. Often, a phisher might try to mimic a legitimate URL by slightly changing the spelling such as creating "goggle.com" inspired by "google.com" or in some cases, a URL might redirect to some other website. Hence, it is important to consider these features due to two reasons: a) These features are often indicative, descriptive and apt in classification and b) The URL is the primary source of information when classifying a website in real-time. These features were also extracted using python scripts; for example to derive the domain from the URL, the following code was employed:

```
def getDomain(url):
        domain = parse.urlparse(url).netloc
        return domain
```

*2) **Lexical Features:*** Often phishers add one or more special characters to benign URLs to confuse the user into believing that the site is a legitimate one. These are the features extracted from lexical analysis of the data, consisting of attributes such as count of special characters like '@','#','!' in the URL as well as the domain, query, fragment and parameters of the URL. These are a subset of **heuristic features** that describe the vocabulary content of the URL. These features are fast in computation, do not require any third-party services for extraction since they are based on simple mathematical formulae and deliver quality information about a URL. Examples of such features include the count of the character '@' in the URL. These features were extracted for the dataset URLs using python code such as:

```
def count(url):
        count@ = 0
        for i in url:
                if i=='@':
                        count@ = count@ +1
        return count@
```

*3) **Rule-based Features:*** These are features extracted using mathematical formulae describing the website, which are based on a predefined set of **fuzzy logic** (IF-THEN rules) such as spans of each component of the URL or presence of some keywords like "server", " https", "client" in the URL. These features are again easily computed and based on essential conclusions made by previous researches in the domain of phishing detection. For example, [27] concluded that function for finding the probability of a URL phishing being phishing can be determined from fuzzy logic used in:

```
def probPhish(url_length):
        if url_length < 54:
                return 0
        else if url_length <= 75:
                return 0.5
        return 1
```

*4) **Script-based Features:*** Phishing websites trick users by tweaking the tags of the code to disable right clicking and viewing source code or display the wrong URL in the address-bar. These are features related to the script of the web page such as status bar customisation to show a different URL to the viewer on hovering over the address bar, ability to save webpage source code and count of

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:11:41 UTC from IEEE Xplore. Restrictions apply.

the times the website has been redirected. These are a subset of **content-based features** that focus on the content and source code of a webpage. We have not included any image or textual similarity-based features from the webpage as they require more computational power and are often unreliable since phishers often update this content on their webpages to pass undetected. However, script-based content is not updated that frequently, and is more reliable and descriptive. Examples of such features include disabling right click or using iframe redirection.

```
def forwarding(response):
    if response == "":
        return 1
    else:
        if len(response.history) <= 2:
            return 0
        else:
            return 1
```

Feature selection and standardisation were performed on this dataset which helped in reducing overfitting, improving accuracy, and reducing training time. ANOVA (f_classif) statistical test was used for selecting the features.

C. Classification

The objective of phishing detection is to ascertain whether a suspicious website is a phishing website. We have used a supervised learning approach for this problem. We analysed how ensemble models performed on the dataset in contrast to the base learners. For our analysis, we took the Decision Tree model, and we compared it with its bagging and boosting equivalent, i.e., Random Forest and XGBoost respectively.

1) ***Classification using Decision Tree:*** Previous research had concluded that tree-based classification models, such as Decision Tree, based on a set of IF-THEN rules, produce excellent results for phishing detection [28] and are fast in computation. Therefore, in our research we employed the Decision Tree classifier with Gini Index as the splitting criterion. Gini Index estimates the degree of a variable getting wrongly classified, when chosen randomly. The formula for Gini Index is as follows:

$$G(x) = 1 - \sum (p_i)^2$$

where p_i is the probability of an element being classified to a particular class.

2) ***Classification using Ensemble Classifiers:*** Ensemble learning models are meta-learning programs that blend different deep learning methods into an individual prediction model. The benefits of using ensemble learning are that they aid in decreasing variance (using bagging techniques) and reducing bias (using boosting techniques). Classification was carried out using ensemble learning equivalents of Decision Tree (DT) model, namely the Random Forest (RF) and the eXtreme Gradient Boosting (XGB) algorithms. Random Forest is a bagging method which creates several decision trees for the data and gives the majority output label from all the individual trees. XGBoost is a boosting method that builds trees sequentially. Comparison of the three models was done based on five classification metrics: Accuracy score, F1 score, AUC score, Precision and Recall.

A statistical comparison was also carried out using the robust 5x2cv combined F-test. This test helped us understand whether there was a significant difference in the classification results. We propose the following hypothesis for the statistical comparison of the classifiers:

> H_0**:** Both methods produce results from the same distribution and perform equally well on the dataset.
>
> H_1**:** Both methods do not produce results from the same distribution and their performance is significantly different.

A significance threshold of $\alpha = 0.05$ is assumed for rejecting the null hypothesis, which means that the null hypothesis for the 5x2 cv *f* test is rejected if the p value is less than 0.05, otherwise it is accepted.

3) ***Classification using different feature groups:*** For determining feature importance, we further analysed the classification results for various combinations of features. The classification was performed using the base model, i.e. the decision tree model. The results were again analysed based on the previously mentioned five classification metrics.

IV. RESULTS

The results for the experimentation carried out in this paper are discussed in this section.

A. Feature Selection Results

The top 38 features selected for classification using ANOVA test are listed in Table I. Our analysis shows a strong dependency on lexical features such as the presence of special characters or keywords in the URL. Many domain-related features like age of domain influence the classification strongly. Use of URL shortening services like tinyURL and web forwarding is also prevalent in phishing URLS. Fig. 3 (a) displays the ranks of the most significant ten features using the ANOVA statistical test. Fig. 3(b) describes the distribution of features into the selected feature groups. We observed that lexical features constitute the highest number of selected features, with 44.7% of the retained features belonging to this category.

(a) (b)

3. (a) Ten most significant features determined by ANOVA test
(b) Descriptive chart of number of features in each group

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:11:41 UTC from IEEE Xplore. Restrictions apply.

TABLE I. Selected features using ANOVA test

Feature Group	Included Features
Lexical based (L)	Count of (.) in URL, Count of (-) in URL, Count of (_) in URL, Count of (=) in URL, Count of (?) in URL, Count of (@) in URL, Count of (#) in URL, Count of (,) in URL, Count of (-) in Domain, Count of (.) in Query, Count of (-) in Query, Count of (_) in Query, Count of (=) in Query, Count of (?) in Query, Count of (@) in Query', Count of (.) in Fragment, Count of (@) in Fragment
Address-based (A)	If URL is shortened, Number of resolved IPs (IPv4), Time to Live, Time since Activation, Time in Expiration, Check TLD, Count of TLDs, DNS Record, Count MX Servers, Count Name Servers
Script-based (S)	I-Frame Redirection, Web Forwarding
Rule based (R)	Valid email, URL Depth, Top Level Domain, Length of URL, Length of Domain, Protocol, Length of Query, Domain Age, End time of Domain

B. *Classification Results*

The results of all the classification experiments are summarised in Table II. We split the dataset into training and test sets (8:2) and carried out 5-fold stratified cross-validation for evaluating the models. The average accuracy, precision, recall, F1 score and ROC-AUC is recorded for each classifier. We observed that the ensemble models, namely, Random Forest and XGBoost performed better than the base classifier - Decision Tree. Among the ensemble models, XGBoost had the better accuracy and hence served as our chosen classification model.

TABLE II. Classification results of tree-based techniques

Model	Accuracy	Precision	Recall	F1	AUC
DT	0.969	0.968	0.969	0.969	0.982
RF	0.974	0.980	0.968	0.974	0.998
XGB	**0.996**	0.996	0.997	0.996	0.999

C. *Statistical Results*

The robust 5x2cv combined F-test was conducted for the ensemble classifiers (XGBoost and Random Forest) versus the base classifier (Decision Tree), and the P-value and F-statistics obtained for the same are displayed in the Table III below.

TABLE III. Statistical comparison of ensemble learning techniques

Decision Tree v/s Random Forest		***Decision Tree v/s XGBoost***	
p	0.000	*p*	0.000
f	92.278	*f*	239.603

Since the P-value is lesser than the significance threshold that had been assumed for the test, in the case of both the experiments, we conclude that the null hypothesis can be rejected and that the base and ensemble classifiers produce results from different distributions. Therefore, the ensemble models perform significantly better than the base decision tree model, which is also supported by the results from our statistical test.

D. *Feature Importance*

To understand the importance of different types of features in classification, we used the XGBoost model to evaluate classification results by considering different collections of features. These results are displayed in Table IV.

TABLE IV. Classification results using different feature groups

Features	Accuracy	Precision	Recall	F1	AUC
L	0.769	0.848	0.654	0.739	0.869
A	**0.951**	0.947	0.953	0.949	0.950
S	0.723	0.646	0.988	0.781	0.822
R	0.906	0.952	0.854	0.900	0.956
ALL	**0.996**	0.996	0.997	0.996	0.999

The results show that while address-based features had the highest individual accuracy, the model performs best when all different categories of features are taken into account. Fig. 4 shows line graphs of the testing accuracy obtained for different feature groups by the selected tree-based classifiers indicating that the highest accuracy is achieved by the XGBoost model considering all the different categories of features.

4. Bar graph of accuracy for different classifiers and feature groups

V. CONCLUSIONS

Phishing has become a quite common form of data theft online over the past few years as shown by [1]. It is widely being used for stealing sensitive information leading to financial and identity theft. In this paper, we have created a dataset of active phishing websites and proposed a pipeline to perform feature engineering and classify URLs as benign or phishing considering a diverse set of features. We evaluated different tree-based learners on the basis of their evaluation metrics as well as statistical tests and concluded that the ensemble learning based XGBoost classifier gave us the best results with 99.6% accuracy. Our analysis showed that address-based features like the presence of a DNS record or the age of associated domain are strong indicators of a URL being a phishing URL. Lexical features like the presence of certain special characters or keywords are also good indicators of a phishing URL. Phishers often tweak HTML tags in the script to display web pages within a webpage. Phishing websites are ones that have been redirected multiple times.

Our model does not depend on any pre-existent third-party blacklists and hence is not affected by the frequency of updation of these lists. It also does not consider computation

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:11:41 UTC from IEEE Xplore. Restrictions apply.

heavy image or textual features and hence can be deployed real time. It gives robust results with enhanced real-time applicability and is an improvement over the existing state-of-the-art phishing detection techniques. The future scope in this domain includes identification of tailored phishing attacks that target specific interest groups or organisations.

REFERENCES

1. "Phishing Activities Trends Report 3rd Quarter 2020." https://docs.apwg.org/reports/apwg_trends_report_q3_2020.pdf (accessed Dec. 31, 2020).
2. L. A. T. Nguyen, B. L. To, H. K. Nguyen, and M. H. Nguyen, "A novel approach for phishing detection using URL-based heuristic," 2014, doi: 10.1109/ComManTel.2014.6825621.
3. Y. Zhang, J. I. Hong, and L. F. Cranor, "Cantina: A content-based approach to detecting phishing web sites," 2007, doi: 10.1145/1242572.1242659.
4. J. Zhang, P. Porras, and J. Ullrich, "Highly predictive blacklisting," 2008.
5. A. Alswailem, B. Alabdullah, N. Alrumayh, and A. Alsedrani, "Detecting Phishing Websites Using Machine Learning," 2019, doi: 10.1109/CAIS.2019.8769571.
6. P. Prakash, M. Kumar, R. Rao Kompella, and M. Gupta, "PhishNet: Predictive blacklisting to detect phishing attacks," 2010, doi: 10.1109/INFCOM.2010.5462216.
7. M. Felegyhazi, C. Kreibich, and V. Paxson, "On the potential of proactive domain blacklisting," 2010.
8. Y. Wang, R. Agrawal, and B. Y. Choi, "Light weight anti-phishing with user whitelisting in a web browser," 2008, doi: 10.1109/TPSD.2008.4562720.
9. M. N. Feroz and S. Mengel, "Phishing URL Detection Using URL Ranking," 2015, doi: 10.1109/BigDataCongress.2015.97.
10. R. S. Rao and S. T. Ali, "PhishShield: A Desktop Application to Detect Phishing Webpages through Heuristic Approach," in *Procedia Computer Science*, 2015, vol. 54, doi: 10.1016/j.procs.2015.06.017.
11. C. L. Tan, K. L. Chiew, and S. N. Sze, "Phishing website detection using URL-assisted brand name weighting system," 2014, doi: 10.1109/ISPACS.2014.7024424.
12. K. L. Chiew, E. H. Chang, S. N. Sze, and W. K. Tiong, "Utilisation of website logo for phishing detection," *Computers and Security*, vol. 54, 2015, doi: 10.1016/j.cose.2015.07.006.
13. Y. S. Chen, H. S. Liu, Y. H. Yu, and P. C. Wang, "Detect phishing by checking content consistency," 2014, doi: 10.1109/IRI.2014.7051880.
14. A. K. Jain and B. B. Gupta, "Phishing detection: Analysis of visual similarity based approaches," *Security and Communication Networks*, vol. 2017. 2017, doi: 10.1155/2017/5421046.
15. X. Zhang, Y. Zeng, X. B. Jin, Z. W. Yan, and G. G. Geng, "Boosting the phishing detection performance by semantic analysis," in *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017*, 2017, vol. 2018-January, doi: 10.1109/BigData.2017.8258030.
16. K. Ding, N. Pantic, Y. Lu, S. Manna, and M. I. Husain, "Towards building a word similarity dictionary for personality bias classification of phishing email contents," 2015, doi: 10.1109/ICOSC.2015.7050815.
17. K. Ding, N. Pantic, Y. Lu, S. Manna, and M. I. Husain, "Towards building a word similarity dictionary for personality bias classification of phishing email contents," 2015, doi: 10.1109/ICOSC.2015.7050815.
18. R. Verma and N. Hossain, "Semantic feature selection for text with application to phishing email detection," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014, vol. 8565, doi: 10.1007/978-3-319-12160-4_27.
19. L. A. T. Nguyen, B. L. To, H. K. Nguyen, and M. H. Nguyen, "Detecting phishing web sites: A heuristic URL-based approach," 2013, doi: 10.1109/ATC.2013.6698185.
20. H. Choi, B. B. Zhu, and H. Lee, "Detecting malicious web links and identifying their attack types," *WebApps*, 2011.
21. D. Zhang, Z. Yan, H. Jiang, and T. Kim, "A domain-feature enhanced classification model for the detection of Chinese phishing e-Business websites," *Information and Management*, vol. 51, no. 7, 2014, doi: 10.1016/j.im.2014.08.003.
22. F. Feng, Q. Zhou, Z. Shen, X. Yang, L. Han, and J. Q. Wang, "The application of a novel neural network in the detection of phishing websites," *Journal of Ambient Intelligence and Humanized Computing*, 2018, doi: 10.1007/s12652-018-0786-3.
23. B. L. To, L. A. T. Nguyen, H. K. Nguyen, and M. H. Nguyen, "A novel fuzzy approach for phishing detection," 2014, doi: 10.1109/CCE.2014.6916759.
24. S. D. Shirsat, "Demonstrating Different Phishing Attacks Using Fuzzy Logic," 2018, doi: 10.1109/ICICCT.2018.8473309.
25. L. A. T. Nguyen, B. L. To, and H. K. Nguyen, "An Efficient Approach for Phishing Detection Using Neuro-Fuzzy Model," *Journal of Automation and Control Engineering*, vol. 3, no. 6, 2015, doi: 10.12720/joace.3.6.519-525.
26. A. K. Singh, "Malicious and Benign Webpages Dataset," *Data in Brief*, vol. 32, 2020, doi: 10.1016/j.dib.2020.106304.
27. R. M. Mohammad, F. Thabtah, and L. McCluskey, "An assessment of features related to phishing websites using an automated technique," 2012.
28. R. Kiruthiga and D. Akila, "Phishing websites detection using machine learning," *International Journal of Recent Technology and Engineering*, vol. 8, no. 2 Special Issue 11, 2019, doi: 10.35940/ijrte.B1018.0982S1119.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:11:41 UTC from IEEE Xplore. Restrictions apply.

Forest Fires in India: A Review

Reuben Sewak[1], Mehak Vashisth[1], Lovleen Gupta[1*]

[*] **Corresponding Author E-mail:** lgupta@dce.ac.in

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

Abstract: Forest fire has become extreme and has increased in frequency worldwide with India being no exception. In view of this, the present-day knowledge about the forest fire condition in India has been reviewed. The forest cover in India is divided into four clusters namely North Himalayan, North-Eastern, Southern, and Central. Of these clusters, it has been found that though the North-Eastern cluster encompasses the maximum area (55%) of all the clusters in India, it accounts for meagre 16% of the burnt forest area. The central cluster has been found to be the most vulnerable to forest fires as ~56% of the total burnt area lies in this region, though it covers only 28% of the forest cover. In North-East fires tend to concentrate in a smaller area that is subject to repeated burns whereas in Central and Southern India they are more expansive. The impacts of forest fires on biodiversity, forest species, climate change, and soil characteristics in India are presented and various strategies adopted by the Indian government in order to detect and control the forest fires are presented. Additionally, the economic impact of forest fires is estimated to be over INR 1101 crore/year.

Keywords: Forest Fire, Clusters, MODIS, Forest Survey of India, Flora, Fauna

1. Introduction

Forests are crucial to life; they sustain biodiversity and provide endless incalculable products and important ecosystem resources such as storage of carbon, flood & landslide control, clean supply of water, medicines, and crops (Bonell et al., 2010; Krishnaswamy et al., 2009). The global forest comprises 4.06 billion hectares of the area which is equivalent to 31% of the total land area (FAO, 2020). Approximately 1.6 billion people across the world depend on these forests, including 300 million people living within them (WWF, 2011). Between the years 2000 to 2010, a decline of the forest area by about 5.2 million hectares per year (0.1% of the total forest area) has been discerned (FAO, 2020). Forests face various aggravations that can cause harmful effects to their health and productivity along with decreasing their ability to supply products and services. Forest fires being one of them, is ubiquitous and can take place in nearly any ecosystem (Archibald et al., 2013). While fire has been used in crop rotation, agricultural plowing, and

the production of pasture for domesticated animals since ancient times, socioeconomic changes have caused traditional uses to be abandoned (Vidal-Macua et al., 2018).

Forest fires are accounted to be a burgeoning phenomenon over the years ravaging ecosystems, communities, and economies (Jhariya & Raj, 2014). Increasing human reliance on forests for a number of reasons leads to their fragmentation which further aggravates fire events (Cochrane, 2003). Various studies (Cochrane, 2003; Grutzen & Andreae, 1990; Penner et al., 1992) mention that forest fires have become extreme and have increased in frequency worldwide. The impact of forest fires have been documented on biodiversity (Granström, 2001), human life (Finlay et al., 2012; Fowler, 2003), climate change (Wotton et al., 2003, 2010) along with their associated economic impacts (González-Cabán, 2013; Kim et al., 2011; Úbeda & Sarricolea, 2016). India with 21.67% (7,12,249 km^2) area under forests is no exception to forest fires. 55% of forests in India are vulnerable to intermittent fires (Forest Survey of India, 2011a) and fires affect around 3.73 million hectares of forests annually (Chandra & Kumar Bhardwaj, 2015). In addition, population growth in India has triggered substantial deforestation, causing further fragmentation of existing forests. “Present biotic pressures like logging, grazing, and collection of non-timber forest products have also rendered Indian forests susceptible to fires” (Narendran et al., 2001).
This study reviews the scientific literature available to date that examines the forest fires in India from an environmental, social, and planning perspective. This study presents the overall forest and forest fire scenario in India, including the impacts and fire prevention strategies. This paper presents a new contribution to the knowledge of forest fire scenario in India.

2. Forest fire in India: general trends

India occupies just only 2.5% of the global geographical area, 1.8% of the overall forest area and it supports 16% of the total human population (Ashutosh & Satendra, 2014). The forest cover in India, according to the Forest Survey of India is 71.22 Mha, constituting 21.67 % of the total area, which represents 3.02% very dense, 9.39% moderately dense and 9.26% as open forest (Forest Survey of India, 2011b). Out of the 647 districts in India, nearly 380-445 districts encountered fires every year from 2003 to 2016 (World Bank, 2018). According to MODIS (Moderate Resolution Imaging Spectroradiometer) forest fire detections, approximately 37,059 fires were detected in 2018 (Forest Survey of India, 2011b). The classification of Indian forests is shown in Table 1, which shows that nearly 4% of the forest area in the country is extremely vulnerable to fire, while more than 6% is found to be highly vulnerable to fire.

Table 1: Classification of Indian Forests

Forest fire prone classes	Forest cover (in km^2)	% of total forest cover
Extremely fire prone	25,617	3.89
Highly fire prone	39,500	6.01
Fire prone	75,952	11.5
Moderately fire prone	96,422	14.7
Less fire prone	4,20,625	63.9
Total	**6,58,116**	**100**

Source (World Bank, 2018)

With respect to the forest area and the fire ecology variation therein, India can be divided into four geographical clusters viz. North Himalayan, North-Eastern,

Southern, and Central as shown in Figure 1. These clusters encompass ~ 90% of the total forest cover in India and encountered 98% of the total fire points detected in India from the year 2003-16. Although fires are spread across the identified clusters, they happen more frequently in certain areas and affect the forest more than in others. The North-Eastern cluster accounts for the maximum forest cover of 36% and the number of fire detections (40%), however, the maximum forest burnt area (56%) is in the Central cluster (Figure 2), which suggests that the Central cluster is the most vulnerable region in India.

The Central cluster encompasses states like Madhya Pradesh and some districts of Chhattisgarh and Rajasthan. With a subtropical climate, this cluster encounters average rainfall from 800 mm to 1,800 mm with the mean annual temperature varying between 22°C to 25°C. It mainly consists of forest species such as Tectona grandis, Shorea robusta, Lagerstroemia parviflora, and fires therein are caused by slashing and burning agricultural practices that help to collect non-timber forest products (Saha, 2002).

Forest fires are common and widespread in this cluster's dry and humid deciduous forest areas in the borderlands of Chhattisgarh, Maharashtra, and Telangana. (World Bank, 2018). States including Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, and Tripura are primarily part of the North-Eastern cluster. It experiences rainfall of 2,100 mm to 3,500 mm. During the winter months (October-January), temperature varies from 11 °C to 24 °C, while during the summer months (March-June) it encounters temperature ranging from 18 °C to 29 °C. Important species include Schima wallichii, Tectona grandis, Callicarpa species, and Pinus kasya, etc.

The maximum number of fires detected (55%) were in the North-Eastern cluster which also accounts for the maximum forest cover of 36% (Figure 2). In this cluster, fires tend to concentrate in a smaller area that is subject to repeated burns. “This repeated pattern of burning on small forest area is consistent with the practice of shifting cultivation (jhum) seen throughout the northeast. Fires in North-Eastern cluster, occur mainly from slash and burn agriculture”(World Bank, 2018). Due to population growth, there has been a reduction in time between fires, is has reduced from 20-30 years to 2-3 years (Puri et al., 2011), breaking the ecosystem's resilience.

Coniferous forests in the North Himalayan cluster region comprise species such as fir, spruce, and pine. Due to the harsh conditions, tree growth in this cluster is minimal. The annual rainfall ranges between 1200 mm to 1500 mm and the temperature varies from sub-zero to 35°C. The North Himalayan cluster is spread over states such as Kashmir, Uttarakhand, and Himachal Pradesh. Important species include Shorea robusta, Pinus roxburghii, Quercus leucotrichophora, and Mallotus philippinensis. Fire is used as a tool by indigenous communities of the area to clear grass that contributes to forest regeneration (Bahuguna & Upadhay, 2002).

Southern Cluster consists of states like Karnataka, Kerala, Tamil Nadu, Andhra Pradesh, Telangana, and some districts of Maharashtra. This cluster encounters average rainfall from 1000 mm to 1,800 mm, while the mean annual temperature varies between 20°C to 26°C. Important species include Dipterocarpus indicus and Vateria indica. Fires in Southern clusters are more expansive and occur due to human accidents, for management of forest, and due to slash and burn practices (Kodandapani et al., 2004).

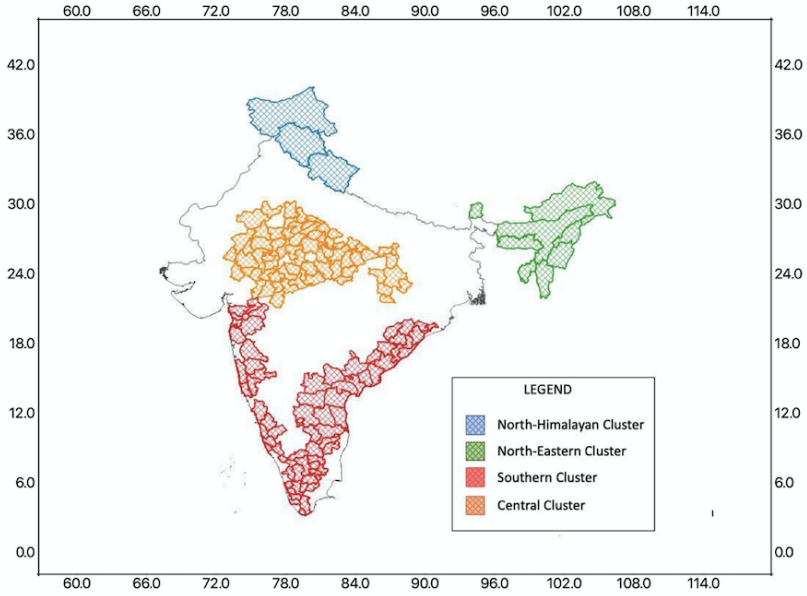


Figure 1: Map showing forest clusters found in India

Of the total forest fire detections in India from the year 2003-16, 20 districts accounted to 40% of all forest fire detections (World Bank, 2018), majority of them lie in the North Eastern cluster. Those 20 districts are listed in Table 2.

Table 2: Total number of fire detections in top 20 districts, 2003-2016

District, State	Cluster	Share of fire detections, 2003-2016 (%)
Lunglei, Mizoram	North East	3.82
Karbi Anglong, Assam	North East	3.48
Dima Hasao, Assam	North East	3.3
Churachandpur, Manipur	North East	3.15
Mamit, Mizoram	North East	2.56
Lawngtlai, Mizoram	North East	2.42
Tamenglong, Manipur	North East	2.32
Aizawl, Mizoram	North East	1.91
Gadchiroli, Maharashtra	Central	1.78
Dhalai, Tripura	North East	1.77
Champhai, Mizoram	North East	1.69
W. Khasi Hills, Meghalaya	North East	1.48
Narayanpur, Chhattisgarh	Central	1.45
Ribhoi, Meghalaya	North East	1.37
Kandhamal, Odisha	Central	1.35
E. Garo Hills, Meghalaya	North East	1.33
Ukhrul, Manipur	North East	1.32
Chandel, Manipur	North East	1.32
Bijapur, Chhattisgarh	Central	1.31
North Tripura, Tripura	North East	1.16
Top 20 total		**40.28**

Source (World Bank, 2018)

Conversely, with regard to the total area affected by forest fires, the top 20 districts which represented 48% of the total area affected by fire between 2003 and 2016, majority of them belong in the central cluster, (World Bank, 2018). These 20 districts in the central cluster are listed in Table 3

Table 3: Total area affected by fire in top 20 districts, 2003-2016

District, State	Cluster	Share of burnt area, 2003-2016 (%)
Gadchiroli, Maharashtra	Central	8.24
Bijapur, Chhattisgarh	Central	5.29
Khammam, Telangana	Southern	3.86
Narayanpur, Chhattisgarh	Central	2.7
Warangal, Telangana	Southern	2.56
Koriya, Chhattisgarh	Central	2.35
Adilabad, Telangana	Southern	2
Chandrapur, Maharashtra	Central	1.95
Surguja, Chhattisgarh	Central	1.9
Kurnool, Andhra Pradesh	Southern	1.8
Amravati, Maharashtra	Central	1.78
Y.S.R., Andhra Pradesh	Southern	1.71
Prakasam, Andhra Pradesh	Southern	1.7
Dakshin Bastar, Chhattisgarh	Central	1.61
Bilaspur, Chhattisgarh	Central	1.6
Raipur, Chhattisgarh	Central	1.56
Betul, Madhya Pradesh	Central	1.46
Champhai, Mizoram	North East	1.42
Lawngtlai, Mizoram	North East	1.35
Dima Hasao, Assam	North East	1.34
Top 20 total		**48.18**

Source (World Bank, 2018)

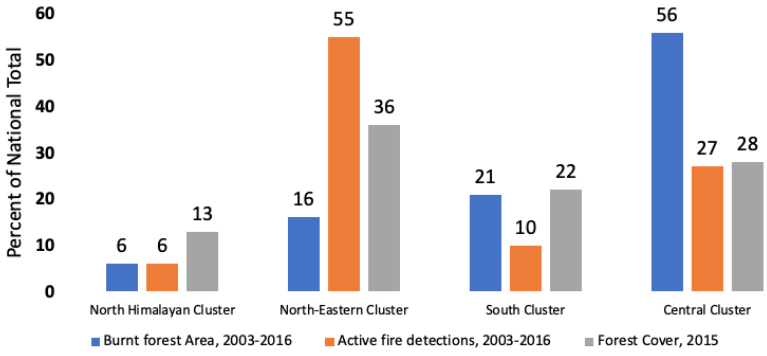


Figure 2: Burnt forest area, Active forest fire detections and Forest Cover from 2003-2016 (World Bank, 2018)

3.0 Impact of forest fires

3.1 Impact on biodiversity

India is rich with rare and complex floral and faunal resources, accounting collectively for 60-70% of the world 's biodiversity (Creswell I, 2018; Producer et al., 2006). Wildfires can have devastating effects on biodiversity by destroying trees, ecosystems, and food supplies and by increasing the resulting susceptibility to predation of surviving animals (Green & Sanecki, 2006; Kodandapani et al., 2008; Letnic & Dickman, 2005; Pelegrin & Bucher, 2010).

Flora

Spatial factors such as the location of a plant's dormant bud, the sub-surface distribution of reproductive structures, and the depths below the surface from which new shoots emerge depict an individual plant's growth behavior post-fire (Flinn & Pringle, 1983; Paula et al., 2009). Depending on the intensity and return period, fires can be beneficial or harmful (Verma & Jayakumar, 2012). According to Cochrane et al., 2009, there has been a "drastic alteration" in the species structure and abundance found in the Nilgiri Biosphere Reserve in the southern cluster, with reduced seedling density observed in the forest area. In the Central cluster Jhariya, M.K. S.S. Bargali, 2014 find similar trends in the Bhoramdeo wildlife dry deciduous forests of Chhattisgarh. Kittur et al., 2014 observe that the regeneration and size structure of economically significant species of plants such as Sal is disrupted by repetitive fires. In the North Himalayan cluster, fire has been a cause for damage to seedlings in Uttarakhand's pine, oak, and mixed deciduous forests (Parashar et al., 2003). Human pressures such as burning, lopping, grazing, and collection of leaf litters have slowly resulted in the transformation of Uttarakhand's oak forests into pine forests, as fires promote the growth of chir-dominated pine forests (Singh et al., 1984).

Fauna

In the North Himalayan cluster, the loss of flora and fauna triggered by wildfires during the summer in these forests is higher than any other human induced factor (Hussain et al., 2018). Evaluating the role of fire (Rodgers, 1986) finds that fires in some areas can be beneficial for wild herbivores. Regulated patchwork burning can improve habitat for grazing species such as swamp deer and chital for small areas of moist grassland. However, the advantages of fire decrease as habitats get drier and fire frequency increases. Also, while fire may be helpful to some degree in promoting habitat for some wild herbivores, not all species benefit. Nests, dens, and eggs can be destroyed by even low-intensity surface fires, killing young animals that cannot escape fast enough. (Rodgers, 1986).

3.2 Impact on soil properties

Forest fires have a significant impact on post-fire forest composition and structure of soil (Jain et al., 2008). It can affect many physical and chemical soil properties including "loss of organic soil and structure, reduced porosity, and increased pH" (Certini, 2005; Debano, 1991). Indirect consequences, include increased water repellence resulting in reduced penetration and increased runoff, which leads to increased erosion (DeBano, 2003). Fires of higher intensity can considerably deplete the soil and alienate it of nutrients and organic matter (Chandra & Kumar Bhardwaj, 2015). In South Kashmir of the North Himalayan cluster (Khaki, B.A., Singh. V.R.R., Wani A.A. & Thakur, 2015) discovered evidence that the overall soil carbon and nitrogen content in burnt areas were lower than unburnt areas, while "phosphorus and potassium" were higher in burnt areas. "The shortening of fire-associated jhum cycles in North-Eastern cluster has had a negative impact on soil fertility". Shortened jhum cycles lower the amount of fallow biomass available for combustion, and offer less time to recuperate soil fertility resulting in lower economic yields and output (Ramakrishnan, 2007).

3.3 Impact on climate change

Climate change has been observed to be on the rise due to the dominant anthropogenic activities that have been exceeding the bounds of natural

variability. One of the numerous consequences associated with climate change is forest fires (Stephens, 2005). Forest fire and climate change possess a feedback effect. Human-instigated climate change has attempted to put more prominent pressure on numerous forest species and has weakened their ability to withstand fires. Within certain regions of the world, more extreme and recurrent wildfires are expected to occur in the future due to climate change (Hemp, 2009; McKenzie et al., 2004; Pitman et al., 2007; Wilson et al., 2010). The appropriate period to find out the fire sensitivity (frequency and intensity) is during the summer when the moisture content and wind speed in the atmosphere is low (Murthy et al., 2019; North et al., 2015; Piñol et al., 1998). This can be regarded as one of the major reasons for how the drought-prone areas are easily drawn to wildfires. The effect of fluctuating climatic patterns on the forest fires of India has also been pointed out by Joseph et al., 2009. They were able to show the shift in the extent of grasslands and forests in response to climate change through their study. (Joseph et al., 2009 and references cited therein).

3.4 Economic impact of forest fires

"Forestry is the second-largest land use after agriculture and accounts for about 1.5 % of the nation's GDP" (World Bank, 2005). With 65 million people identified as tribal groups, they rely on the collection of non-wood forest products from forest areas for their livelihoods and are directly affected by forest fire. (Ashutosh & Satendra, 2014). The economic losses associated with forest fires can be much greater for smallholders than for large holders, because they rely on the forest for a wide range of services and uses (Daniel Nepstad Adriana Moreira & 1999a). The existing estimates of the potential cost of forest fires in India are likely underestimated at approximately INR 1,101 crore (US$ 164 million, 2016 prices) per year (Bahuguna, 1999; World Bank, 2018). The loss of biodiversity, timber, increased carbon sequestration, soil moisture, and nutrient loss, etc., are not taken into account in this estimate. Furthermore, there is a lack of comprehensive data in India that can indicate forest losses in terms of area burned, values, and volume and regeneration destroyed by fire. The available forest fire estimates are not correct because the measured fire numbers and the area burned are underestimated. The rationale behind this is due to the absence of accountability (Bahuguna, V.K. & Singh, 2001).Monetary damage from forest fires is usually only measured in terms of timber value for the loss of standing (natural or planted) trees, which is usually negligible in the case of low-intensity surface fires.(World Bank, 2018). Estimates may be enhanced by considering the direct and indirect effects on other industries like, for example, transport, utilities, loss of environmental services, etc. Thus, the estimate of INR 1101 crore/year is underestimated and the actual economic impact of forest fires is likely to be much higher.

4. Fire management practices in India

4.1 Detection practices

Globally, various wireless sensors, satellite systems, and neural network-based techniques are being employed in order to detect or manage forest fires. However, in India, currently, fires are detected using satellites and ground-based measurements.

Using Satellites and Remote Sensing

In India, active forest fire detection systems rely on the MODIS (Moderate Resolution Imaging Spectroradiometer) instrument and the Visible Infrared Imaging Radiometer Suite (VIIRS) for satellite-based hotspot observation (World Bank, 2018). Using satellite detection, the forest fire alert system was first introduced in India in 2007-2008. Madhya Pradesh (Central Cluster) pioneered its Fire Warning Messaging System to distribute warnings to field workers through SMS text alerts (FAMS). After undergoing a significant number of improvements since 2004, the Forest Fire Alert System has shown a 10-fold upgrade of its version to Forest Fire Alert System Versions 3.0 (Fast 3.0) in the year 2019. This system is based on collaborative arrangements between NASA-ISRO (National Aeronautics and Space Administration- Indian Space Research Organization) and ISRO- FSI (Indian Space Research Organization-Forest Survey of India) (Forest Survey of India, 2011a).

Ground-based detection of fires by field staff

Even with advancements in remote sensing technology and warning systems, ground monitoring of forest fires will continue to be important. However, this mechanism is under-resourced in India. The forest service employs seasonal fire watchers from the local community to help with fire detection and response. These locally employed watchers are facing payment delays or shortages. No proper training instructions or equipment is provided to these fire watchers (World Bank, 2018). Thus, its important on the Indian government's part to strengthen the ground-based fire detection system.

Some of the commonly used fire detection techniques employed in India have been compared in terms of their cost and practicality in Table 4.

Table 4: Comparison Amongst Different Forest Fire Detection Techniques

Name of the Fire Detection Technique	Cost	Practicality	References
Wireless Sensor Based Techniques	Medium	High	(Aslan et al., 2012), (Díaz-Ramírez et al., 2012), (Bayo et al., 2010), (Bouabdellaha et al., 2013)
Satellite Based Systems	Very High	Very High	(Oliva & Schroeder, 2015), (Filizzola et al., 2016), (Coppo, 2015), (Koltunov et al., 2016)
Digital Camera Sensor	High	Medium	(Narasimha Rao et al., 2018), (Wong & Fong, 2014), (Çetin et al., 2013), (Q. X. Zhang et al., 2018)
Air Borne Techniques	High	High	(Krüll et al., 2012), (Cruz et al., 2016), (Allison et al., 2016),(Dennison & Roberts, 2009)
Neural Network Based	Low	Low	(Soliman et al., 2010), (Chauhan et al., 2013) (Giriraj et al., 2008), (Zhang et al., 2009)

4.2 Prevention Strategies

Formulation of National Forest Policy in 1988, The National Master Plan for Forest Fire Control, and Guidelines in 2018 are some of the key initiatives that the Government of India has taken to manage the fire. Common methods for prevention of fire used in India include - fire line clearance and controlled burning to limit fuel loads, silvicultural practices such as selective thinning and fire-adapted tree species planting in fire-prone areas, early warning, and fire hazard rating systems. The implementation of Joint Forest Management (JFM) village level committees was another significant step in the micro-level management of forest fires in India.

The knowledge in systematic forest fire control is limited to the short term. There are weaknesses in the evaluation, ranging from fire detection to the coordination of preventive steps. At this time, the need to establish an adaptive management plan is critical. Sector-wise, with short- and long-term visions, clear points are suggested for the implementation of a holistic fire prevention plan in the light of climate change. Some of them include mitigation measures to minimize the uncertainty in baseline data, strengthen present fire-fighting programs, developing precautionary measures, integrating institutional efforts, publicity, extension, and training, legal measures, and funding of more programs aimed to improve the current status.

5. Conclusion

India with 21.67% (7,12,249 km^2,) area under forests is no exception to wildfires. 55% of forests in India are vulnerable to intermittent fires and wildfires affect around 3.73 million ha of forests per year in India. Wildfires in India impact the nation on different fronts, creating three major kinds of problems.

Environmental: The four geographic clusters in North-east, North-West Himalayan Region, Central, and Southern India best describe the fire ecology variation within India. Northeast accounts for the greatest no. of fire detections (nearly 55%) while the Central region accounts for the largest area affected by forest fire (nearly 56%). In the Western Himalayas, the impact of forest fire causes the loss of flora and fauna during the summer and is greater in these forests than any other anthropogenic factor. In the case of the North East cluster, in addition to other stresses, such as intensive grazing and browsing, frequent fires in short succession reduces species diversity and damages natural regeneration. Lastly, shorter duration fires occurred in the Central and Southern regions of the country have a detrimental impact on the composition and structure of forest, and diversity of species.

Social: In addition to reducing their ability to supply goods and services, forests face multiple aggravations that can cause adverse effects on their health and productivity. Being one of them, forest fires are ubiquitous and can take place in almost every eco-environment. Although fire has been used since ancient times in crop rotation, agricultural plowing, and pasture production for domesticated animals, socio-economic changes have caused traditional uses to be abandoned.

Economic: For small holders, the economic losses associated with forest fires may be much greater than for large holders, since a wide variety of resources and uses depend on the forest. At an annual rate of approximately INR 1,101 crore, the current estimates of the economic costs of forest fires in India are almost certainly underestimated (US$164 million, 2016 prices). However, not much data has been reviewed regarding the economic losses during the present forest fire phenomenon due to a lack of accountability. Many aims would be fulfilled by a more detailed accounting of the economic costs and benefits of forest fires in India.

Finally, with varying strategies and techniques adopted for forest fire detection and prevention, the appreciable difference has been discerned in the forest fire trends of the country.

Nomenclature

WHO: World Health Organization
FAO: Food and Agriculture Organization
MODIS: Moderate Resolution Imaging Spectroradiometer

Mha: Million hectares
FSI: Forest Survey of India

Acknowledgments

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

References:

Allison, R. S., Johnston, J. M., Craig, G., & Jennings, S. (2016). Airborne optical and thermal remote sensing for wildfire detection and monitoring. Sensors (Switzerland), 16(8). https://doi.org/10.3390/s16081310

Archibald, S., Lehmann, C. E. R., Gómez-Dans, J. L., & Bradstock, R. A. (2013). Defining pyromes and global syndromes of fire regimes. Proceedings of the National Academy of Sciences of the United States of America, 110(16), 6442–6447. https://doi.org/10.1073/pnas.1211466110

Ashutosh, D. K., & Satendra. (2014). FOREST FIRE DISASTER MANAGAMENT National Institute of Disaster Management.

Aslan, Y. E., Korpeoglu, I., & Ulusoy, özgür. (2012). A framework for use of wireless sensor networks in forest fire detection and monitoring. Computers, Environment and Urban Systems, 36(6), 614–625. https://doi.org/10.1016/j.compenvurbsys.2012.03.002

Bahuguna, V. K. (1999). Forest fire prevention and control strategies in India. International Forest Fire News, 20(20), 5–9.

Bahuguna, V. K., & Upadhay, A. (2002). Forest fires in India: policy initiatives for community participation. International Forestry Review, 4(2), 122–127. https://doi.org/10.1505/ifor.4.2.122.17446

Bahuguna, V.K. & Singh, S. . 2001. (2001). The forest fire situation in India. Int. Forest Fire News 26.

Bayo, A., Antolín, D., Medrano, N., Calvo, B., & Celma, S. (2010). Early detection and monitoring of forest fire with a wireless sensor network system. Procedia Engineering, 5, 248–251. https://doi.org/10.1016/j.proeng.2010.09.094

Bonell, M., Purandara, B. K., Venkatesh, B., Krishnaswamy, J., Acharya, H. A. K., Singh, U. V., Jayakumar, R., & Chappell, N. (2010). The impact of forest use and reforestation on soil hydraulic conductivity in the Western Ghats of India: Implications for surface and sub-surface hydrology. Journal of Hydrology, 391(1–2), 47–62. https://doi.org/10.1016/j.jhydrol.2010.07.004

Bouabdellaha, K., Noureddine, H., & Larbi, S. (2013). Using wireless sensor networks for reliable forest fires detection. Procedia Computer Science, 19(Seit), 794–801. https://doi.org/10.1016/j.procs.2013.06.104

Certini, G. (2005). Effects of fire on properties of forest soils: A review. Oecologia, 143(1), 1–10. https://doi.org/10.1007/s00442-004-1788-8

Çetin, A. E., Dimitropoulos, K., Gouverneur, B., Grammalidis, N., Günay, O., Habiboğlu, Y. H., Töreyin, B. U., & Verstockt, S. (2013). Video fire detection - Review. Digital Signal Processing: A Review Journal, 23(6), 1827–1843. https://doi.org/10.1016/j.dsp.2013.07.003

Chandra, K. K., & Kumar Bhardwaj, A. (2015). Incidence of Forest Fire in India and Its Effect on Terrestrial Ecosystem Dynamics, Nutrient and Microbial Status of Soil. International Journal of Agriculture and Forestry, 5(2), 69–78. https://doi.org/10.5923/j.ijaf.20150502.01

Chauhan, A., Semwal, S., & Chawhan, R. (2013). Artificial neural network-based forest fire detection system using wireless sensor network. 2013 Annual IEEE India Conference, INDICON 2013. https://doi.org/10.1109/INDCON.2013.6725913

Cochrane, M. A. (2003). Fire science for rainforests. Nature, 421(6926), 913–919. https://doi.org/10.1038/nature01437

Cochrane, M. A., Kodandapani, N., Cochrane, M. A., & Sukumar, R. (2009). Forest fire regimes and their ecological effects in seasonally dry tropical ecosystems in the Western Ghats, India. Tropical Fire Ecology, 335–354. https://doi.org/10.1007/978-3-540-77381-8_12

Coppo, P. (2015). Simulation of fire detection by infrared imagers from geostationary satellites. Remote Sensing of Environment, 162, 84–98. https://doi.org/10.1016/j.rse.2015.02.016

Creswell I, M. H. (2018). Cresswell ID, Murphy H (2016). Biodiversity: Biodiversity. In: Australia state of the environment 2016, Australian Government Department of the Environment and Energy, Canberra, https://soe.environment.gov.au/theme/biodiversity, DOI 10.4226/94/58b65ac8288.

Cruz, H., Eckert, M., Meneses, J., & Martínez, J. F. (2016). Efficient forest fire detection index for application in Unmanned Aerial Systems (UASs). Sensors (Switzerland), 16(6). https://doi.org/10.3390/s16060893

Daniel Nepstad Adriana Moreira, A. A., & 1999a. (n.d.). Flames in the Rain Forest – Origins , Impacts and Alternatives to Amazonian Fires. 22, 1998–1999.

Debano, L. F. (1991). The effect of fire on soil properties. General Technical Report - US Department of Agriculture, Forest Service, INT-280, 151–156.

DeBano, L. F. (2003). The role of fire and soil heating on water repellency. Soil Water Repellency: Occurrence, Consequences, and Amelioration, 232, 193–202. https://doi.org/10.1016/B978-0-444-51269-7.50020-5

Dennison, P. E., & Roberts, D. A. (2009). Daytime fire detection using airborne hyperspectral data. Remote Sensing of Environment, 113(8), 1646–1657.

https://doi.org/10.1016/j.rse.2009.03.010
Díaz-Ramírez, A., Tafoya, L. A., Atempa, J. A., & Mejía-Alvarez, P. (2012). Wireless Sensor Networks and Fusion Information Methods for Forest Fire Detection. Procedia Technology, 3(4322), 69–79. https://doi.org/10.1016/j.protcy.2012.03.008
FAO, 2020. (2020). Global Forest Resources Assessment 2020 (Vol. 8). https://doi.org/10.1097/01.gim.0000223467.60151.02
Filizzola, C., Corrado, R., Marchese, F., Mazzeo, G., Paciello, R., Pergola, N., & Tramutoli, V. (2016). RST-FIRES, an exportable algorithm for early-fire detection and monitoring: description, implementation, and field validation in the case of the MSG-SEVIRI sensor. Remote Sensing of Environment, 186, 196–216. https://doi.org/10.1016/j.rse.2016.08.008
Finlay, S. E., Moffat, A., Gazzard, R., Baker, D., & Murray, V. (2012). Health impacts of wildfires. PLoS Currents, NOVEMBER 2012, 1–23. https://doi.org/10.1371/4f959951cce2c
Flinn, M. A., & Pringle, J. K. (1983). Heat tolerance of rhizomes of several understory species. Canadian Journal of Botany, 61(2), 452–457. https://doi.org/10.1139/b83-052
Forest Survey of India. (2011a). India State of Forest Report 2011. India State of Forest Report 2011, 236–240. https://doi.org/http://www.fsi.org.in/cover_2011/uttarakhand.pdf
Forest Survey of India. (2011b). India State of Forest Report 2011. In India State of Forest Report 2011: Vol. II. https://doi.org/http://www.fsi.org.in/cover_2011/uttarakhand.pdf
Fowler, C. (2003). Human Health Impacts of Forest Fires in the Southern United States: A Literature Review. Journal of Ecological Anthropology, 7(1), 39–63. https://doi.org/10.5038/2162-4593.7.1.3
Giriraj, A., Irfan-Ullah, M., Murthy, M. S. R., & Beierkuhnlein, C. (2008). Modelling spatial and temporal forest cover change patterns (1973-2020): A case study from South Western Ghats (India). Sensors, 8(10), 6132–6153. https://doi.org/10.3390/s8106132
González-Cabán, A. (2013). The economic dimension of wildland fires. Vegetation Fires and Global Change – Challenges for Concerted International Action. A White Paper Directed to the United Nations and International Organizations, January 2013, 229–237. http://treesearch.fs.fed.us/pubs/44383
Granström, A. (2001). Fire management for biodiversity in the European boreal forest. Scandinavian Journal of Forest Research, 16(1), 62–69. https://doi.org/10.1080/028275801300090627
Green, K., & Sanecki, G. (2006). Immediate and short-term responses of bird and mammal assemblages to a subalpine wildfire in the Snowy Mountains, Australia. Austral Ecology, 31(6), 673–681. https://doi.org/10.1111/j.1442-9993.2006.01629.x
Grutzen, P. J., & Andreae, M. O. (1990). Biomass burning in the tropics: Impact on atmospheric chemistry and biogeochemical cycles. Science, 250(4988), 1669–1678. https://doi.org/10.1126/science.250.4988.1669
Hemp, A. (2009). Climate change and its impact on the forests of Kilimanjaro. African Journal of Ecology, 47(SUPPL. 1), 3–10. https://doi.org/10.1111/j.1365-2028.2008.01043.x
Hussain, A., Negi, A. K., & Todaria, N. P. (2018). History of Forest Fire around Globe, India and Uttarakhand-A Brief Scenario. Journal of Agriculture and Technology, 5(1), 7–14.
Jain, T. B., Gould, W. A., Graham, R. T., Pilliod, D. S., Lentile, L. B., & González, G. (2008). A soil burn severity index for understanding soil-fire relations in tropical forests. Ambio, 37(7–8), 563–568. https://doi.org/10.1579/0044-7447-37.7.563
Jhariya, M. K., & Raj, A. (2014). Effects of wildfires on flora, fauna and physico-chemical properties of soil-An overview. Journal of Applied and Natural Science, 6(2), 887–897. https://doi.org/10.31018/jans.v6i2.550
Jhariya, M.K. S.S. Bargali, S.L. Swamy, B. Kittur, Kiran Bargali, and G. V. P. 2014. (2014). "Impact of Forest Fire on Biomass and Carbon Storage Pattern of Tropical Deciduous Forests in Bhoramdeo Wildlife Sanctuary, Chhattisgarh. June 2018.
Joseph, S., Anitha, K., & Murthy, M. S. R. (2009). Forest fire in India: A review of the knowledge base. Journal of Forest Research, 14(3), 127–134. https://doi.org/10.1007/s10310-009-0116-x
Khaki, B.A., Singh. V.R.R., Wani A.A. & thakur, R. K. (2015). Effects of forest fire on soil nutrients in blur pine (pinus wallichiana A.B. Jackson) ecosystems. Indian Forester, 141(4), 355–360.
Kim, M., Harris, T., & Alevy, J. (2011). Measuring Regional Economic Impacts from Wildfire : Case Study of Southeast Oregon Cattle-Ranching Business. January.
Kittur, B. H., Swamy, S. L., Bargali, S. S., & Jhariya, M. K. (2014). Wildland fires and moist deciduous forests of Chhattisgarh, India: divergent component assessment. Journal of Forestry Research, 25(4), 857–866. https://doi.org/10.1007/s11676-014-0471-0
Kodandapani, N., Cochrane, M. a., & Sukumar, R. (2004). Conservation Threat of Increasing Fire Frequencies in the Western Ghats, India. Conservation Biology, 18(6), 1553–1561. https://doi.org/10.1111/j.1523-1739.2004.00433.x
Kodandapani, N., Cochrane, M. A., & Sukumar, R. (2008). A comparative analysis of spatial, temporal, and ecological characteristics of forest fires in seasonally dry tropical ecosystems in the Western Ghats, India. Forest Ecology and Management, 256(4), 607–617. https://doi.org/10.1016/j.foreco.2008.05.006
Koltunov, A., Ustin, S. L., Quayle, B., Schwind, B., Ambrosia, V. G., & Li, W. (2016). The development and first validation of the GOES Early Fire Detection (GOES-EFD) algorithm. Remote Sensing of Environment, 184, 436–453. https://doi.org/10.1016/j.rse.2016.07.021
Krishnaswamy, J., Bawa, K. S., Ganeshaiah, K. N., & Kiran, M. C. (2009). Quantifying and mapping biodiversity and ecosystem services: Utility of a multi-season NDVI based Mahalanobis distance surrogate. Remote Sensing of Environment, 113(4), 857–867. https://doi.org/10.1016/j.rse.2008.12.011
Krüll, W., Tobera, R., Willms, I., Essen, H., & Von Wahl, N. (2012). Early forest fire detection and verification using optical smoke, gas and microwave sensors. Procedia Engineering, 45, 584–594. https://doi.org/10.1016/j.proeng.2012.08.208

Letnic, M., & Dickman, C. R. (2005). The responses of small mammals to patches regenerating after fire and rainfall in the Simpson Desert, central Australia. Austral Ecology, 30(1), 24–39. https://doi.org/10.1111/j.1442-9993.2004.01410.x

McKenzie, D., Gedalof, Z., Peterson, D. L., & Mote, P. (2004). Climatic change, wildfire, and conservation. Conservation Biology, 18(4), 890–902. https://doi.org/10.1111/j.1523-1739.2004.00492.x

Murthy, K. K., Sinha, S. K., Kaul, R., & Vaidyanathan, S. (2019). A fine-scale state-space model to understand drivers of forest fires in the Himalayan foothills. Forest Ecology and Management, 432(January), 902–911. https://doi.org/10.1016/j.foreco.2018.10.009

Narasimha Rao, G., Jagadeeswara Rao, P., Duvvuru, R., Bendalam, S., & Gemechu, R. (2018). Fire detection in Kambalakonda Reserved Forest, Visakhapatnam, Andhra Pradesh, India: An Internet of Things Approach. Materials Today: Proceedings, 5(1), 1162–1168. https://doi.org/10.1016/j.matpr.2017.11.197

Narendran, K., Indu, M. K., S., S. H., S., D. H., H., R. N., & R., S. (2001). Nontimber forest product extraction, utilization and valuation: a case study from the Nilgiri Biosphere reserve, Southern India. Economic Botany, 55(4), 528–538.

North, B. M. P., Stephens, S. L., Collins, B. M., Agee, J. K., Aplet, G., Franklin, J. F., & Fulé, P. Z. (2015). Reform forest fire management. Science 349(6254):1280–1281. Science.

Oliva, P., & Schroeder, W. (2015). Assessment of VIIRS 375m active fire detection product for direct burned area mapping. Remote Sensing of Environment, 160, 144–155. https://doi.org/10.1016/j.rse.2015.01.010

Parashar, B. S. (2003). The Impact Of Forest Fire On Forest Biodiversity In The Indian Himalayas (Uttaranchal). Proceeding of XII world Forestry Congress, Canada 0358-B1.

Paula, S., Arianoutsou, M., Kazanis, D., Tavsanoglu, Ç., Lloret, F., Buhk, C., Ojeda, F., Luna, B., Moreno, J. M., Rodrigo, A., Espelta, J. M., Palacio, S., Fernández-Santos, B., Fernandes, P. M., & Pausas, J. G. (2009). Fire-related traits for plant species of the Mediterranean Basin. Ecology, 90(5), 1420. https://doi.org/10.1890/08-1309.1

Pelegrin, N., & Bucher, E. H. (2010). Long-term effects of a wildfire on a lizard assemblage in the Arid Chaco forest. Journal of Arid Environments, 74(3), 368–372. https://doi.org/10.1016/j.jaridenv.2009.09.009

Penner, J. E., Dickinson, R. E., & O'Neill, C. A. (1992). Effects of aerosol from biomass burning on the global radiation budget. Science, 256(5062), 1432–1434. https://doi.org/10.1126/science.256.5062.1432

Piñol, J., Terradas, J., & Lloret, F. (1998). Climate warming, wildfire hazard, and wildfire occurrence in coastal eastern Spain. Climatic Change, 38(3), 345–357. https://doi.org/10.1023/A:1005316632105

Pitman, A. J., Narisma, G. T., & McAneney, J. (2007). The impact of climate change on the risk of forest and grassland fires in Australia. Climatic Change, 84(3–4), 383–401. https://doi.org/10.1007/s10584-007-9243-6

Producer, S., Books, C., & Nature, O. N. (2006). Hotspots revisited. Choice Reviews Online, 43(05), 43-2786-43–2786. https://doi.org/10.5860/choice.43-2786

Puri, K., Areendran, G., Raj, K., Mazumdar, S., & Joshi, P. K. (2011). Forest fire risk assessment in parts of Northeast India using geospatial tools. Journal of Forestry Research, 22(4), 641–647. https://doi.org/10.1007/s11676-011-0206-4

Ramakrishnan, P. S. (2007). Traditional forest knowledge and sustainable forestry: A north-east India perspective. Forest Ecology and Management, 249(1–2), 91–99. https://doi.org/10.1016/j.foreco.2007.04.001

Rodgers, W. A. (1986). The Role of Fire in the Management of Wildlife Habitats: A Review'. Indian Forester, 112: 845-857. Indian Forester, 112(10).

Saha, S. (2002). Anthropogenic fire regime in a deciduous forest of central India. Current Science, 82(9), 1144–1147.

Singh, J. S., Rawat, Y. S., & Chaturvedi, O. P. (1984). Replacement of oak forest with pine in the Himalaya affects the nitrogen cycle. Nature, 311(5981), 54–56. https://doi.org/10.1038/311054a0

Soliman, H., Sudan, K., & Mishra, A. (2010). A smart forest-fire early detection sensory system: Another approach of utilizing wireless sensor and neural networks. Proceedings of IEEE Sensors, January 2017, 1900–1904. https://doi.org/10.1109/ICSENS.2010.5690033

Stephens, S. L. (2005). Forest fire causes and extent on United States Forest Service lands. International Journal of Wildland Fire, 14(3), 213–222. https://doi.org/10.1071/WF04006

Úbeda, X., & Sarricolea, P. (2016). Wildfires in Chile: A review. Global and Planetary Change, 146, 152–161. https://doi.org/10.1016/j.gloplacha.2016.10.004

Verma, S., & Jayakumar, S. (2012). Impact of forest fire on physical, chemical and biological properties of soil: A. Proceedings of the International Academy of …, 2(3), 168–176. http://www.iaees.org/publications/journals/piaees/articles/2012-2(3)/impact-of-forest-fire.pdf

Vidal-Macua, J. J., Ninyerola, M., Zabala, A., Domingo-Marimon, C., Gonzalez-Guerrero, O., & Pons, X. (2018). Environmental and socioeconomic factors of abandonment of rainfed and irrigated crops in northeast Spain. Applied Geography, 90(September 2017), 155–174. https://doi.org/10.1016/j.apgeog.2017.12.005

Wilson, A. M., Latimer, A. M., Silander, J. A., Gelfand, A. E., & de Klerk, H. (2010). A Hierarchical Bayesian model of wildfire in a Mediterranean biodiversity hotspot: Implications of weather variability and global circulation. Ecological Modelling, 221(1), 106–112. https://doi.org/10.1016/j.ecolmodel.2009.09.016

Wong, A. K. K., & Fong, N. K. (2014). Experimental study of video fire detection and its applications. Procedia Engineering, 71, 316–327. https://doi.org/10.1016/j.proeng.2014.04.046

World Bank, 2018. (2018). Strengthening Forest Fire Management in India". World Bank, Washington DC. www.fsi.nic.in
World Bank. (2005). India Unlocking Opportunities for Forest-Dependent People in India. I(34481).
Wotton, B. M., Martell, D. L., & Logan, K. A. (2003). Occurrence in Ontario. Climatic Change, 275–295.
Wotton, B. M., Nock, C. A., & Flannigan, M. D. (2010). Forest fire occurrence and climate change in Canada. International Journal of Wildland Fire, 19(3), 253–271. https://doi.org/10.1071/WF09002
WWF, I. (2011). Forests for a living planet. WWF Living Forests Report, 1–37. http://www.wwf.se/source.php/1359561/Living Forests Report_Chapter 1_2011.pdf
Zhang, D., Han, S., Zhao, J., Zhang, Z., Qu, C., Ke, Y., & Chen, X. (2009). Image based Forest fire detection using dynamic characteristics with artificial neural networks. IJCAI International Joint Conference on Artificial Intelligence, 2, 290–293. https://doi.org/10.1109/JCAI.2009.79
Zhang, Q. X., Lin, G. H., Zhang, Y. M., Xu, G., & Wang, J. J. (2018). Wildland Forest Fire Smoke Detection Based on Faster R-CNN using Synthetic Smoke Images. Procedia Engineering, 211, 441–446. https://doi.org/10.1016/j.proeng.2017.12.03

Proceedings of the Fifth International Conference on Trends in Electronics and Informatics (ICOEI).
IEEE Xplore Part Number:CFP21J32-ART; ISBN:978-1-6654-1571-2

FPGA Implementation of a pipelined and pseudo-randomized TDES algorithm

Arnab Roy
Dept. of Electronics and Communication Engineering
Delhi Technological University
Delhi,India
arnab6749@gmail.com

Anubhav Gadhok
Dept. of Electronics and Communication Engineering
Delhi Technological University
Delhi,India
anubhav2199@gmail.com

Deva Nand
Dept. of Electronics and Communication Engineering
Delhi Technological University
Delhi,India
devkamboj07@gmail.com

2021 5th International Conference on Trends in Electronics and Informatics (ICOEI) | 978-1-6654-1571-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICOEI51242.2021.9452952

***Abstract*—In today's world where data breaching and misuse of confidential data are so common, an end user's utmost priority should be to keep one's data secure. This is where information security and cryptography comes into the picture and in this paper, a pipelined and pseudo randomized implementation of TDES encryption standard on FPGA using Verilog HDL is designed which can encrypt and decrypt a given user's messages with high throughput and the pseudo randomizer randomizes the key with the help of a Linear Feedback Shift Register which makes the encrypted output random thereby making it more secure than the traditional TDES systems. The target device used is xc7vx330t-3ffg1157 from Virtex-7 FPGA family offered by Xilinx on which a maximum frequency of 608.224 MHz and maximum throughput of 38.92 Gigabits per second was achieved. The proposed design is also faster when compared to previous works on their respective platforms. The proposed system was designed and analyzed thoroughly using the Xilinx ISE Design Suite v14.7.**

***Keywords*—Cryptography,Encryption,TDES,FPGA,Pipelining, Throughput,Information Security**

I. Introduction

With the approach of innovation, the information which is sent and put away in advanced world is expanding step by step. On the other hand, this data is also at a high risk of being accessed unethically. The threat to data security continues to increase exponentially as new types of advanced attacks on data are emerging regularly. To prevent all these mishaps, data needs to be protected so that it can be accessed by only those people who are authorized to do so. This is where encryption comes into picture. Encryption is the process of converting data into some other form which is unrecognizable by unauthorized or normal people and the reverse process is called Decryption [1]. The data is converted into a form which is unreadable by normal people and only the authorized people can convert this unreadable form into readable form. A lot of encryption methodologies are used in the current digital world like DES, TDES, AES, RSA, AES, ECC, Blowfish, Twofish, Threefish, RC5, IDEA [2] [3]. All these algorithms have their own advantages and disadvantages which decide their usage in real time applications.

This paper implements a pseudo randomized pipelined version of TDES in ECB mode with a modified key generation function to make it more secure and efficient than the regular TDES. In order to implement TDES [4], DES algorithm [5] is implemented first followed by introduction of pipelined architecture into the existing algorithm [6] [7] which in return makes the encryption and decryption process faster than before with higher throughput. In order to improve the security, a modified key generation algorithm is employed which consists of sequence of pseudo random numbers generated by Linear Feedback Shift Registers [8] [9] masking the generated key thereby producing a pseudo randomized enciphered output which is completely foreign for an unauthorized user. On the other hand, an authorized user who has access to the master key as well as the seed to the LFSR can safely access the data. These modified DES blocks (both encryption and decryption) are then used to construct a pipelined version of TDES which is faster and more secure than many of the current DES and TDES based systems [10] [11] [12].Cryptography has come a long way over years [13] and in the present world this system can be used in e-governance and banking security [14], smart card applications [15], cloud security [16], etc. The paper organization is as follows: Section II describes the methodology used to design the pipelined DES block. Section III explains the modified key generation employed in this paper which helps in the increasing security of the proposed system. Section IV elucidates the detailed implementation of the final design. Detailed comparisons, relevant results and conclusions are drawn in the final sections of the paper.

II. Pipelined DES Implementation

The Data Encryption Standard (DES) is a symmetric key encryption algorithm which is used for encryption and protection of computerized information. Despite the fact that its key length is short (56 bits) which makes it highly insecure for applications, it has been profoundly influential in the progression of cryptography. In order to implement a pipelined TDES design , it is necessary to construct a coherent pipelined DES algorithm first. The complete design and description of the DES algorithm is presented in [5]. This section will show the pipelining introduced in the existing DES algorithm.

Pipelining increases the throughput of the system by means of parallel computation [7]. Instead of processing intermediate data one by one, pipelining helps in processing intermediate

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:31:16 UTC from IEEE Xplore. Restrictions apply.

data within a system simultaneously. This leads to increase in data processed per unit time which is better known as throughput. A 16 stage pipeline as shown in Figure 1 is employed in this paper which means that each feistel round will act as one stage of pipelining. Registers are added after every round which act as output of the previous round for the current clock cycle and input for the next round in the next clock cycle. One register is also added after the final permutation which gives output only when the system is enabled. The system takes 16 clock cycle latency for the initial result and then it encrypts 64 bits of data per clock cycle moving forward. Each feistel round is taken care of with a

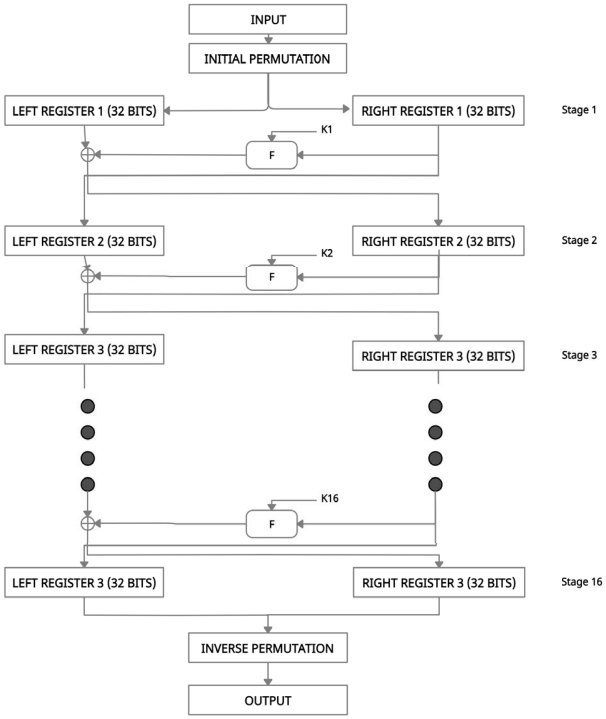


Fig. 1. Pipelining in DES

subkey which is obtained by left shift operations in the 56 bit new key obtained from the original key (64 bit) only by dropping off few bits. The procedure for generation of 16 sub keys is described in detail in [5]. The sub key generation process is also pipelined in accordance with the pipelining in the feistel rounds which can be seen in Figure 2. Each sub key is subject to be stored in registers after the corresponding left shift operation pertaining to the concerned round. These stored values are passed on to next stages to ensure a smooth pipelined design.

Both, the encrypting and decrypting blocks for DES are same as before but the only difference between the DES encrypt and DES decrypt schemes are the sequence in which the 16 sub-keys generated are fed to the Feistel rounds. The sequence in which sub keys are inserted in decryption is exactly reverse of the order in encryption. In non pipelined decryption of DES one may generate all the sub keys first and then apply them in reverse order i.e K16 to round 1, K15 to round 2 and so on. But K16 is generated only after all other sub keys are generated thereby disturbing the timing of the feistel rounds. So the traditional method of left shift operations to generate sub keys will not fit into the pipelined architecture of DES decryption. In order to generate sub keys in reverse order i.e K16, K15, K14 and so on, this paper performs right shift operations on the 56 bit initial key obtained from the user instead of left shift operations. Depending on the round the number of bits by which the keys have to be shifted will change. The number of bits shifted right in the decryption cycle is same as the number of bits shifted to the left in the encryption for the corresponding rounds with just round 1 being an exception where there is no shifting of the bits.

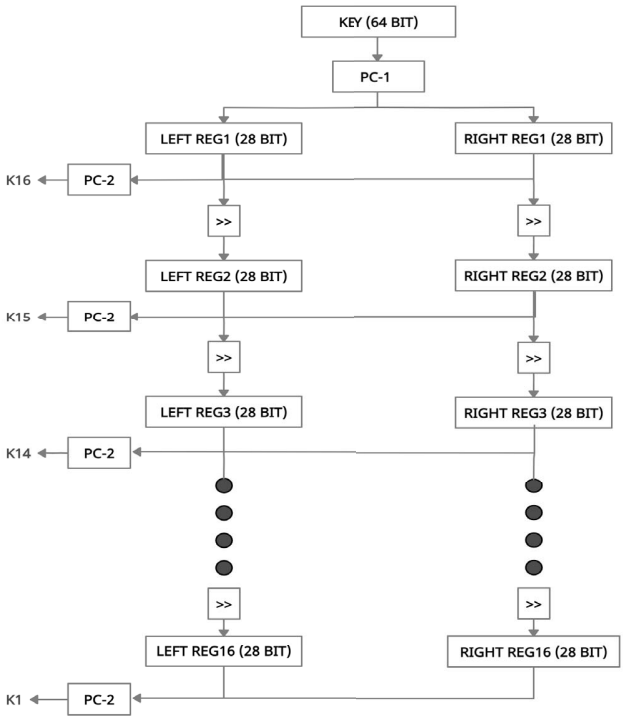


Fig. 2. Pipelined sub key generation in DES decryption

III. Pseudo-randomized Key generation

A. Linear Feedback Shift Registers

Linear Feedback Shift Registers are widely used in cryptographic applications, digital broadcasting , generation of counters, VLSI testing etc. A linear feedback shift register (LFSR) is a configuration of shift registers in which the input bit has a linear dependence on the previous stage [8] [9]. The XOR operation is the most widely used linear function. The input bit of the Linear Feedback Shift Register is the resultant of the XOR operation of few selected bits of the previous state values of the overall LFSR. Firstly the LFSR is loaded with an initial seed value which can be anything except all 0s. Then the LFSR is clocked and enabled i.e. when the input

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:31:16 UTC from IEEE Xplore. Restrictions apply.

message is changed and the edge (positive or negative) of the clock is encountered, it starts shifting data sequentially and performs XOR operation with the current value at the flip flops specified resulting in a pseudo random pattern of 1s and 0s. The few selected bits which are used in XOR operation define the characteristic polynomial of the LFSR. For example in Figure 3 the characteristic polynomial is

$$\phi(x) = x^4 + x + 1 \tag{1}$$

x^4 + 1 term indicates the presence of the loop and rest of the terms define the bits which are involved in the XOR operation (D1 is involved here so x^1 is the additional term). For LFSRs with large number of bits, the characteristic polynomial proves to be useful for representation and analysis as well. The bigger the LFSR is used, bigger is the length of pseudo random sequence thereby making it highly unpredictable and very secure. The pseudo randomizer used in this paper is achieved with the help of a 56 bit Linear Feedback Shift Register. The characteristic polynomial for the LFSR used in this paper is

$$\phi(x) = x^{56} + x^{48} + x^{36} + x^{24} + x^{12} + 1 \tag{2}$$

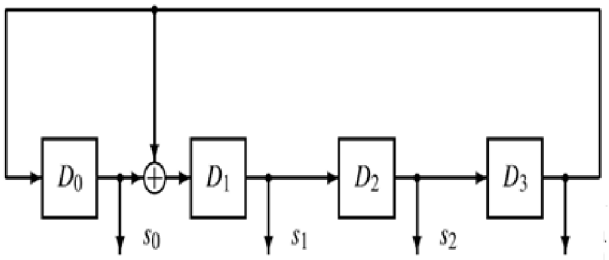


Fig. 3. A 4 bit LFSR

B. Implementation of the modified key generator

The aim of randomizing the subkeys is to increase the security. If the key is kept same for a stream of message bits then the security can be breached easily. In order to make these attacks by hackers face difficulty, it is ensured that for each different 64 bit message the key must be different and random. The original 56 bit key and the output of 56 bit LFSR are put under a specific function in order to get a new and random key from which subkeys are generated for Feistel rounds. The process is described below and is also depicted in Figure 4.

- Obtain the 56 bit key (Koriginal) by dropping off 8 select bits from 64 bit key(Kuser)
- Knew,i= f(Koriginal, Ri) (where Ri is ith 56 bit random number generated by LFSR)
- The function f used applies the following conditions for output:
  - If Ri[n] is 1 then complement the nth bit of Koriginal
  - If Ri[n] is 0 then retain the nth bit of Koriginal
- The obtained Knew,i is used to generate the subkeys for the feistel rounds

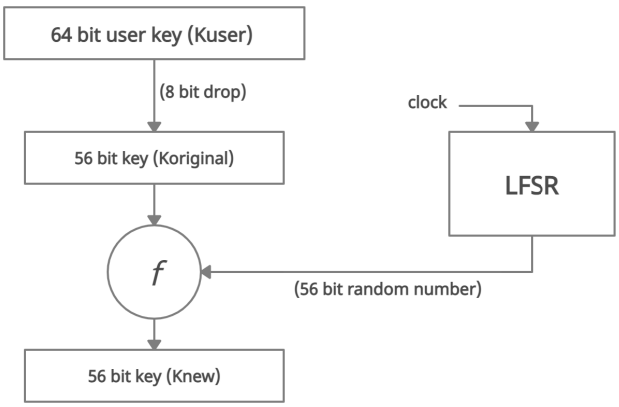


Fig. 4. Key randomization process

The value of LFSR changes with every change detected in message. Usually a stream of bits is sent over a communication system and hence after every clock the LFSR value is subject to change. The change in LFSR after every clock cycle also changes the new key to be used which in turn reduces the correlation between two consecutive encryption cycles thus making it more secure. The LFSR operation can be understood with the help of flowchart shown in Figure 5 and pseudocode discussed below:

Always @(posedge clock)
If(reg1 == plaintext) // plaintext is same as previous one
Lfsr_enable ⇐ 0; // lfsr value remains same
Else begin // plaintext changes
Lfsr_enable ⇐ 1; // lfsr value changes
Reg1 ⇐ plaintext; // new value of plaintext loaded for future comparisons

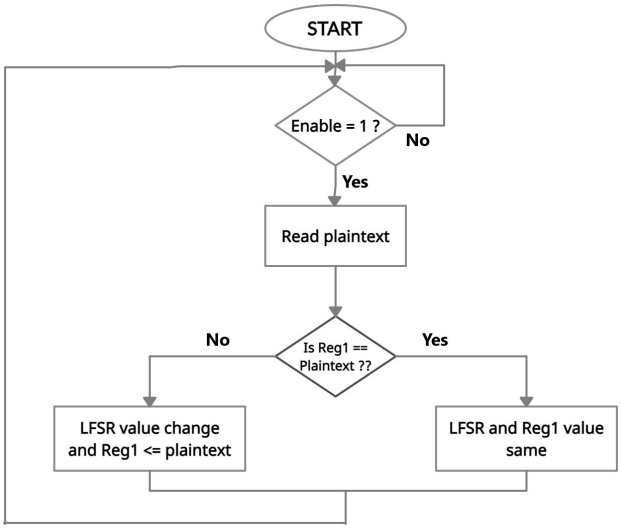


Fig. 5. Flowchart for LFSR operation.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:31:16 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Trends in Electronics and Informatics (ICOEI).
IEEE Xplore Part Number:CFP21J32-ART; ISBN:978-1-6654-1571-2

IV. Pseudo-Randomized Pipelined TDES

A. Encryption and Decryption

The TDES encryption process is pretty simple once DES algorithm is thoroughly understood. It is the connection of a DES encrypt block followed by a DES decrypt block then another DES encrypt block in series. With the help of structural coding practices in Verilog one can easily instantiate the DES blocks three times and connect them using wires. The objetive behind using three DES blocks is to enhance the security by increasing the key length and overcome the drawback of short key length in single DES system. The input keys are also connected to the three blocks using the key distribution process shown in Figure 6. The DES blocks which are instantiated here are the ones with pseudo randomized key generation block. The middle block uses an inverted primary clock (negative edge triggered with respect to primary clock) whereas the start and the end block use primary clock only(positive edge triggered with respect to primary clock). This is done to provide half clock cycle span for modified key function to generate its values correctly. Therefore the TDES discussed in this paper gives its first output after 16+ 0.5 +16 + 0.5 +16 = 49 clock cycles and thereafter gives the encrypted output after every clock cycle. In TDES there are three DES blocks, so in order to further improve the security user can feed different seed values to the three LFSRs to reduce the correlation between the intermediate processes.The TDES decrypting process is just the opposite of the TDES encrypting process. Therefore, a DES decrypt block followed by a DES encrypt block then another DES decrypt block are used for decryption. Since TDES is also a symmetric key algorithm it uses the same keys as the encryption process.

B. Key Management

A 192-bit key is provided by the user which is split up into 3 keys whose sizes are 64 bit each which act as the keys for each of the 3 DES blocks. Once again the problem encountered here is the lack of orderliness. To get over it the method so employed is key buffering i.e. use of buffers to synchronize the sets of 64 bit keys in the orientation as intended so that correct result is obtained. The first 64 bit key is fed into the first DES block as soon as it is given by the user. The next 64 key bit key is given to the second DES block after a 16 clock cycles buffer so that the first block can complete its computation since it takes an initial 16 clock latency. Lastly the third 64 bit key is fed to the third DES block after 33 clock cycles buffer so that the computations of first and second block is already complete when the third block's computation initiates. This ensures that if key changes at a particular instant, then the data which is already being processed is not fed with wrong keys and hence the results are not affected. Figure 6 shows the implementation of the TDES algorithm proposed in this paper.

V. Results and Discussion

A. Simulation Results

The simulations have been carried out in Xilinx ISE simulator better known as ISIM and are shown in Figure 7. The

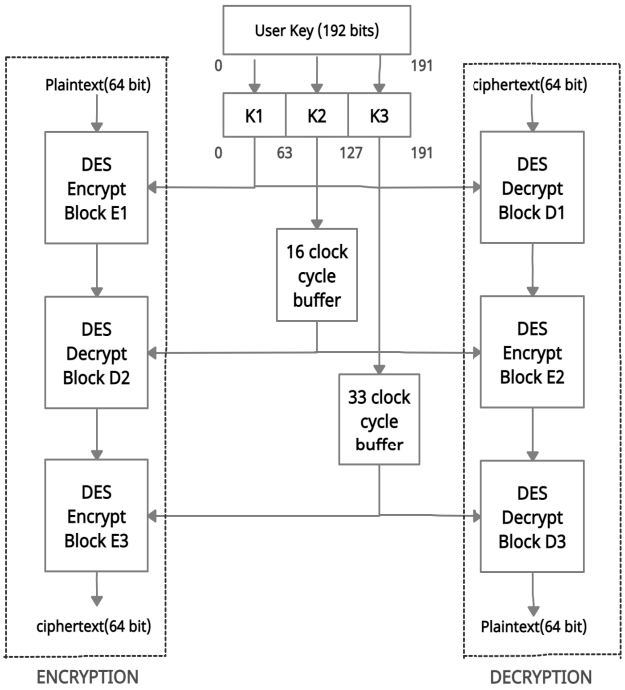


Fig. 6. TDES Block Diagram

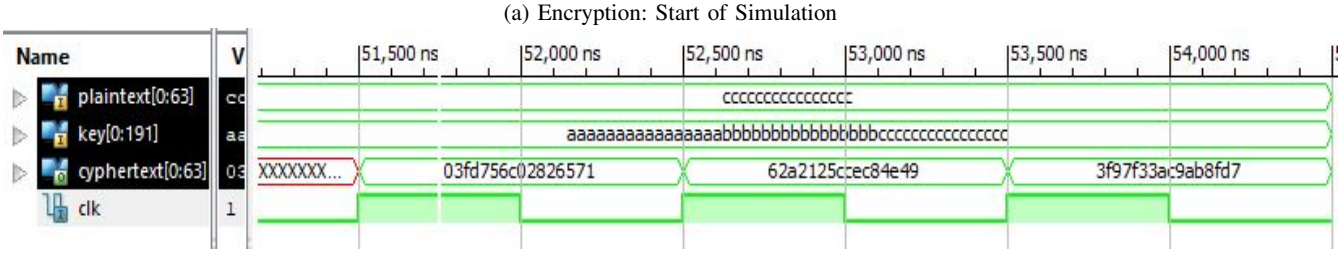
message and master key are provided to the proposed design by the user in hexadecimal format at the start of the simulation. Three plaintexts are applied consecutively after each clock cycle. The sequence of plaintext applied is aaaaaaaaaaaaaaaa (1st clock) ,bbbbbbbbbbbbbbbb(2nd clock) , ccccccccccccccc(3rd clock). The 192 bit key is also fed which is "aaaaaaaaaaaaaaaa-bbbbbbbbbbbbbbbb-ccccccccccccccc". These plaintexts will be encrypted as per the TDES algorithm proposed in this paper. The first encrypted ciphertext appears after an initial latency of 49 clock pulses there after encryption occurs after each clock pulse. The sequence of encrypted ciphertext is 03fd756c02826571 (49th clock) , 62a2125ccec84e49(50th clock), 3f97f33ac9ab8fd7(51st clock). In order to verify decryption scheme of the proposed TDES algorithm, the ciphertext sequence is fed into the decrypt module in the same order i.e 03fd756c02826571 (1st clock) , 62a2125ccec84e49(2nd clock), 3f97f33ac9ab8fd7(3rd clock)with the same key which is fed in encryption. After 49 clock pulses the initial sequence of plaintext applied at the encryption is obtained back at the output i.e aaaaaaaaaaaaaaaa (1st clock) ,bbbbbbbbbbbbbbbb(2nd clock) , ccccccccccccccc(3rd clock).

B. Synthesis Results

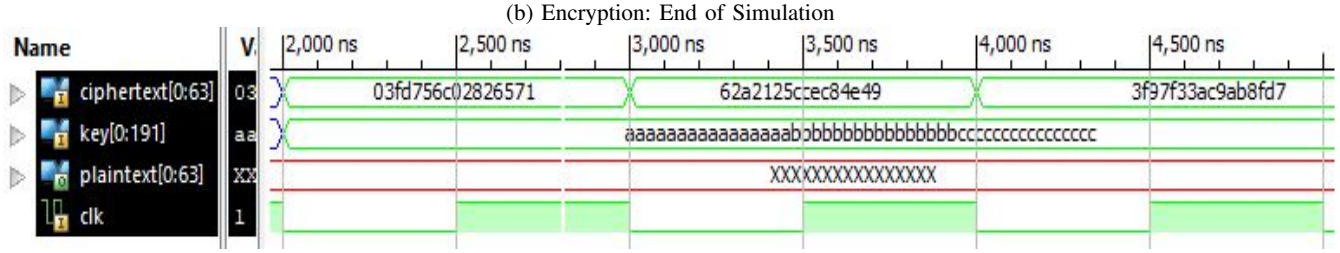
In order to observe the improvements attained in the design, the non-pipelined and pipelined implementations of TDES were also done on the same device i.e xc7vx330t-3ffg1157. It is evident from Table I that both the pipelined and pseudo randomized TDES implementations are faster than the non

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:31:16 UTC from IEEE Xplore. Restrictions apply.

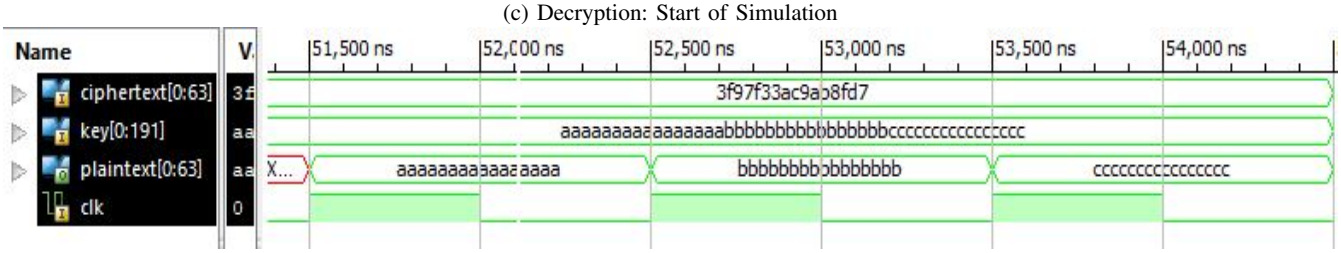
(a) Encryption: Start of Simulation



(b) Encryption: End of Simulation



(c) Decryption: Start of Simulation



(d) Decryption: End of Simulation

Fig. 7. Simulation Results

pipelined implementation but the resources used are more in the pipelined versions. Between pseudo randomized TDES and simple pipelined TDES, the latter is faster by a marginal amount which can be afforded as pseudo randomized TDES offers much better security. The resources used in all the three implementations are also well within the limits of the device used and further changes, if required can be made in the design without any difficulty.

TABLE I. Comparison between Non pipelined , Pipelined and pseudo randomized TDES

Type of Encrytion Used	Non-pipelined TDES	Pipelined TDES	Pseudo-randomized pipelined TDES
Maximum Frequency(MHz)	11.89	640.16	608.22
Maximum Throughput(Mbps)	761	40,970	38,926
Slice Registers Used	0	6120	6853
Slice LUTs used	8322	5377	6220
Delay(ns)	84.1	1.56	1.644

TABLE II. Comparison with previous works

Device Used	Authors	Maximum Frequency (MHz)	Maximum Throughput (Mbps)
xc4vlx25-12-ff676 (Virtex-4)	Fang Ren et al. [10]	215.165	860.66
	Proposed work	235.673	15083.072
xc7a100t-3csg324 (Artix-7)	Guruprasad S.P et al. [11]	348.947	797.593
	Proposed Work	391.654	25065.856
xc6vlx240t-3ff1156 (Virtex-6)	Saeid Taherkhani et al. [12]	294.031	18817.98
	Proposed work	566.370	36247.68

In order to reassure the fastness of the pseudo randomized TDES algorithm discussed in this paper, the speed parameters of the proposed work were compared with the previous works discussed in [10] [11] and [12]. The proposed design was implemented on the same device with which these works have been carried out and the results are shown in Table II. It is observed that proposed design is faster than the previous works which is depicted in Figure 8. It is also seen that the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:31:16 UTC from IEEE Xplore. Restrictions apply.

work presented in this paper gives tremendous increase in throughput (processes more data in less amount of time) as compared to all the works which have been taken for reference as can be seen in Figure 9. This is because the proposed work will process each batch of 64 bit data after every single clock cycle whereas the previous works [10] and [11] take multiple clock cycles to process a single batch of 64 bit data. It is also worthy to note that all these improvements in the proposed design are coming along with improvement in security due to the pseudo randomized key generation.

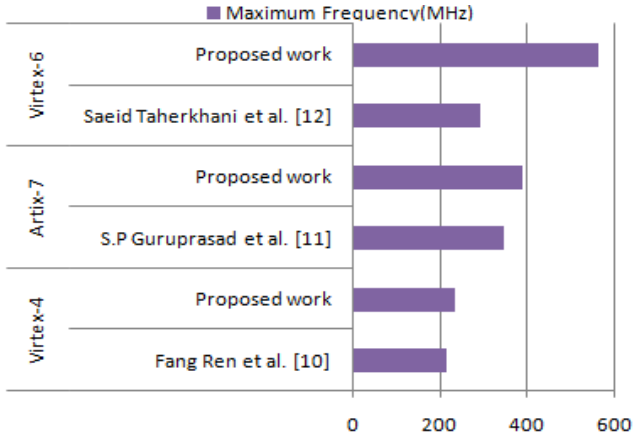


Fig. 8. Graphical representation of comparison of maximum frequency with previous works

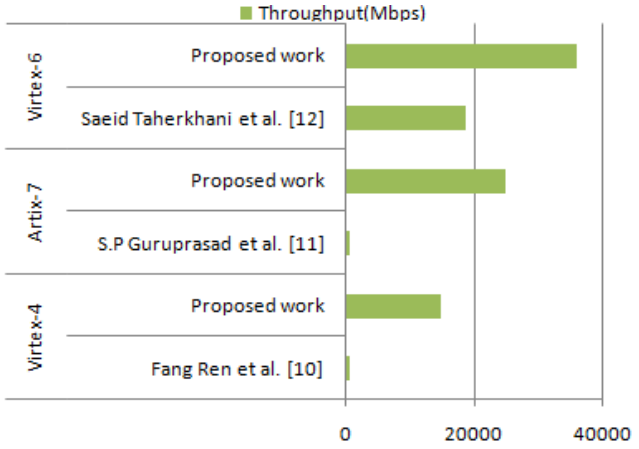


Fig. 9. Graphical representation of comparison of maximum throughput with previous works

VI. Conclusion

A highly efficient and secure implementation of TDES algorithm is presented in this paper. The proposed work offers enhanced speeds with the help of the pipelining explained in the previous sections and is way secure than any traditional TDES system because of its modified key generation which employs a pseudo randomizer (with the help of an LFSR) which makes the encrypted output random thereby increasing the sense of security. The proposed system is faster and provides enhanced throughput when compared to previous works [10] [11] and [12] on respective devices. The design is implemented on the device xc7vx330t-3ffg1157 (Virtex – 7) using which maximum frequency of operation achieved is 608.224 MHz and at the maximum throughput so achieved is 38.926 Gbps. This variant of TDES encryption can be used in various real time applications like electronic payment systems, SMS encryption, cloud security, smart card applications etc. because of its increased sense of security and speed.

References

[1] M. E. Smid, D. K. Branstad, “Data Encryption Standard: past and future,” in Proceedings of the IEEE, vol. 76, no. 5, pp. 550-559, May 1988
[2] R. Bhanot, R. Hans, “A Review and Comparative Analysis of Various Encryption Algorithms,” International journal of security and its applications, vol. 9, no. 4, 2015, pp. 289-306
[3] A. Nadeem, M. Y. Javed, “A Performance Comparison of Data Encryption Algorithms,” 2005 International Conference on Information and Communication Technologies, Karachi, Pakistan, 2005, pp. 84-89.
[4] E. Barker, N. Mouha, “Recommendation for the Triple Data Encryption Algorithm (TDEA) Block Cipher,” NIST Special Publication, National Institute of Standards and Technology, Gaithersburg, 2017, pp. 800-867.
[5] Data Encryption Standard (DES), National Institute of Standards and Technology, Gaithersburg, 1999
[6] E. D. Rosal, S. Kumar, “A Fast FPGA Implementation for Triple DES Encryption Scheme,” Circuits and Systems, vol. 8, 2017, pp. 237-246.
[7] S. Kilts, “Advanced FPGA Design: Architecture, Implementation and Optimization”, Wiley-IEEE Press, June 2007, pp. 1-16.
[8] A. K. Panda, P. Rajput, B. Shukla, “FPGA Implementation of 8, 16 and 32 Bit LFSR with Maximum Length Feedback Polynomial Using VHDL,” 2012 International Conference on Communication Systems and Network Technologies, Rajkot, Gujarat, India, 2012, pp. 769-773.
[9] A. Ajane, P. M. Furth, E. E. Johnson, R. L. Subramanyam, “Comparison of binary and LFSR counters and efficient LFSR decoding algorithm,” IEEE 54th International Midwest Symposium on Circuits and Systems (MWSCAS), Seoul, Korea (South), 2011, pp. 1-4.
[10] F. Ren, L. Chen., T. Zhang, “3DES Implementation Based on FPGA,” Emerging Research in Web Information Systems and Mining, 2011, pp. 218–224.
[11] S. P. Guruprasad, B. S. Chandrasekar, “An evaluation framework for security algorithms performance realization on FPGA,” 2018 IEEE International Conference on Current Trends in Advanced Computing (ICCTAC), Bangalore, India, 2018, pp. 1-6.
[12] S. Taherkhani, E. Ever, O. Gemikonakli, “Implementation of Non-Pipelined and Pipelined Data Encryption Standard (DES) Using Xilinx Virtex-6 FPGA Technology,” 2010 10th IEEE International Conference on Computer and Information Technology, Bradford, UK, 2010, pp. 1257-1262.
[13] K. Chitra, V. P. Venkatesan. ”An antiquity to the contemporary of secret sharing scheme,” Journal of Innovative Image Processing (JIIP) ,vol. 2, no. 1, 2020, pp. 1-13.
[14] I. R. S. Reddy and G. Murali, ”A novel triple DES to enhance E-governance security,” 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, India, 2017, pp. 2443-2446
[15] S. Shrivastava, S. Palnitkar, S. Pathak, M. Shinde, ”Secure Encryption of data using Symmetric and Asymmetric Cryptographic Algorithm,” Mukt Shabd Journal, vol. 9, no. 6, June 2020, pp. 1100-1109
[16] S. Kaur and L. Jain, “A Hybrid Cryptographic Scheme for Improving Cloud Security Using ECC and TDES Algorithms”, Current Journal of Applied Science and Technology, vol. 39, no. 47, pp. 27-34, December 2020.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:31:16 UTC from IEEE Xplore. Restrictions apply.

FPGA Implementation of CNN based Fingertip Digit Recognition

MD Shahbaz Khan, Niharika, Priya Yadav, Rishabh Verma, Indu Sreedevi

Abstract*: The paper proposes a novel framework for fingertip digit recognition using Convolutional Neural Network (CNN) and its hardware implementation on FPGA. In the above-mentioned framework, tracking of the fingertip is done using Learning Adaptive Discriminative Correlation Filter (LADCF) algorithm using C++. From the tracked image, digit recognition is done using CNN. CNN model was trained on MNIST dataset of 50000 images of handwritten digits and tested on 10000 such samples with the accuracy rate of 98.4% while recognising an unseen digit. The predicted digit was displayed on seven-segment using SystemVerilog. As per the outcomes of VOT 2018 challenge, LADCF based tracking algorithm has an accuracy of 90.7%. The proposed algorithm is fast in processing as it uses LADCF based fingertip tracking and CNN for digit recognition. Existing works in this field uses image classification tree to recognise digits[12], whose accuracy is 95.8% which is much less than the proposed algorithm in this paper.

Keywords* : CNN, Fingertip-based digit recognition, FPGA implementation, image processing.

I. INTRODUCTION

Computer vision finds its wide application in day-to-day life. Character recognition is an important task of computer vision algorithm. With the progress in algorithms, computational efficiency is increasing manifold thus making human-computer interface smooth, faster and more precise. Advancement in machine learning algorithms has also developed few computational efficient computer vision algorithms and one such is Convolutional Neural Network (CNN). Fingertip based gesture system plays a vital role in the life of physically disabled people. Using fingertip gestures, one can communicate in the vacuum without any medium of writing. Such systems should be computationally faster with higher precision. The paper proposes a noble methodology of fingertip based digit recognition. Using camera input, video of the movement of a fingertip is recorded using SystemVerilog and C++ language. On applying the image processing algorithms in C++ language, the gesture of digit is extracted from video into one frame. That frame is further applied on a trained model for digit recognition. The digit predicted by the CNN based algorithm is displayed on seven Segment display, connected to the FPGA board.

Revised Manuscript Received on April 25, 2020.

***** Correspondence Author

MD Shahbaz Khan*, Dept. of Electronics & Communication Engineering, Delhi Technological University, New Delhi, India. Email: mdshahbaz_bt2k16@dtu.ac.in

Niharika, department, Dept. of Electronics & Communication Engineering, Delhi Technological University, New Delhi, India. Email: niharika_bt2k16@dtu.ac.in

Priya Yadav, Dept. of Electronics & Communication Engineering, Delhi Technological University, New Delhi, India. Email: priyayadav_bt2k16@dtu.ac.in

Rishabh Verma, Dept. of Electronics & Communication Engineering, Delhi Technological University, New Delhi, India. Email: rishabhverma18@gmail.com

Dr. S. Indu, Dept. of Electronics & Communication Engineering, Delhi Technological University, New Delhi, India. Email: sindu@dtu.ac.in

II. LITERATURE LAYOUT

Several algorithms have come up to make fingertip-based digit recognition faster. However, accuracy for recognising digits is still low. Algorithms like [12], has an accuracy of 95% only, as compared to ours of 98.4%. Some of the algorithms [2], [3] uses image processing and decision tree based digit recognition, which are computationally less efficient with lower accuracy. In contrary to this, our algorithm uses FPGA implementation of CNN which makes it much faster as compared to the algorithms running on CPUs or GPUs.

Paper like [7] uses R-CNN based algorithm for fingertip tracking. However as per the result of VOT2018 challenge [21], LADCF tracking algorithm out throws the other tracking algorithms. It has an efficiency of 90.7% for tracking an object in continuous frames at a fps of 50. Along with LADCF based fingertip tracking, the paper proposes CNN based handwritten digit recognition system. CNN model for digit recognition is trained on 50,000 handwritten images and been tested on 10,000 images which yielded an accuracy of 98.4%. The paper deals in hardware implementation using FPGA board of digit recognition using CNN. As per the white paper published by Microsoft [5], training of any neural network on FPGA is much faster than the its training on CPUs and GPUs.

III. PROBLEM STATEMENT

The paper is aimed to find the solution for the fingertip digit recognition system. The intent is to implement the CNN based recognition system on FPGA board as to increase its computational efficiency. Earlier work in this field [13] was carried out using Python language on CPUs, which are not as fast as FPGAs. The proposed algorithm also enhances the accuracy of digit prediction as it uses CNN for handwritten digit recognition. Works like [11], [12] deals in decision tree and SVM based classifier for handwritten digit recognition, whose accuracy is not as high as CNN based algorithm. Moreover, on implementing algorithm on FPGA, training speed of model also increases manifold.

IV. EXTRACTING FOREGROUND

The very step in the field of fingertip tracking is the removal of background from each frame. Background subtraction is an important image pre-processing technique to extract the region of interest. This preprocessing of the image is mainly done to subtract the foreground from the background. For the abovementioned task, background segmentation algorithm (based on Gaussian Mixture) was used. This algorithm is based on two research papers [6],[9] given in the year 2004 and 2006 respectively. For each pixel, the algorithm picks the Gaussian distributions by itself. It makes the algorithm flexible for diurnal changes. Moreover, it proves to efficient for shadow detection and its removal from the foreground. This background subtraction algorithm makes the whole framework independent of diurnal changes and shadow and hence enhances its efficiency.

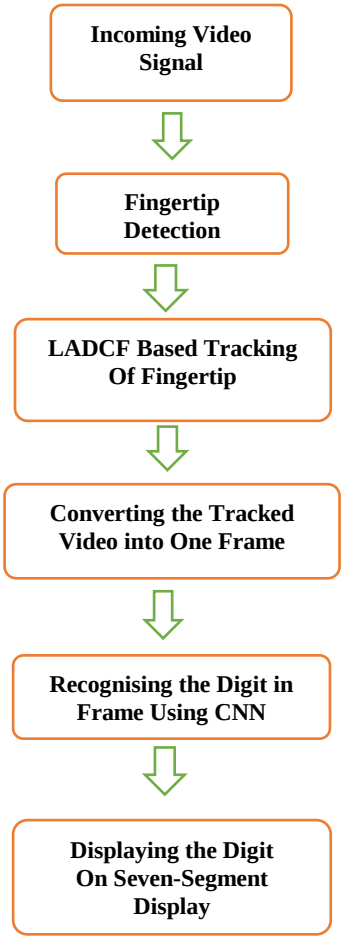


Fig. 1.Flowchart of proposed algorithm.

V. FINGERTIP DETECTION

Detecting a fingertip in an RGB video, is a tough and challenging task as the dimension of fingertip is small compared to the whole frame. However, recent CNN-based algorithms have come up to detect fingertip precisely and accurately. The paper implements an algorithm [7] which segments the hand region, followed by localization of its centroid. From the segmented region, fingertip is localized with the help of an entropy function, termed as distance-weighted curvature entropy function.

In the above-mentioned entropy function, distance of each contour points from the centroid is calculated and stored in a 1-D matrix. Then the weighted distance is calculated by multiplying the matrix with the distance of curvature points from the centre of hand. Using contours detecting algorithm as proposed by [16], contours of the hand are detected. Let p be the contours of hand of length k. On sampling the contours into f samples, sampling interval is given by,

$$\Delta p = \frac{k}{f} \tag{1}$$

The turning angle θ, is defined as the change in direction of tangent of two points on contour. While curvature c is the change in tangent direction of points on moving along the curve and is given by,

$$c(p) \approx \frac{\theta}{\Delta p} \tag{2}$$

Hence the curvature entropy u which is locally proportional to $c(p)$ (curvature of the contour) is approximated by,

$$u\big(c(p)\big)\theta - \cos\big(\Delta p.c(p)\big) \tag{3}$$

The other factor influencing entropy function is the distance between contour points and the centroid (x', y') which is denoted by $s(p)$. The complete function is denoted by $\chi(p)$ and is defined as the

following,

$$\chi(p) = u\big(c(p)\big).s(p)^{\tau} \quad (4)$$

Where τ is the weighted distance as defined earlier.

It is known that the region of fingertip will have maximum value for curvature entropy along with maximum distance from centroid. Thus, the entropy function $\chi(p)$ will have its maxima at the fingertip coordinates, which may be expressed as,

$$x_o.y_o = p(arg\ max\ \chi(p)) \quad (5)$$

Where (x_o, y_o) are the coordinates of fingertip.

VI. FINGERTIP TRACKING USING LADCF

As discussed in [20], LADCF algorithm is based on temporal consistency preserving spatial feature selection method for appearance modelling. On detecting the fingertip in a frame, the tracking algorithm(LADCF) starts its execution. It uses single-channel features which can be further extended to multi-channel. Direction Method of Multiplier (ADMM) is used for optimizing the algorithm.

A. Spatially Feature Selection Technique

Selection of features is done using discriminative and descriptive information stored in filter 'h'.

$$h_{\alpha} = diag\,(\alpha)h \quad (6)$$

Where, α is the feature vector.

$diag\,(\alpha)$ is the diagonal matrix of α.

Spatial feature vector is dependent on input by,

$$h_{\alpha}^{'} f = h_{\alpha}^{'} f_{\alpha} \quad (7)$$

In learning stage, spatial feature vector can be given as,

$$h' = argmin_{h,\alpha} \|h \odot f - \beta\|_2^2 + \lambda_1 \|f\|_0 \quad (8)$$

For the approximation of sparsity, convex envelope l_1 - norm and l_0 - norm (non-convex) were obtained and was restricted to current template 'x'.

$$\|\alpha - \alpha_{model}\|_0 < x \quad (9)$$

Using l_1 - norm, spatial feature is given by,

$$argmin_h \|h \odot f - \beta\|^2 + \lambda_1 \|h\|_1 + \lambda_2 \|h - h_{model}\|_1 \quad \ldots\ldots(10)$$

Where λ_1 & λ_2 are tuning parameters.

& $\lambda_1 \ll \lambda_2$.

Sparsity of h, is increased, by increasing the similarity between the estimated h and the template model h_{model}.

On taking its l_1 - norm, it is found to be,

$$argmin_h \|h \odot f - \beta\|_2^2 + \lambda_1 \|h\|_1 + \lambda_2 \|h - h_{model}\|_2^2 \quad \ldots\ldots(11)$$

B. Multi-channel Feature Selection Model

In this framework, the appearance model was characterised by multi-channel feature representation. Multi-channel input is denoted as,

$$h = \{h_1, h_2, \ldots, h_k\} \quad (12)$$

Where k is the number of channels.

Where

$$f_i = \left[f_i^1, f_i^2, f_i^3, \ldots, f_i^o\right]' \quad (13)$$

In case of multi-channel model, the objective function is given by,

$$g(h) = \sum_{i=1}^{k} \|h_i \odot f_i - \beta\|_2^2 + \lambda_1 \sum_{j=1}^{d^2} \sqrt{\sum_{i=1}^{k} (h_i^j)^2} + \lambda_2 \sum_{i=1}^{k} \|h_i - h_{model_i}\|_2^2 \quad \ldots\ldots(14)$$

Where h_i^j is the jth element of ith channel feature vector.

C. Optimisation

To optimize the above derived equation, eq^n (14), convex optimization was employed,

$$h = argmin_{h,h'} \sum_{i=1}^{k} \|h_i \odot f_i - \beta\|_2^2 + \lambda_1 \sum_{j=1}^{d^2} \sqrt{\sum_{i=1}^{k} (h_i^j)^2} + \lambda_2 \sum_{i=1}^{k} \|h_i - h_{model_i}\|_2^2 \quad \ldots.(15)$$

We define a set of Lagrange multipliers, given by

$$\eta = \{\eta_1, \eta_2, \eta_3, \ldots, \eta_k\} \quad (16)$$

With a penalty controlling parameter μ & μ>0.

For updating h, for a given h', η & μ. We optimize h using OCF learning method.

$$h = argmin_h \mathcal{L}(h, h', \eta, \mu) \quad (17)$$

$$h' = argmin_{h'} \mathcal{L}(h, h', \eta, \mu) \quad (18)$$

$$\eta = argmin_{\eta} \mathcal{L}(h, h', \eta, \mu) \quad (19)$$

TABLE I. TRACKING ALGORITHM

LADCF
Prerequisites Input image I_i, filter (model) h_{model}, centre coordinate (target) c_{x-1} & frame size b x l of frame $i-1$. Tracking:- 1. Search windows were extracted of frame size s from input image at its center coordinate. 2. Corresponding feature representation was obtained as $[F(s)]_{x-1}$. 3. Response scores 'r' was calculated. 4. Centre coordinate and scale size were set. Observation:- 1. Multichannel features were obtained and represented as 'F', depending upon the current target. 2. 'α' was optimized for 'k' iterations. Update:- 1. Filter model 'h_{model}' was updated. Output:- 1. Bounding box was drawn using centre coordinate and frame size. 2. Updated filter model 'h_{model}' for the given frame 'i'.

VII. DIGIT RECOGNITION

After tracking the fingertip movement in consecutive frames, single frame is constructed from it as shown in the figure 23123. The image thus formed is further used to recognise the numeral character present in it. For recognising the handwritten digit in the frame, we have deployed CNN based algorithm. As proposed in [18], CNN is a computationally faster algorithm as compared to other multi-layer perceptron. CNN works on 2D convolution on the input pixel values. CNN based algorithm implemented in our paper for digit recognition task can be divided into following subtasks: -

A. Convolution Layer

The very first layer of a CNN is the convolution layer. An RGB image consists of three channels i.e., Red, Green and Blue. However, convolution layer can only work on a single channel at a time. Thus, 3 channel RGBV image was converted into one channel by taking average of it. Hence, in this layer we have a 2-D matrix of 28 x 28 pixels. We picked three filters randomly of 5x5 from the image. Convolution of all the three filters with the image was performed resulting into three 2-D matrices of 28 x 28 pixels. The whole task could be summed up in a table: -

TABLE II. CONVOLUTION LAYER ALGORITHM

CONVOLUTION LAYER ALGORITHM
Step 1: RGB image was converted into a single channel image. Step 2:- Three filters of 5 x 5 pixel was picked from the image randomly. Step 3:- Each filter was multiplied by the corresponding pixel of 28 x 28 image. Step 4:- All the elements of resultant 5 x 5 matrix was summed and divided by 25. Step 5:- Filter was moved and centered at the next pixel of image and step 2 to 4 was repeated for each filter separately.

Fig. 2.Three Filters used in Convolution.

B. Activation Layer

The resultant matrices of convolution layer may or may not consists of negative values. However, a negative value as the pixel of an image, holds no meaning. In order to remove negative values from the pixel matrix of filtered images, we applied an activation function. An activation function only lets the positive values to pass through it. There are many activation functions as such sigmoid function, Rectified Linear Unit (ReLU) function, tanh function etc.

In our algorithm we applied ReLU as an activation function for the filtered images. A ReLU only activates a node if its value is greater than 0 and deactivates the negative node. A ReLU layer can be mathematically defined as follows: -

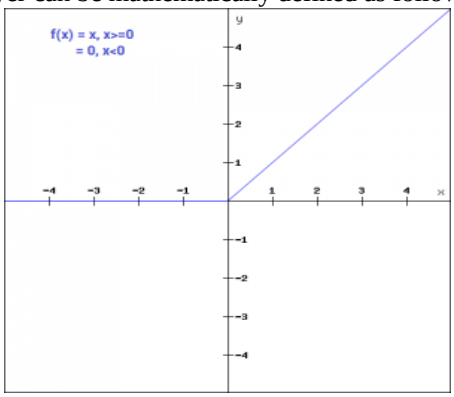


Fig. 3.ReLU Activation Function.

C. Max Pool Layer

After passing the image matrices from ReLU layer, the dimension of the matrix is very high i.e., 28 x 28 that makes up to 724 elements. In order to reduce the size of image matrix, we created a window of 2 x 2 size and put it at the beginning of image pixel matrix. Maximum value of that window was picked and replaced in the image matrix. Window was slided over the whole matrix, thus reducing its size to its half.

After pooling layer, the whole process starting from convolution to pooling, was repeated thrice. This resulted in 4 x 4 matrices of three channels. All the elements of three channels were arranged in a single vector/list in a row major form. The vector thus formed after training, was unique for each digit and were 10 in number.

Unknown handwritten digits had to undergo all the above-mentioned steps, which yielded a 1-D vector for them. The resultant vector was compared with all the 10 vectors that were formed during training of model. The unknown digit was predicted on the basis of highest resemblance of its vector with the precalculated 10 vectors.

VIII. DISPLAY ON SEVEN SEGMENT

After the fingertip drawn digit is recognized by CNN algorithm, it is displayed on seven-segment display board. This is task is performed using the modules created in SystemVerilog. Steps performed for displaying the digit are as follows :

TABLE III. DISPLAYING THE RECOGNISED DIGIT

STEPS FOR DISPLAYING THE PREDICTED DIGIT
Step 1: Read the predicted digit using 'readmemb()' function Step 2: Assigned the data read to a register named 'digits'. Step 3:Converted the binary data to the anode and cathode connection values using 'case()'. Step 4: Produced the data if data lies in the range 0-9 else turned off the 7-segment display.

IX. RESULT

The above proposed algorithm was trained on MNIST datasets containing 60,000 images of handwritten digits. Dataset was divided into training and testing data in the ratio 5:1. The CNN model was trained with the following parameters :-

- Number of epochs = 17
- Learning rate = 0.003
- Number of hidden layers = 2
- Optimizer used = Stochastic Gradient Descent

On testing the trained model on 10,000 images, accuracy of the algorithm was 98.4% which is much higher than [7] which was 95%.

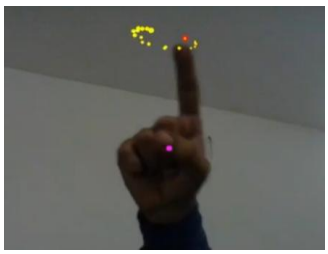


Fig. 4.Fingertip Tracking.

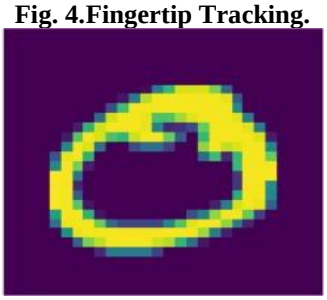


Fig. 5.Fingertip Digit fed into CNN algorithm.

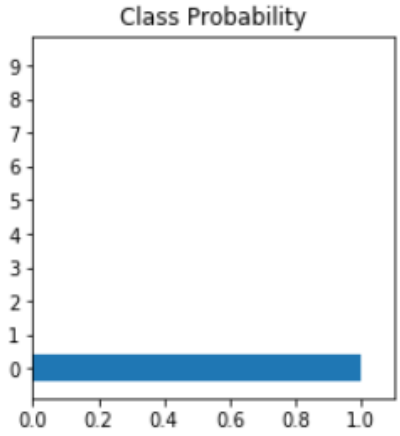


Fig. 6.Digit predicted by the CNN algorithm.

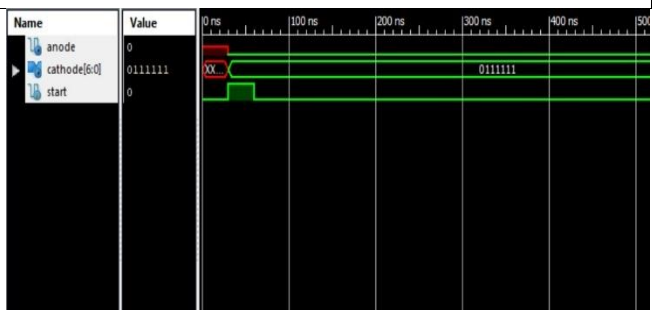


Fig. 7.Digit displayed on Seven-Segment.

```
Simulator is doing circuit initialization process.
Finished circuit initialization process.
Digit Predicted is :
0000
Anode value of Seven Segment :
0
Cathode value of Seven Segment :
0111111
```

Fig. 8.SystemVerilog output of the predicted digit.

X. CONCLUSION

The paper has proposed an algorithm for fingertip writing digit recognition system using CNN. Proposed recognition system processes both pre-processed image algorithms and fingertip writing numbers recognition algorithms in hardcode. The CNN model has been trained on 50,000 handwritten digits provided by an MNIST and has been tested on 10,000 images. The computational efficiency of this algorithm is very high as it implements FPGA for digit recognition and the accuracy of digit recognition is 98.4%. The proposed algorithm makes use of both SystemVerilog as well as C++ for fingertip hand recognition. SystemVerilog is used just to communicate between hardware and software. Direct programming interface facility of SystemVerilog has been used to read an image and to convert it in its RGB value using C++ program. That corresponding RGB value has been further used for digit recognition and the recognised digit was displayed on SevenSegment using SystemVerilog. The future expectation of this algorithm is the betterment of the recognition system and increasing its computational efficiency manifold. In our coming papers, we will be proposing the algorithm to implement the CNN on hardware only as it will increase the computational efficiency and will make the algorithm faster with a high accuracy rate. So that our main motive of helping humankind by making use of machines will come true.

ACKNOWLEDGMENT

We express our deepest gratitude to all those who helped us make this research complete and meaningful in all sense. First of all, we would like to thank our faculty in-charge, Prof. S. Indu, whose contribution to this research work is incredible. She helped us to get through this research work by providing her valuable suggestions and sparing her precious time. Furthermore, we acknowledge the contribution of other staffs of the ECE department (DTU), who helped us by providing their valuable suggestions related to our topic. Last but not least, we are highly thankful to our parent and friends, who provided us moral support in times of distress and failure and played a vital role to make this work complete in all sense.

REFERENCES

1. Ching-Long Shih, et al, (2015), FPGA Implementation of Vision-Based Fingertip-Writing Digits Recognition System, IEEE International Conference on Systems, Man, and Cybernetics.
2. O.D. Trier, et. Al, Feature extraction methods for character recognition – a survey, Pattern Recognition, Vol. 29, No 4, pp. 641-662, 1996.
3. R. Plamondon, S. N. Srihari, On-line & off-line Handwriting Recognition: A Comprehesive Survey, IEEE Transactions on Pattern Analysis & Machine Intelligence, Vol 22, No 1, pp. 63-84, 2000.
4. Sohom Mukherjee, et al, (2018), Fingertip Detection and Tracking for Recognition of Air-Writing in Videos.
5. M. Kristan, R. Pflugfelder, A. Leonardis, J. Matas, F. Porikli, L. Cehovi, G., G. Fernandez and T. Vojir. 2018 The Sixth Visual Object Tracking VOT2018 Challenge Results, pp 3-53.
6. K. Outcharov, et. Al, Accelerating Deep Convolutional Neural Networks Using Specialized Hardware, (Feb 2015)
7. V. R. Rudra.,K. Bhavanish., Handwritten Digit Recognition using CNN, Int. J. of Innovative Sc. & Research Tech., Vol. 4, Iss. 6.
8. L.S.Oliveira, R. Sabourin, Support vector machines for handwritten numerical string recognition, Int.Workshop on Frontiers in Handwriting Recognition, Tokyo, Japan (2004).
9. Iuliana Chiuchisan, (2013), A New FPGA-based Real-Time Configurable System for Medical Image Processing, 4th IEEE International Conference, Romania.
10. Ahinus H, et al, A New Object Detection and Tracking Using FPGA, IOSR Journal of Electronics & Communication Engineering.
11. Feldman J., Singh M.,Information along contours and object boundaries, Psychological Review, Vol. 112, pp. 243-252.
12. Tianyang X., Zhen-Hua F., Xiao-Jun W., Josef K. (2019). Learning Adaptive Discriminative Correlation Filter via Temporal Consistency Preserving Spatial Feature Selection for Robust Visual Object Tracking. IEEE trans. on Img. Proc, Vol. 8, Iss. 11, pp 5596 - 5609.
13. X. Yang, J. pu, MDig: Multi-digit Recognition using Convolutional Neural Network on Mobile, 14th International Conference on Natural Computation, Huangshan, China.

AUTHORS PROFILE

MD Shahbaz Khan currently pursuing B. Tech in Electronics & Communication branch from Delhi Technological University, Delhi India. His fields of interest for research are Computer Vision, Machine Learning and FPGA.

Niharika currently pursuing B. Tech in Electronics & Communication branch from Delhi Technological University, Delhi India. Her field of research are Machine Learning and Data Science.

Priya Yadav currently pursuing B. Tech in Electronics & Communication branch from Delhi Technological University, Delhi India. Her fields of interest are Machine Learning. FPGA and Verilog.

Rishabh Verma currently pursuing B. Tech in Electronics & Communication branch from Delhi Technological University, Delhi India. His fields of research are FPGA and Verilog.

Dr. Indu Sreedevi did her PhD in the area of Visual Sensor Networks from University of Delhi, Delhi, India. Currently she is working as Dean (student Welfare) and Professor of ECE Department of Delhi Technological University. She has guided around 40 M. Tech and 11 PhD students under her. She has published around 150 papers in reputed Journals and National and International conferences. Her area of research interest is Computer Vision, Sensor Networks and Image Processing. She is recipient of Best Branch Councillor award from IEEE USA and also recipient of Outstanding Branch Councillor award of IEEE Delhi section for 5 consecutive years from 2013-2018.

2021 2nd International Conference for Emerging Technology (INCET)
Belgaum, India. May 21-23, 2021

From Classical to Quantum: A Review of Recent Progress in Reinforcement Learning

Kunal Kashyap
Department of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
kunalkashyap1107@gmail.com

Lalit
Department of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
lalit31999@gmail.com

Daksh Shah
Department of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
dakshshah0605@gmail.com

Lokesh Gautam
Department of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
lokeshgautam21@gmail.com

2021 2nd International Conference for Emerging Technology (INCET) | 978-1-7281-7029-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/INCET51464.2021.9456218

***Abstract*—Standard machine learning algorithms have taken shape well due to increase in computational power. There is a special focus on reinforcement learning, which offers solutions to concepts like autonomous vehicles and robotics, among others. On the other hand, research in the fields of quantum computing and associated machine learning has catapulted in the last few years. This work analytically compares and contrasts between the classical and quantum approaches to two reinforcement learning problems, and tries to put focus on future works to advance both fields further.**

Keywords—reinforcement learning, quantum computing, q-learning, variational quantum circuits

I. INTRODUCTION

Reinforcement learning (RL) refers to a class of machine learning problems which make use of learning the steps for maximizing a numerical reward, specifically by learning about situations and actions and the possible mappings between each [1]. These are most distinguished by being crucially closed-loop, not being informed about possible actions at any state, and how a particular action can lead out to multiple subpaths over repeated experiments. Some intuitive examples may include a game of chess, an adaptive controller at a refinery, and a mobile robot.

One can segregate an RL system into four main sub-components (excluding the agent and the environment):

- *Policy:* It describes, at a given time, the learning agent's way of behaving. Psychologically, it can be termed as "a set of stimulus response rules or associations". Often called the core of an RL agent, it alone is sufficient to ascertain the agent's behavior with respect to the environment.
- *Reward signal:* It is essentially the goal in an RL problem. The agent is being incentivized to maximize the cumulative reward received during each iteration. In a way, it describes the positive and negative events for the agent.
- *Value function:* For an arbitrary state, its *value* is the total amount of reward the agent can hope to collect over the future, starting from that state. Thus, the value function signifies the long-term desirability of states, where rewards determine the immediate and intrinsic desirability.
- *Model:* Used in some RL systems, it can mimic the behavior of the environment. In other words, it can allow reasonings to be furnished on how the environment will behave.

On the other hand, the field of quantum machine learning (QML), a subset of quantum computing [2] has been a major focus of research in recent years, partly due to development of quantum computers by prolific technology firms such as IBM [3] and Google [4]. The mix of quantum computing, which offers infinite superpositions of quantum bits (qubits) and quantum information transfer (instantaneous transfer), with learning algorithms offers possibilities of predictions more accurate than ever [5].

This work is an effort to analytically compare and contrast between two approaches to the problem of machine learning: the first being Q-learning, a standard model-free technique for reinforcement learning, and the second being an RL system made using variational quantum circuits on noisy intermediate-scale quantum (NISQ) devices [6], [7]. It also endeavors to comment on future trends based on the approaches presented, and hopefully guide research in the right direction.

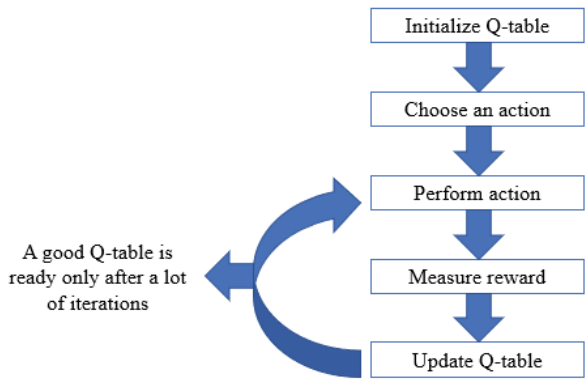


Fig. 1. Flowchart showing the working of Q-learning algorithm.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:26:14 UTC from IEEE Xplore. Restrictions apply.

The final Q-table generated is good enough for testing use only after a considerable number of iterations.

II. Q-Learning

Q-Learning is one of the recent reinforcement learning methods which is used for learning the optimal policy in a Markov Decision Process [1]. In reinforcement learning, after every action, a state is reached and accordingly the next step has to be taken. This state-action pair is taken care by the Q-values. The main desideratum of the Q-learning is to maximise the total reward or the return after every action taken by the algorithm.

The Q-function for a particular policy takes in a state and action pair, and gives out an expected reward from performing that particular action in that particular state, following the particular policy all along. It is based on the following Bellman's Equation:

$$Q^{\pi}(s_t,a_t)=E[R_{t+1}+\gamma R_{t+2}+\gamma^2 R_{t+3}+\ldots][s_t,a_t], \quad (1)$$

where the given state and action pair is multiplied with the expected discounted cumulative reward so that Q-values are created for that pair.

After every action, there is an iteration and the Q-values are updated. This iteration process continues till we get an optimal set of Q-values in the Q-Table. This table gives the idea of the Q-values with respect to various actions taken and returns received. Initially all the values in Q-table are 0 and then after every iteration, the values are updated.

The flow diagram for the functioning of the Q-Algorithm is shown in Fig.1. When we are able to get the maximum return, a good Q-table is achieved.

The algorithm works on the concept of Exploration and Exploitation. When it is just the start of the process and there is no idea of current surroundings, the process of exploration is executed so as to study the environment. Once the reinforcement learning takes place and the idea of return on future action is known, exploitation has to be done. A good Q-learning algorithm makes a balanced use of exploration as well as exploitation.

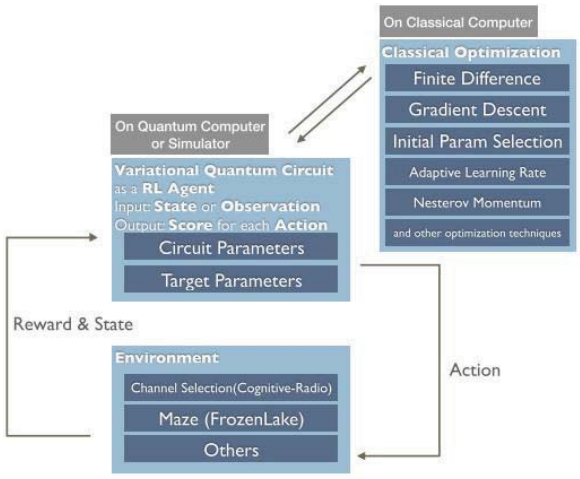


Fig. 2. High-level view of variational quantum circuits for deep reinforcement learning [8].

A hybrid quantum-classical approach is described to solve the RL problem. The framework suggested in [8] can be expected to solve more complex tasks when quantum machines are available in a larger scale.

III. Variational Quantum Circuits for RL

Reference [8] was the first to reshape concepts more akin to the classical approach to deep reinforcement learning, e.g. experience replay and target network, into an actualization of variational quantum circuits. Also, a computational basis encoding scheme was used to reduce the number of parameters controlling the model required for deep reinforcement learning in the classical scenario.

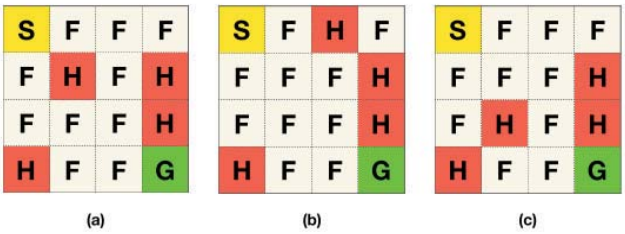


Fig. 3. Frozen-lake environment [8].

Frozen-lake environment is a simple maze environment in openAI Gym [9]. The RL agent is supposed to go from the start position (marked S) to the goal position (marked G) one step at a time while avoiding holes (marked H).

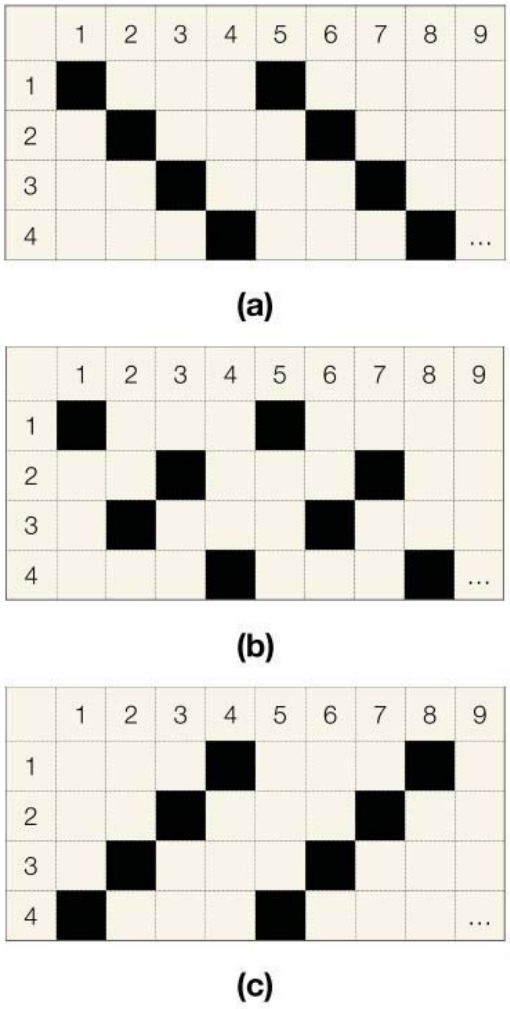


Fig. 4. Cognitive-radio environment [8].

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:26:14 UTC from IEEE Xplore. Restrictions apply.

Cognitive-radio environment is used to simulate a real-world application. The agent is supposed to select a free channel not occupied or in use by a primary user (shaded black). To avoid complexity, three simple periodic patterns were utilized.

Two testing environments were considered for the ensuing experiments: frozen-lake and cognitive-radio. The frozen-lake environment is a fairly simple and commonly tested example in reinforcement learning (see Fig.3), and feasible simulations are possible with available quantum machines if the problem size dimension is not too large (with reasonable time consumption for the experiment). Using this environment, it was also established that quantum circuits have the capability to learn a policy, or a process where decisions are to be made in sequence. The cognitive-radio environment (see Fig.4), while equivalent to the frozen-lake environment in terms of complexity, helped establish the capability of these circuits to execute a real-world application.

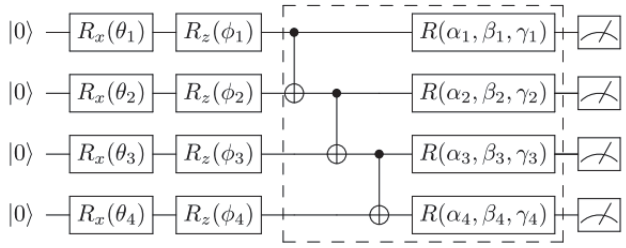


Fig. 5. VQ-DQN proposed for the frozen-lake testing environment [8].

Rotation about the *x*-axis and *z*-axis is provided by the single-qubit gates at the beginning of the circuit. The CNOT gates entangle the quantum states. The general single-qubit unitary gates provide the iterative optimization.

The variational quantum circuit is an approach adopting the strengths of both classical and quantum aspects of computing. It has tunable parameters optimized in an iterative manner by a classical computer, which are equivalent to weights in an artificial neural network. This approach is flexible with respect to the depth of circuits and somewhat resistant to noise. Therefore, by combining a variational quantum circuit, QML algorithms can bypass complex quantum errors present in current devices, even when quantum error correction and fault-tolerant computation is not available in current quantum machines.

The reinforcement learning agent (Variational Quantum Deep Q-Network – VQ-DQN) proposed was divided into quantum and classical parts (see Fig.2), the former outputting measurement results based on two sets of parameters, determining possible actions to take. The latter part focuses on optimizing the circuit by deciding the new values of parameters. The same circuit architecture was used to construct two sets of parameters. Optimization was achieved by calculating gradients of expectation values of quantum measurements using the unchanged circuit architecture and slenderly different parameters. Each state was further encoded with computational basis encoding, i.e., each state was represented in binary and then corresponding quantum state was calculated from that result. The encoding was done using the single-qubit unitary rotation method.

The variational quantum circuit is flexible in circuit depth, which means that a shallow circuit can still approximately achieve goals set by the solution space, but a deep, more sophisticated circuit may perform better. Thus, a strongly entangling low-depth circuit was used twice for both testing environments. The number of qubits used in the aforementioned circuit can be modified to fit the problem posed by the respective testing environment as well as the quantum computer or simulator.

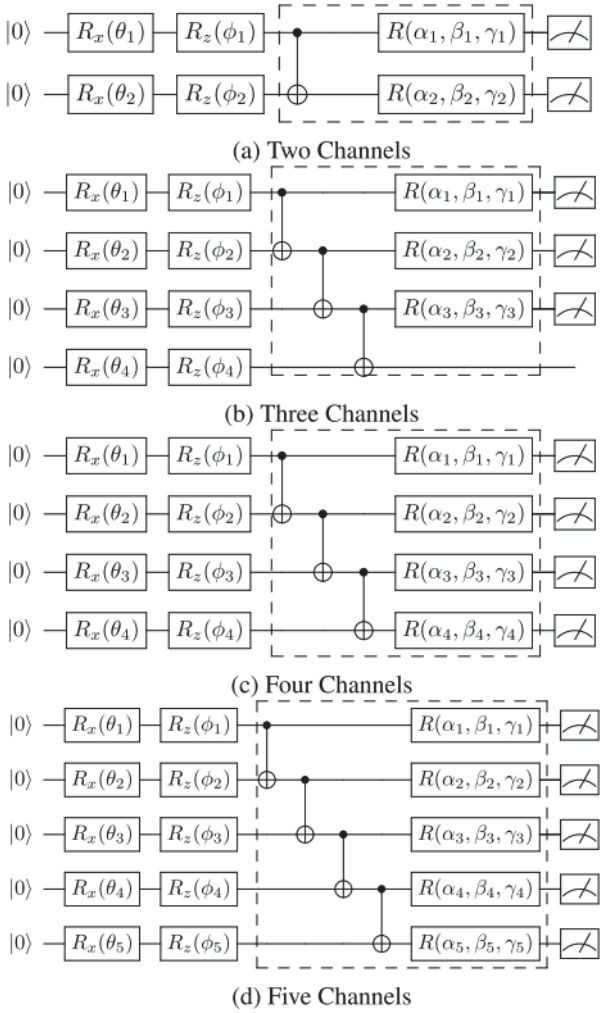


Fig. 6. VQ-DQN proposed for the cognitive-radio environment [8].

A four-qubit circuit was proposed for the three-channel case because the number of environment states (i.e., 9) is greater than the number of computational basis states (i.e., 8). The networks can be seen as an extension for that proposed for the frozen-lake environment.

The VQ-DQN for both testing environments can be seen in Fig.5 and Fig.6. For an *n*-qubit quantum circuit, the measurement output was taken to be the ensemble average (i.e., expectation) value for a single qubit. The output for *n* qubits was stored in an *n*-tuple. The index of the above measurement output port/wire which has the largest expectation value was used as the action for the circuit. The subsequent action that the agent then chooses was determined by the corresponding output for a qubit, and not some random outcome in each single run of the experiment. Therefore, the discrete probability distribution can be learned in O(*n*) time complexity. Quantum circuits also provide the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:26:14 UTC from IEEE Xplore. Restrictions apply.

TABLE I. COMPARISON BETWEEN CLASSICAL Q-LEARNING AND PROPOSED VQ-DQN FOR FROZEN-LAKE AND COGNITIVE-RADIO TESTING ENVIRONMENTS.

Environment	2-Channels	3-Channels	4-Channels	5-Channels	Frozen-Lake
Q-Learning	$2 \times 2 \times 2$	$3 \times 3 \times 3$	$4 \times 4 \times 4$	$5 \times 5 \times 5$	$4 \times 4 \times 4$
VQ-DQN	$2 \times (3 \times 2 + 1)$	$3 \times (3 \times 2 + 1)$	$4 \times (3 \times 2 + 1)$	$5 \times (3 \times 2 + 1)$	$4 \times (3 \times 2 + 1)$

flexibility of parallel measurement across each qubit, which means that the number of repeated experiments can be a fixed number and not dependent on n. Moreover, because the focus is on the qubit having the highest expectation value, the algorithm proposed was comparatively potent against noise and errors present in real quantum machines.

Two kinds of simulations were performed as part of the research: one being with software packages PennyLane [10] and PyTorch [11] (for simulation acceleration), the other with IBM Qiskit-Aer simulation software as the backend to incorporate noise from a real quantum computer.

The lack of quantum error correction in current NISQ machines implies that they are highly unsuitable for deep quantum circuit architectures. A rough corollary can be that when the number of quantum gates increases in a circuit, there will exist more errors. The circuit proposed in [8] thus cannot have parameters commensurate to classical neural networks, but was able to show comparable performance for the environments considered with fewer iterative parameters.

IV. COMPARISON BETWEEN APPROACHES

Table 1 shows for both testing environments, the number of parameters in classical Q-learning is of the form $n \times n \times n$, where n is a dimension of the environment [8]. This means that as the environment gets more complex, the number of parameters increases, making the RL problem more computationally expensive.

On the other hand, the VQ-DQN approach makes use of two-qubit CNOT gates and parameterized single-qubit unitary gates, leading to a total of $n \times (3 \times 2 + 1)$ parameters in its learning process. So, in general, the circuit can represent the n^2 states the testing environment requires with an n-qubit system, providing 2^n computational basis states in the process.

An exception is $n = 3$, where the number of states required is greater than the number of computation basis states. Here, the proposal was to build a four-qubit circuit but leave out the parameterization and measurement of the fourth qubit. A graphical comparison can be seen in Fig.7.

From Table 1 as well as Fig.7, the VQ-DQN approach (for the cognitive-radio environment) uses less parameters for cases with large number of channels than the classical Q-learning approach.

V. ANALYSIS

We saw from the two perspectives of solving the testing environments the efficiency of variational quantum circuits, which implies that this approach can be applied on near-term NISQ machines while bypassing difficulties intrinsic to quantum computing such as quantum error correction.

Current research in the quantum computing field is limited, mostly due to the level of complexity involved in building and maintaining a quantum computation device. In spite of this limitation, considerable progress has been made, surpassing existing applications in terms of implementations and creating new ones in the process [12]–[15].

Future scope with the quantum approach to reinforcement learning can be seen with everyday devices edge-connected with quantum machines to provide precise judgments and decisions for future steps. A notable mention in this scope is autonomous vehicles, which if connected to a quantum computer able to provide real-time actions may not ever have to deal with choices such as the runaway trolley paradox. This is, of course, only feasible if and when quantum machines are built to scale.

In conclusion, the problems faced with quantum computing at present is because conclusive research on this topic is only at its nascent stage, so we can expect big disruptions in the field in the near future.

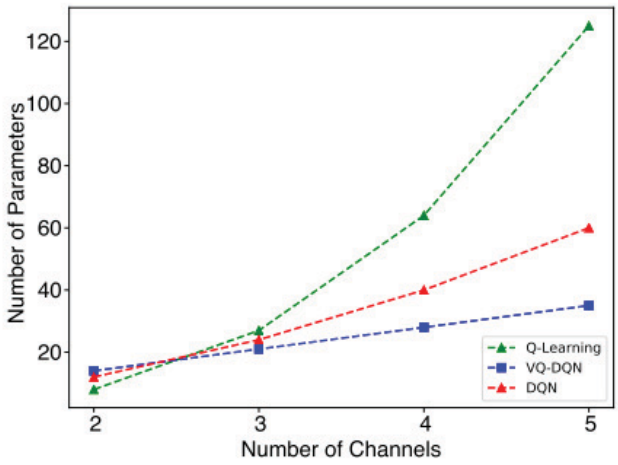


Fig. 7. Comparison of number of parameters required for solving the n-channel cognitive-radio testing environment [8].

The VQ-DQN approach is able to compute more efficiently with lesser number of parameters as the number of channels increases.

ACKNOWLEDGMENT

K. K., L. and D. S. thank Srijan Saxena and Sudhanshu Kumar for constructive discussion and great support.

REFERENCES

[1] P. R. Montague, "Reinforcement Learning: An Introduction, by Sutton, R.S. and Barto, A.G.," *Trends Cogn. Sci.*, vol. 3, no. 9, 1999, doi: 10.1016/s1364-6613(99)01331-5.

[2] IBM, "What is quantum computing ?," *IBM*, 2020.

[3] IBM, "IBM Q Experience," *Quantumexperience.Ng.Bluemix.Net*, 2019. .

[4] F. Arute *et al.*, "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, no. 7779, 2019, doi: 10.1038/s41586-019-1666-5.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:26:14 UTC from IEEE Xplore. Restrictions apply.

[5] D. Dong, C. Chen, H. Li, and T. J. Tarn, "Quantum reinforcement learning," *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 38, no. 5, pp. 1207–1220, 2008, doi: 10.1109/TSMCB.2008.925743.

[6] Y. Du, M.-H. Hsieh, T. Liu, and D. Tao, "Expressive power of parametrized quantum circuits," *Phys. Rev. Res.*, vol. 2, no. 3, 2020, doi: 10.1103/physrevresearch.2.033125.

[7] J. Preskill, "Quantum computing in the NISQ era and beyond," *Quantum*, vol. 2, 2018, doi: 10.22331/q-2018-08-06-79.

[8] S. Y. C. Chen, C. H. H. Yang, J. Qi, P. Y. Chen, X. Ma, and H. S. Goan, "Variational Quantum Circuits for Deep Reinforcement Learning," *IEEE Access*, vol. 8, pp. 141007–141024, 2020, doi: 10.1109/ACCESS.2020.3010470.

[9] G. Brockman *et al.*, "OpenAI Gym," pp. 1–4, 2016, [Online]. Available: http://arxiv.org/abs/1606.01540.

[10] V. Bergholm *et al.*, "PennyLane: Automatic differentiation of hybrid quantumclassical computations," *arXiv*. 2018.

[11] S. Zagoruyko *et al.*, "A multipath network for object detection," in *British Machine Vision Conference 2016, BMVC 2016*, 2016, vol. 2016-September, doi: 10.5244/C.30.15.

[12] A. D. Corcoles *et al.*, "Challenges and Opportunities of Near-Term Quantum Computing Systems," *Proc. IEEE*, vol. 108, no. 8, 2020, doi: 10.1109/JPROC.2019.2954005.

[13] A. Ajagekar and F. You, "Quantum computing for energy systems optimization: Challenges and opportunities," *Energy*, vol. 179, 2019, doi: 10.1016/j.energy.2019.04.186.

[14] C. Outeiral, M. Strahm, J. Shi, G. M. Morris, S. C. Benjamin, and C. M. Deane, "The prospects of quantum computing in computational molecular biology," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 11, no. 1. 2021, doi: 10.1002/wcms.1481.

[15] Y. Cao *et al.*, "Quantum Chemistry in the Age of Quantum Computing," *Chemical Reviews*, vol. 119, no. 19. 2019, doi: 10.1021/acs.chemrev.8b00803.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:26:14 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the ASME 2021
International Mechanical Engineering Congress and Exposition
IMECE2021
November 1-5, 2021, Virtual, Online

IMECE2021-69933

GENERATION AND PARAMETERIZATION OF FORCED ISOTROPIC TURBULENT FLOW USING AUTOENCODERS AND GENERATIVE ADVERSARIAL NETWORKS

Kanishk[1], Tanishk Nandal[1], Prince Tyagi[1], Raj Kumar Singh[1]

[1]Delhi Technological University, New Delhi, India

ABSTRACT

Autoencoders and generative neural network models have recently gained popularity in fluid mechanics due to their spontaneity and low processing time instead of high fidelity CFD simulations. Auto encoders are used as model order reduction tools in applications of fluid mechanics by compressing input high-dimensional data using an encoder to map the input space into a lower-dimensional latent space. Whereas, generative models such as Variational Auto-encoders (VAEs) and Generative Adversarial Networks (GANs) are proving to be effective in generating solutions to chaotic models with high 'randomness' such as turbulent flows. In this study, forced isotropic turbulence flow is generated by parameterizing into some basic statistical characteristics. The models trained on pre-simulated data from dependencies on these characteristics and the flow generation is then affected by varying these parameters. The latent vectors pushed along the generator models like the decoders and generators contain independent entries which can be used to create different outputs with similar properties. The use of neural network-based architecture removes the need for dependency on the classical mesh-based Navier-Stoke equation estimation which is prominent in many CFD softwares.

Keywords: Machine Learning, Fluid Mechanics, Generative Models, Auto-encoders, Turbulence Flow Generation.

1. INTRODUCTION

The advancements in computational and simulation capabilities in recent years have led to better than ever analysis of complex fluid flows. This in turn has led to vast amounts of multi-dimensional data representing the intricate details of these chaotic flows. Due to the challenges posed by the vastness of this ever-increasing data in analysing complex numerical simulation problems, extraction of important physical parameters from this data is indispensable. Characteristics of such high-fidelity DNS data can be contained in a low dimensional latent subspace. Model reduction is used to extract coherent key features and reduce data and computation costs. Several Model Order Reduction techniques have been employed and refined to analyse flows with high spatial and temporal complexity.

Various traditional model reduction techniques including proper orthogonal reduction, dynamic mode decomposition, global stability analysis, Koopman analysis, and Resolvent analysis have been utilized to analyse fluid flows [1]. POD and DMD are based on Singular value decomposition. POD is generally used along with projection techniques such as Galerkin Projection. However, it is found that Galerkin projection models are unstable under many circumstances [2,3], and this sparks interest to look into non-projection-based approaches. Also, the rapid advancements in machine learning and especially neural networks have accelerated the study of fluid flows and have shown great potential in modelling and study of these flows due to their handling of nonlinear phenomena. Hence it is not surprising that various reduced-order modelling techniques based on machine learning (ML-ROMs) have recently been proposed [5-7]. Neural network-based dimensionality reduction techniques often employ autoencoders.

In recent years, there have been many developments in the field of deep learning due to the availability of massive computing power. Variational autoencoders [8] are generative models that encode the input as a probabilistic distribution. Since their proposal, they have been consistently leveraged. For instance, Gregor et al. [9] combined an attention mechanism with a sequential variational auto-encoding framework that generates highly realistic images indistinguishable from the naked eye. Other enhanced models such as Importance Weighted Autoencoders [10] and Ladder Autoencoders [11] have also been presented. Another variant of autoencoders that has been utilized heavily for engineering problems is convolutional autoencoders. Convolutional autoencoders differ from traditional autoencoders due to weight sharing. Hence, they have been demonstrated to work well and are extensively used in image processing

 © 2021 by ASME

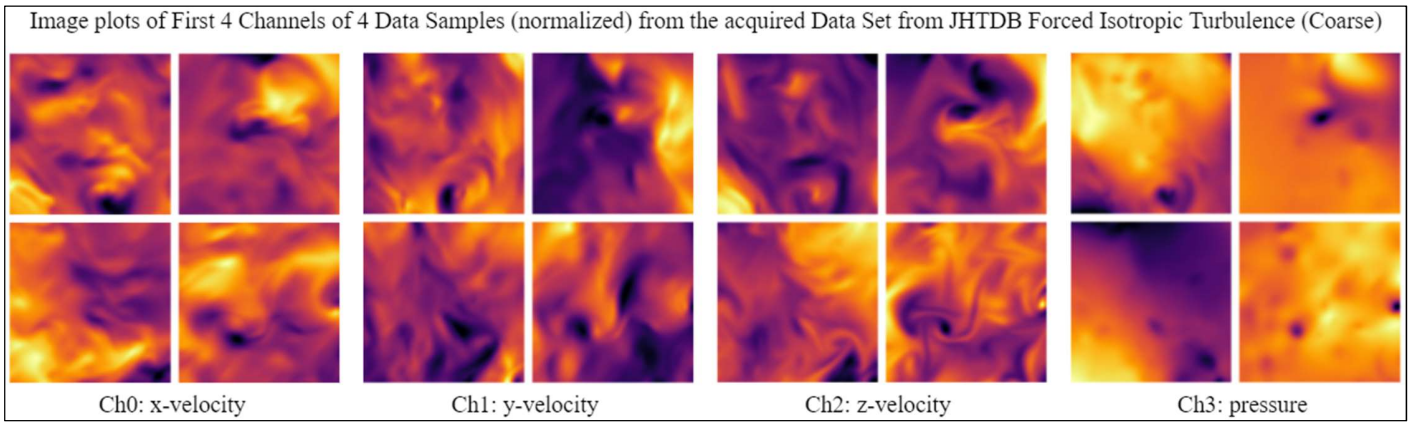


FIGURE 1 EXTRACTED DATA SAMPLES FROM THE ORIGINAL DATASETS

applications. Maski et al. [12] proposed the idea of stacked CAEs for hierarchical feature extraction. Du et al. [13] proposed a model by stacking denoising autoencoders in a convolutional manner. Research for advancements in autoencoders is thus ongoing.

Autoencoders have proved to be of great use in the order reduction of fluid flows. Fukami et al. [14] presented the assessment of an autoencoder which consisted of a convolutional neural network and multi-layer perceptron and evaluated the model based on parameters like number of latent modes, choice of activation function, and number of weights. Mohan et al. [15] proposed a physics-embedded convolution-based autoencoder (PhyCAE), by enforcing the incompressibility of the fluid flow within CNNs. Srinivasan et al. [16] demonstrated the potential of LSTM to predict turbulent statistics. Similarly, Nakamura et al. [17] presented a CNN-based AE combined with LSTM to predict the temporal turbulent flow. Cheng et al. [18] proposed a hybrid deep adversarial autoencoder by combining a variational autoencoder with a GAN for nonlinear flow prediction. With the proposal of these different autoencoder models, there arises a need to compare how accurately and efficiently do these different autoencoders encode and predict various fluid data. This study aims to parameterize a forced isotropic turbulence flow into statistical characteristics of turbulence by testing different models of convolutional and variational autoencoders along with Generative Adversarial Network. This study also aims to create models for reconstruction and generation of the turbulent flow from the extracted parameters to steer the turbulence modelling techniques towards neural network based systems instead of rigorous mathematical formulations and derivations.

2. MATERIALS AND METHODS

2.1 Dataset and Data preprocessing

The dataset used in training the models is from the JHTDB Forced Isotropic Turbulence Simulation [25,26,27]. 2400 64x64 random slices from the total of 1024x1024x1024 cells are extracted containing flow values in various channels. The extracted data is of the size 2400x16x64x64. The 16 channels contain values of x, y, z velocities (u, v, w), pressure (p), gradients (u_x, u_y, u_z, v_x, v_y, v_z, w_x, w_y, w_z) and pressure gradients (p_x, p_y, p_z). These are selected in order to extract their dependencies on each other and help in estimation of the flow parameters and eventually help in reconstruction and generation of the flow after passing it through the proposed Neural Network models.

The simulation parameters as given in [27] are as following:
Simulation Domain: 2π x 2π x 2π
Viscosity (ν) = 0.000185
Simulation time-step Δt = 0.0002
δt = 0.002
Time step stored between t=0 and 10.056: (5028 time samples separated by δt)

There are 9 characteristic flow parameters that are calculated for individual data slices of 16x64x64. Following characteristic parameters were calculated for each slice (ν = 0.000185 $E_{k_{int}}$=0.40864) [27]:

1. Total Kinetic Energy: $E_{tot} = \frac{1}{2}\langle u_i u_i \rangle$
2. Dissipation Factor: $\varepsilon = 2\nu\langle S_{ij} S_{ij} \rangle$, where S_{ij} are the gradient values of velocity
3. RMS Velocity: $u' = \sqrt{\frac{2}{3} E_{tot}}$
4. Taylor Micro Scale: $\lambda = \sqrt{15\,\nu\,u'^2/\varepsilon}$
5. Taylor-Scale Reynolds No.: $R_\lambda = u'\lambda/\nu$
6. Kolmogorov time scale: $\tau_n = \sqrt{\frac{\nu}{\varepsilon}}$
7. Kolmogorov length scale: $\eta = \nu^{3/4}\varepsilon^{-1/4}$
8. Integral scale: $L = \frac{\pi}{2u'^2} E_{k_{int}}$
9. Large eddy turnover time: $T = L/u'$

These parameters are stored for each data slice and forms a vector of 2400x9. This matrix is normalized using min-max

© 2021 by ASME

normalization and the values are mapped between -1 and 1 for the integration with the models. Same goes for the main dataset as well.

2.2 Autoencoders

Autoencoders are neural networks that are trained to match their output to the input. An autoencoder that passes the input through a code with dimensions less than the input is called undercomplete autoencoder (Figure 1). It has two parts: An encoder function $z = f(x)$ that encodes the input to a latent space and a decoder function $r = g(z)$ that reconstructs the input from this latent space and maps it to the output. The bottleneck layer forces the autoencoder to learn the most useful properties of the input. It is trained simply by minimizing the loss function $\mathcal{L}(x, g(z))$ such as MSE:

$$\mathcal{L}_{MSE}(x, g(z)) = \frac{1}{n}\left|\left|g(z) - x\right|\right|_2^2 \quad (1)$$

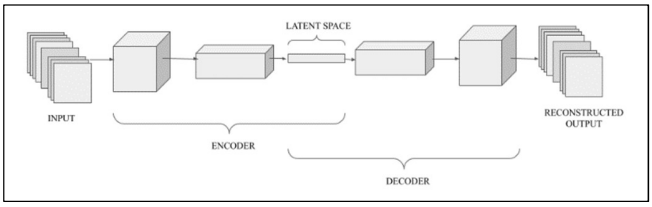


FIGURE 2: BASIC REPRESENTATION OF AN UNDERCOMPLETE AUTOENCODER

2.2.1 Variational Autoencoders

A variational autoencoder is a stochastic generative model that consists of a probabilistic decoder given by a likelihood function $p_\theta(x|z)$ and an encoder $q_\varphi(z|x)$ which approximates the posterior $p_\theta(z|x) \propto p_\theta(x|z)p(z)$ which is intractable, where $p(z)$ is a prior distribution over latent z. The decoder and encoder learn simultaneously by maximizing the variational lower bound, i.e.,

$$\mathcal{L}(x; \theta, \varphi) = \mathbb{E}_{q(Z|X)} \log p(x|z) - D_{KL}(q(z|x)||p(z)) \quad (2)$$

This equation represents the combination of reconstruction loss and the KL divergence of the approximate posterior and the prior of the latent variable z.

2.2.2 Convolutional Autoencoders

Convolutional Autoencoders combine the traditional convolutional networks with an autoencoder architecture. The encoder consists first of convolutional and max pooling layers as described earlier. To further reduce dimensions, the tensor is then reshaped and fed into a series of linear layers that lead up to the latent layer.

The decoder then maps this latent layer to the output via a series of deconvolutional (transposed convolutional) or upsampling layers that upsamples the code to the original dimensions. In our model, upsampling layers were used as the deconvolutional layers seemed to introduce artifacts. The reconstructed output is checked against a loss function to minimized the per pixel reconstruction loss.

2.3 Generative Models

To generate new types of flows with the same initialization parameters, Generative models are used. Generative models are generally used to create new data instances with the help of some initialization variables and/or some predetermined inputs. They tend to capture the probability of the transformation of the inputs to the desired output and generate new data instances based on the probability distributions in the input vector space.

Here, we are aiming to generate new flow distributions for a certain set of flow parameters extracted from the JHTB dataset. We are aiming to utilize Conditional Deep Convolutional Generative Adversarial Networks with and without relativistic discrimination and Pix2Pix networks (as described here, [1]) with and without relativistic discrimination.

2.3.1 Conditional GAN

Generative Adversarial Networks usually have two models, the Generator (G) and the Discriminator (D). Both the Generator and the Discriminator are nonlinear functions like a multi-layer perceptron or a CNN, which maps the inputs to respective outputs.

In a classic GAN, G usually receives inputs **z** and tries to map the probability of getting the data **x**, D takes the mapped output of G and classifies it as to whether the output comes from the data **x** or generated by G. The aim is to train a G such that D makes a lot of mistakes in identifying the generated data from the original. [22]

The generator and the discriminator models are trained together and the aim is to adjust the parameters of G such that the (1- loss of D with the generated output of G) is to be minimized and the loss of D with real data points is also minimized at the same point.

In Conditional Adversarial Networks, an extra information **y** is fed to both G and D. Both are conditioned on the extra information **y**. Let the generator is fed a noise of the distribution $p_z(z)$, along with **y**. While training, G tries to maximize the following expression, and D tries to minimize it:

$$\mathbb{E}_x\left[\log\left(D(x|y)\right)\right] + \mathbb{E}_z\left[\log\left(1 - D\left(G(z|y)\right)\right)\right] \quad (3)$$

In the model that is adopted for the generation of flows, the input **y** is a vector containing a normalized array of flow parameters that were extracted by processing individual instances of elements from the main data set **x.**

The generator G is made of both hidden layers as well as a CNN-based upscaling model for image creation. The G gets **y** and **z** as the input, where **z** is the latent variable from a normal distribution with a mean of 0 and a standard deviation of 1. The discriminator D gets **y** and either the image slice **x** or G(z|y) which is the generator output as its inputs and outputs a single Float value as the validity indicator.

© 2021 by ASME

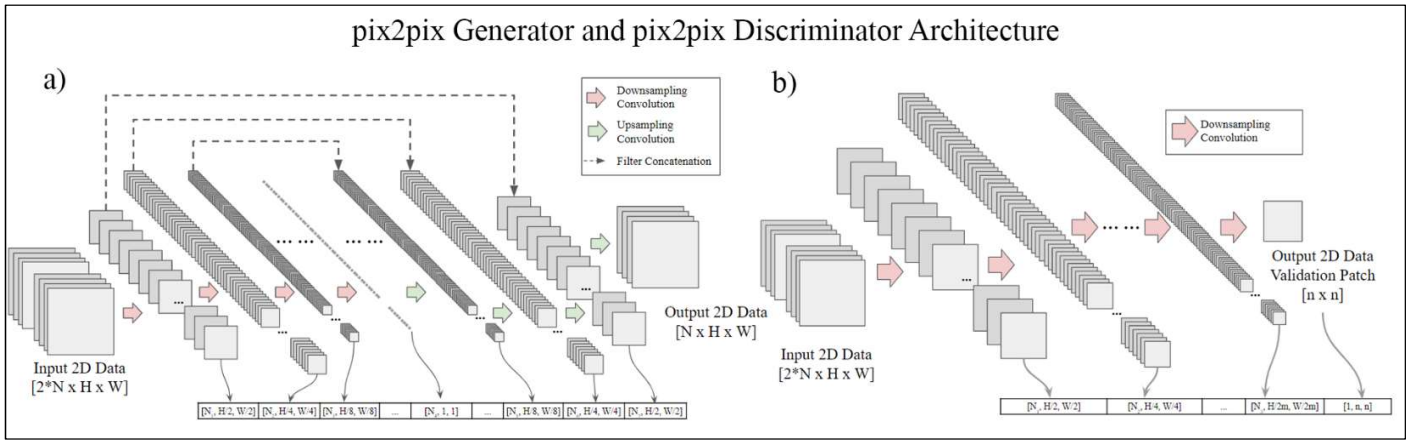


FIGURE 3: PIX2PIX ARCHITECTURE. A) DESCRIBES GENERATOR ARCHITECTURE ALONG WITH THE LAYERS THAT ARE ARRANGED TO FORM THE OUTPUT. B) DESCRIBES THE DISCRIMINATOR. N, W, H REPRESENTS THE NUMBER OF CHANNELS, THE WIDTH AND THE HEIGHT IN THE INPUT IMAGE DATA RESPECTIVELY.

In our implementation, y is replaced by parameters p which along with some normally distributed noise z are used as input for the generator G to produce 'fake' data. The Generator loss can be expressed as:

$$L_G = \mathbb{E}_{adv}[D(G(z|p), p), V] \tag{4}$$

where V is a validity vector of ones as its elements and $\mathbb{E}_{adv}$ represents the adversarial error function which is a mean Squared Error function in our case. L_G is then backpropagated through G using the Adam optimizer at a suitable learning rate.

As for Discriminator D, the fake and real losses can be expressed as:

$$V_{real} = D(x, p) \tag{5}$$
$$V_{fake} = D[G(z|p), p] \tag{6}$$
$$L_{real} = \mathbb{E}_{adv}(V_{real}, V) \tag{7}$$
$$L_{fake} = \mathbb{E}_{fake}(V_{fake}, F) \tag{8}$$
$$L_D = \frac{1}{2}(L_{real} + L_{fake}) \tag{9}$$

where F is a zero vector to account for 'truly fake' data. L_D is then backpropagated through D using the Adam optimizer to modify the model parameters. Both G and D run against eachother and should settle to a stable L_G and L_D values to appropriately train the model.

2.3.2 Pix2Pix Architecture

Convolution-based image to image generative networks have become immensely popular nowadays. They usually consist of a Convolutional network that is trained to 'morph' the input 2D image similar to the desired output. pix2pix is one such GAN architecture [23]. Pix2pix is generally used for style transfer, image colorization, label identification, image reconstruction from edge maps, etc. Here we are using this approach to translate a randomly generated noise map to generate turbulent flows.

The pix2pix implementation in the previously mentioned reference is based upon 'U-Net' architecture [24]. The U-Net approach utilizes skip connections in between convolutional layers to retain the previously encoded layers.

Let the input image instance be x, let z be the random noisy input and y be the desired output image instance that the generator G needs to map x and z towards y or G: {x, z} → y. The Discriminator is also trained parallelly to distinguish the real images from the fake or the generated ones.[24] The objective function of such conditional GAN can be described as

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log(D(x|y))] + \mathbb{E}_{x,z}[\log(1 - D(x|\, G(x|z)\,))] \tag{10}$$

where G is trying to minimize the above objective function and D is trying to maximize it. An unconditioned variant in which the discriminator doesn't observe x is also taken into consideration as

$$\mathcal{L}_{GAN}(G, D) = \mathbb{E}_{y}[\log(D(y))] + \mathbb{E}_{x,z}[\log(1 - D(G(x|z))\,)] \tag{11}$$

From experimentation, it was concluded that if we use L2 loss, the data output was not very clear to comprehend and most of its details were lost in doing so. A direct L1 loss relation was utilized to produce more clear data to 'fool' the discriminator:

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}\left[\left||y - G(x|z)|\right|_1\right] \tag{12}$$

The final objective function becomes,

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G) \tag{13}$$

In our implementation, we are using a similar model structure as the one given in original reference [23], which has a CNN based generator with skip layers in between.

For each instance of training, two sets of data are created, one with the original data set **x** and other **x'** which is $\mathbf{x} + \mathbf{z_n}$ where

© 2021 by ASME

$\mathbf{z_n}$ is a 2D noise distribution. Generator also receives some external parameters which like data can be written as **p** and **p'** as the unmodified and modified parameters respectively. The generator output can be written as G(x|p) or G(x'|p') depending upon the inputs given. The discriminator D is straightforward another neural network-based model which receives two data instances and itself gives out an n x n dimensional array.

Let the modified data instance be **x'**, **p'** and the unmodified data be **x**, **p** and let these variables represent the 'real' data. Then the overall GAN loss and pixelwise loss can be written as

$$L_{GAN} = \mathbb{E}_{GAN}[D(\, G(x'|p'')\, , x), V] \tag{14}$$

$$L_{px} = \mathbb{E}_{px}[G(x'|p'), x] \tag{15}$$

where V is the validity matrix of ones with [n x n] dimensions. $\mathbb{E}_{GAN}$ and $\mathbb{E}_{px}$ is the error functions used, where $\mathbb{E}_{GAN}$ is the Mean Squared error function or L2 loss and $\mathbb{E}_{px}$ is calculating the absolute error or the L1 loss.
The total generator loss becomes,

$$L_G = L_{GAN} + \lambda\, L_{px} \tag{16}$$

L_G is then backpropagated through the Generator model using Adam optimizer.

For the Discriminator loss, the discriminator's prediction for real and fake data are compared through the $\mathbb{E}_{GAN}$ function to generate the real and fake loss. Mathematically this can be expressed as,

$$P_{real} = D(x,\ x') \tag{17.1}$$
$$L_{real} = \mathbb{E}_{GAN}(P_{real},\ V) \tag{17.2}$$
$$P_{fake} = D(G(x', p'),\ x') \tag{17.3}$$
$$L_{fake} = \mathbb{E}_{GAN}(P_{fake},\ F) \tag{17.4}$$

Where V again is an [n x n] matrix of ones and F is an [n x n] matrix of zero to signify a truly fake data instance. The total Discriminator loss can be expressed as,

$$L_D = \frac{1}{2}\left(L_{real}\ +\ L_{fake}\right) \tag{18}$$

Where L_D is again backpropagated through the discriminator model through Adam optimizer.

The variation in the training can be done by implementing various data input techniques during training like changing the $\mathbf{z_n}$ distribution function by either changing the sample range or by changing the noise generator altogether. The amount of original information that is going into the GAN along with the noise can also be varied to get different kinds of outputs.

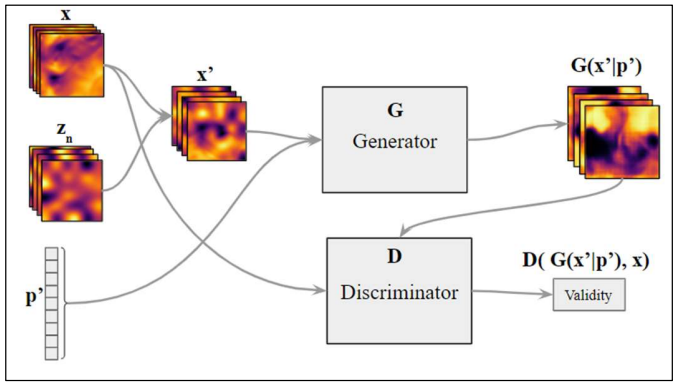


FIGURE 4: PIX2PIX DATA FLOW THOUGH THE GENERATOR G AND THE DISCRIMINATOR D

3 RESULTS AND DISCUSSION

3.1 Autoencoders

3.1.1 Convolutional Autoencoder

Convolutional Encoder takes in the 64 × 64 input with 16 channels and applies convolution and max pooling layers to reduce its size and increase the channel width. After going through multiple layers, it splits into parametric and latent vector space. The decoder concatenates these to pass through linear layers similar to the encoder and finally the output with the required four channels is reconstructed through a series of UpSample layers. Other channels play their part in gradient computation. The Convolutional Autoencoder model can be more clearly understood in the figure.

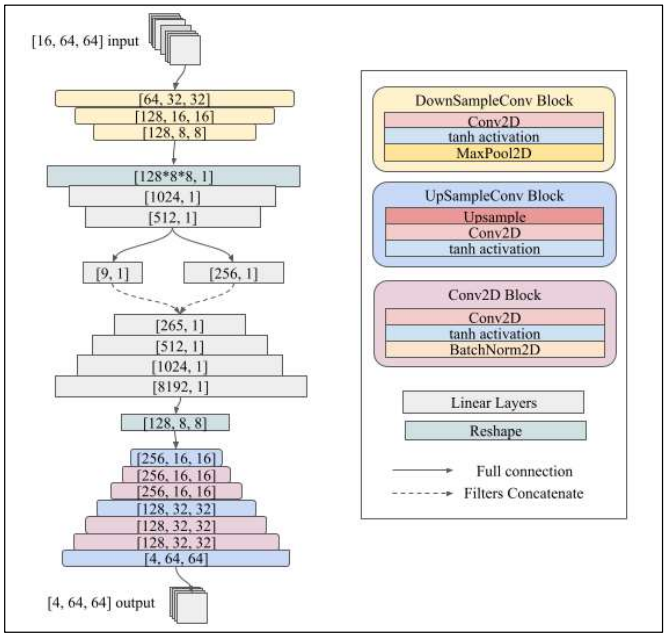


FIGURE 5: CONVOLUTIONAL AUTOENCODER MODEL

© 2021 by ASME

The summary and various hyperparameters are summarized in Table 1. Schedulers for the model and the encoder are also used to vary the learning rate after few epochs of stagnation.

Figure 10 shows the reconstruction and parameter losses for the convolutional autoencoder. Figure 8 shows the reconstruction of 5 random images from the dataset. The reconstruction seems to capture the predominant features in most cases, barring some outliers, though there is little detail in the reconstructions. Individual channel reconstructions can be seen in Figure 8.

Using the convolutional autoencoder as a generative model is demonstrated in Figure 5, which shows the linear interpolation between two images which generates new images at different parameter values.

3.1.2 Variational Autoencoder

Variational Autoencoder also uses convolutional and upsampling layers but works on a probabilistic principle as described earlier. Figure 7 shows the structure of the Variational autoencoder in detail. It also employs the concept of skip connections in its encoder and decoder blocks as shown in Figure 7. The summary and various hyperparameters are also summarized in Table 1.

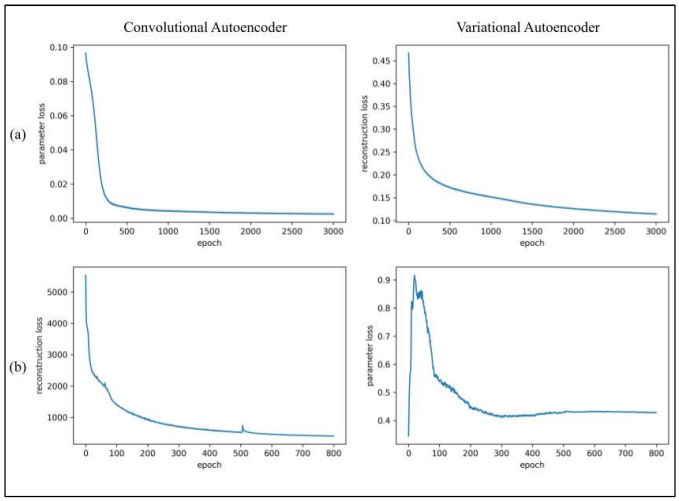


FIGURE 6: (a) RECONSTRUCTION AND (b) PARAMETER LOSS CURVES FOR CONVOLUTIONAL AND VARIATIONAL AUTOENCODERS

Figure 6 shows the reconstruction and parameter losses for the variational autoencoder. The reconstruction loss is the sum of losses of all individual pixels in the four channels. 5 random images from the dataset are fed into the autoencoder and the results are shown in Figure 16. The reconstructed images have good detail and capture most of the features, though some images stand out as “bad” reconstructions. Velocity and pressure channel reconstructions can be seen in Figure 8.

Interpolation in variational autoencoder results in images with moderately detailed features that smoothly transition into one another. It outperforms the convolutional autoencoder in terms of feature clarity as shown in Figure 8.

Figure 10 shows the reconstruction and parameter losses for the convolutional autoencoder. Summary of the results can be seen in Figure 8. Figure 8(a) shows the reconstruction of velocity and pressure channels of a random image from the

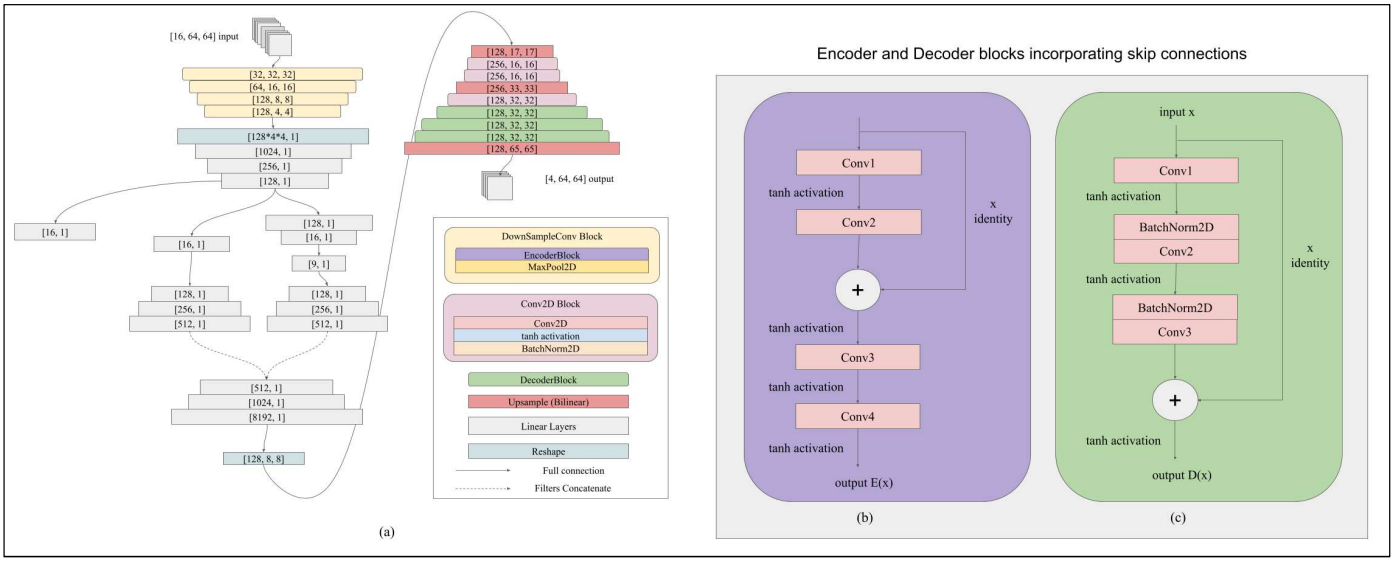


FIGURE 7: (a) VARIATIONAL AUTOENCODER MODEL, (b) ENCODER BLOCK AND (c) DECODER BLOCK

© 2021 by ASME

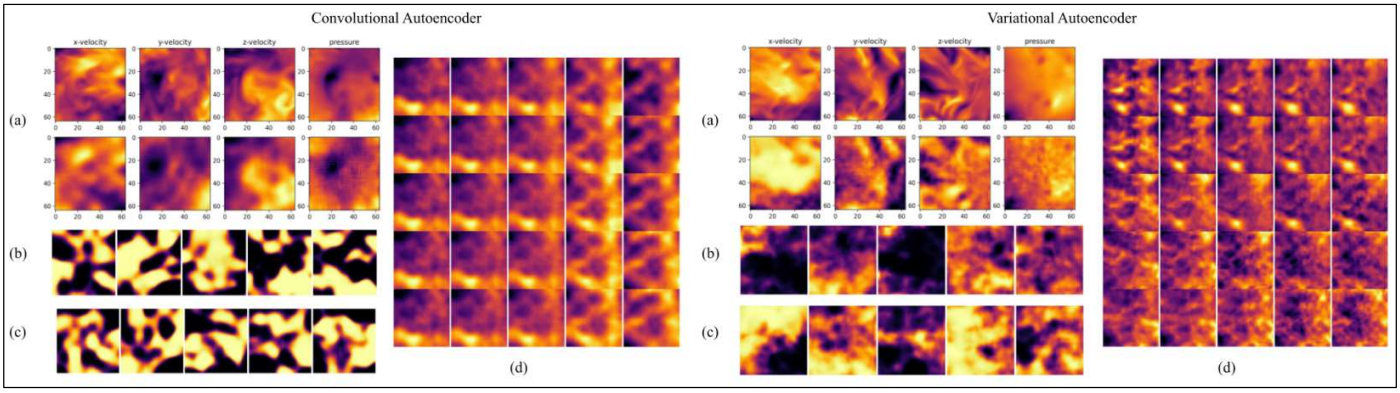


FIGURE 8 RESULTS OF CONVOLUTION AND VARIATIONAL AUTOENCODERS. A) IMAGE RECONSTRUCTION, B) GENERATED SAMPLES (FIXED PARAMETERS), C)GENERATED RANDOM SAMPLES, D) CH:1 OUTPUT BY LINEAR INTERPOLATION OF PARAMETERS

dataset. The reconstruction seems to capture the predominant features in most cases, barring some outliers, though there is little detail in the reconstructions. Using the convolutional autoencoder as a generative model is demonstrated in Figure 8(b) and 8(c), which shows the linear interpolation between two images which generates new images at different parameter values.

TABLE 1 TRAINING DETAILS FOR AUTOENCODER BASED MODELS

Model	CNN-AE	VAE
Number of epochs	3000	800
Optimizer	SGD	Adam
Model Learning Rate	0.001	0.00001
Encoder Learning Rate	0.0005	0.000005
Batch size	5	1
Criterion	L1 Loss	MSE Loss
Activation function(s)	tanh	tanh
Weight decay	1E-5	1E-5

3.2 Generative Adversarial models

Training generative models is a very difficult process as two losses are adversarial and are running against each other. Sometimes these models tend to become very 'explosive' and tip towards favouring either the generator or the discriminator especially in cases where the inputs and outputs have very little defined correlation and are almost like a form of noise distribution itself

3.2.1 Conditional GAN

The conditional GAN architecture as explained before consists of a Generator and a Discriminator. The Generator is itself a combination of MLP and CNN models. The discriminator is MLP based model which checks the validity of the input data and distinguishes it from the 'real' and 'fake' data.

The actual model architecture can be explained by the Fig 9,10.

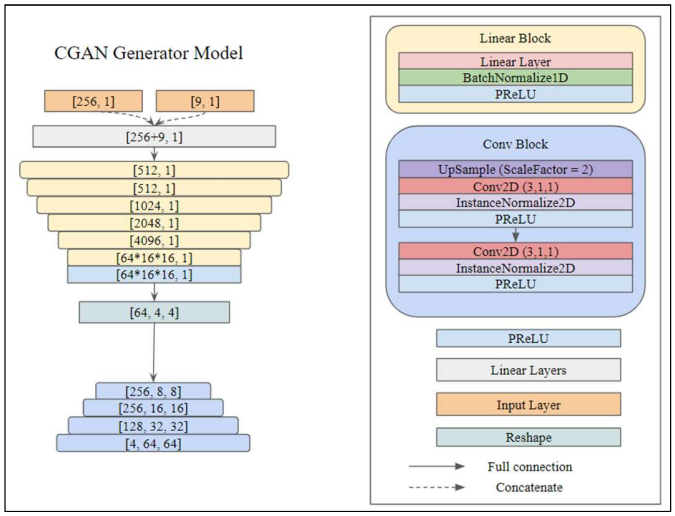


FIGURE 9: CGAN GENERATOR MODEL

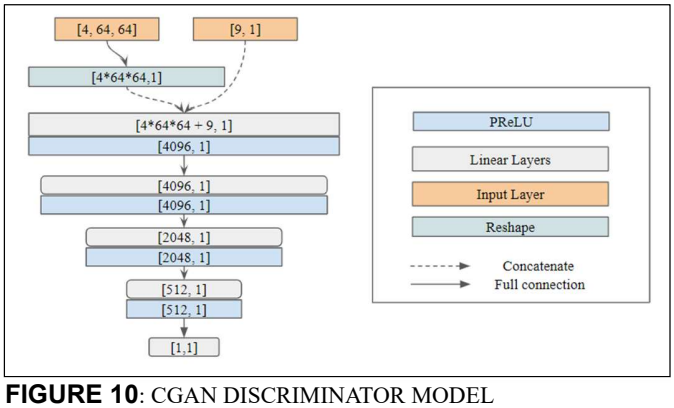


FIGURE 10: CGAN DISCRIMINATOR MODEL

© 2021 by ASME

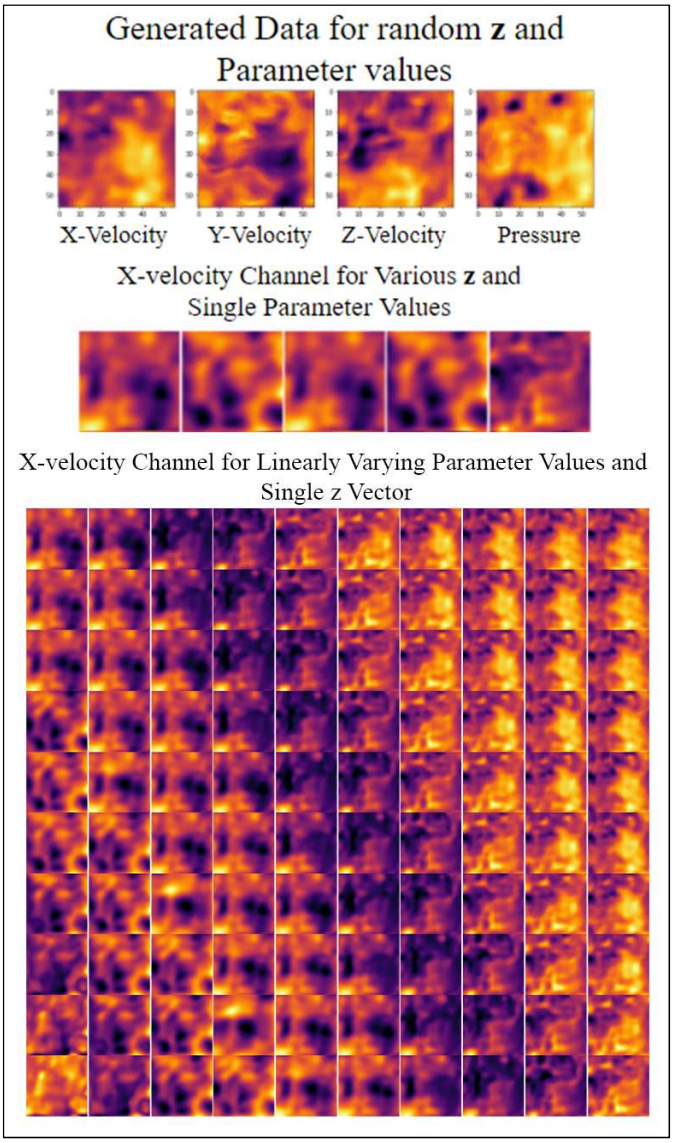


FIGURE 11: CGAN RESULTS

The generator outputs the x-velocity, y-velocity, z-velocity, and pressure of a single instance. The results can be seen from the Figure 11. The latent vector here is a normal distribution of 256 random numbers with a mean 0 and standard deviation of 1. The 9 flow parameters are also randomized. As it is trained on normalized parametric data, the results should mimic the pattern in the original data Fig 1.

By varying the parameter p, various conditions of flow can be created. From Figure 10, the conversion of flows into one another can be seen clearly. The overall performance of this model can be called satisfactory but not great.

3.2.2 pix2pix GAN Implementation

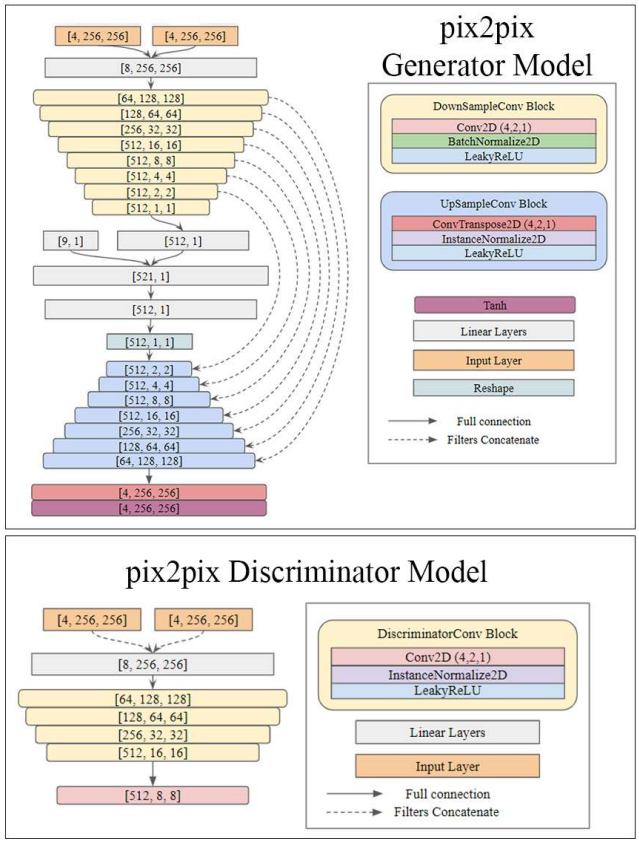


FIGURE 12 DETAILED PIX2PIX GENERATOR AND DISCRIMINATOR MODEL WITH MENTIONED OUTPUT LAYER SIZES AND TYPES

As mentioned earlier, pix2pix architecture has been utilized to learn the flow patterns and retain information of x, y, z velocity, and pressure of various data slices and produce a new set of data by taking the original data and a 2D random distribution. The architecture of the models is given here in the figures

This kind of architecture demands 2 data inputs to transfer the 'style' from one to the other. As mentioned earlier, X and x' are the inputs of the model where x' is the noise added image. x' can be expressed as a combination of the images from the dataset x and z_n which is the noise distribution, which can be expressed as

$$x' = \alpha x + \beta z_n \tag{19}$$

Where α and β are set manually to train different models. Three such training schemes have been demonstrated (Fig 13):

1. Model **M1** ($\alpha = 0$, $\beta = 1$)
   z_n is a 4x8x8 grid of random values of a normal distribution with mean = 0 and standard deviation = 1. z_n is upscaled to 4x256x256 by 'bicubic' interpolation.

© 2021 by ASME

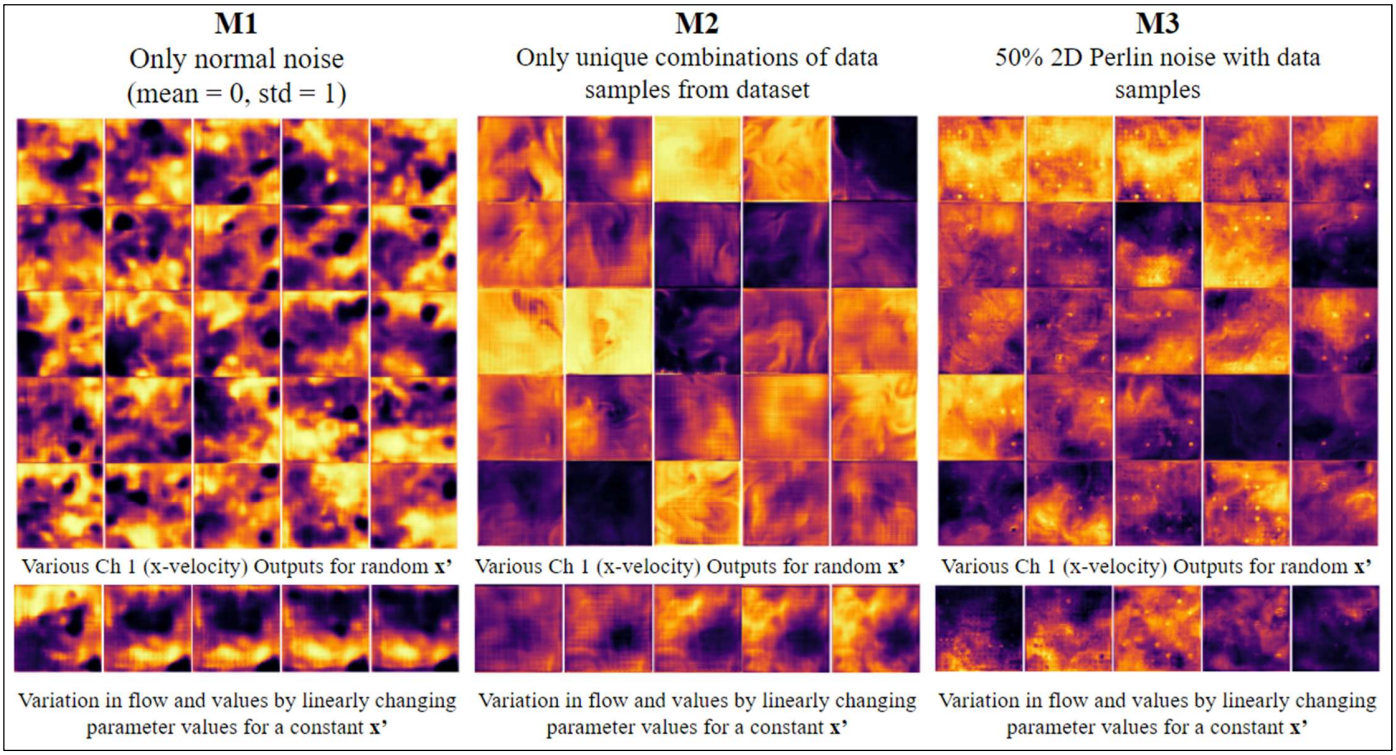


FIGURE 13 PIX2PIX RESULTS

2. Model **M2** ($\alpha = 1$, $\beta = 0$)
   Different instances (different from the 2nd input) from the data are fed as the 'modified' input.
3. Model **M3** ($\alpha = 0.5$, $\beta = 1$)
   Here, z_n is an 8x8 grid of 2D Perlin noise which produces procedural organic textures in the 8x8 grid for every 4 channels. z_n is upscaled to 4x256x256 by 'bicubic' interpolation. 50% of additional data slices from the dataset is added to z_n to form complete 'modified' input.

After adding the noise functions, it could be observed that the M1 model suffers severely from the influence of the initiation noise **x'**. M2 and M3 produce natural-looking flow velocity patterns and are comparable to the original images of the data set (Fig 1). Since M2 only had 2 different sets of inputs from the dataset itself, the generation is well defined and preserves details as well. The 'style transfer' properties of the pix2pix architecture can be seen easily. Fine flow lines and vortices can be noticed easily. M2 model suffers from low range values reproductions which signifies that the flow values are stuck in some narrow range that produces low contrast images. M3 produces excellent details and a greater value range in the generator output, but it suffers from 'artifact' creation in some cases. This can be reduced by training M3 for longer epochs.
The influence of the flow parameters can be seen on every model (Fig 13). Slight variance in the flow parameters can change the outputs in a very drastic manner.
Training details for the Generative models are shown in Table 2

TABLE 2 GENRATIVE MODELS TRAINING SUMMARY

Model	CGAN	pix2pix M1	pix2pix M2	pix2pix M3
Optimizer	Adam	Adam	Adam	Adam
Learning rate	0.0002	0.0002	0.0002	0.0002
Batch size	5	5	10	10
Epoch	2000	1500	1500	1500
Generator Loss (avg)	0.17154	0.46542	0.28369	0.25869
Discriminator Loss (avg)	0.48476	0.10318	0.25096	0.25012

4 CONCLUSION

Various autoencoder and GAN based models have been proposed and studied to reconstruct and generate natural looking fluid flows. The proposed autoencoders can be used to compress the information of the input fluid field to a few normalized parametric values which can be passed on to the decoders to generate new outputs based on the compressed parameters. In both the CNN and Variational CNN autoencoders, the decoder can work independently as a flow generator and is significantly dependent upon the 9 flow parameters extracted from the data. Similar results can be seen in both GAN implementations. CGAN generator produces comparable flow patterns to the dataset. The 'pix2pix' model which is generally used for 'style-transfer' has also proved to be very efficient in extracting information from the parameters and generating unique flows using an initiating noise distribution. The M3 implementation has proved to produce highly detailed generated flows but suffers from 'artifact creation' in some images.

© 2021 by ASME

It can be concluded that the proposed neural network based models provide a fairly less computationally expensive framework to generate and reconstruct turbulent flows based on statistical parameters. Similar models can be utilized in many areas including professional CFD applications to simulate fluid flows and in generating professional textures and effects in video game creation and VFX simulation using more elaborate 3D implementation.

Future work consists of making the strongest performing models better by more rigorous training and with a larger dataset.

ACKNOWLEDGEMENTS

The authors would like to thank Fluid Systems Laboratory, Delhi Technological University for providing computing facilities. We are thankful to our professor Dr. Raj Kumar Singh who provided expertise which greatly assisted us in the research and improved the paper significantly. We would also like to thank other members of the Fluid Mechanics Group for directly or indirectly helping during the research work.

REFERENCES

[1] Taira, Kunihiko, et al. "Modal analysis of fluid flows: Applications and outlook." *AIAA journal* 58.3 (2020): 998-1022.

[2] Akhtar, Imran, Ali H. Nayfeh, and Calvin J. Ribbens. "On the stability and extension of reduced-order Galerkin models in incompressible flows." *Theoretical and Computational Fluid Dynamics* 23.3 (2009): 213-237.

[3] Iollo, Angelo, Stéphane Lanteri, and J-A. Désidéri. "Stability properties of POD–Galerkin approximations for the compressible Navier–Stokes equations." *Theoretical and Computational Fluid Dynamics* 13.6 (2000): 377-396.

[4] Gupta, Priyam, Prince Tyagi, and Raj Kumar Singh. *Analysis of Generative Adversarial Networks for Data Driven Inverse Airfoil Design*. No. 5975. EasyChair, 2021.

[5] San, Omer, and Romit Maulik. "Extreme learning machine for reduced order modeling of turbulent geophysical flows." *Physical Review E* 97.4 (2018): 042322.

[6] San, Omer, Romit Maulik, and Mansoor Ahmed. "An artificial neural network framework for reduced order modeling of transient flows." *Communications in Nonlinear Science and Numerical Simulation* 77 (2019): 271-287.

[7] Pawar, Suraj, et al. "Data-driven recovery of hidden physics in reduced order modeling of fluid flows." *Physics of Fluids* 32.3 (2020): 036602.

[8] Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." *arXiv preprint arXiv:1312.6114* (2013).

[9] Gregor, Karol, et al. "Draw: A recurrent neural network for image generation." *International Conference on Machine Learning*. PMLR, 2015.

[10] Burda, Yuri, Roger Grosse, and Ruslan Salakhutdinov. "Importance weighted autoencoders." *arXiv preprint arXiv:1509.00519* (2015).

[11] Sønderby, Casper Kaae, et al. "Ladder variational autoencoders." *arXiv preprint arXiv:1602.02282* (2016).

[12] Masci, Jonathan, et al. "Stacked convolutional auto-encoders for hierarchical feature extraction." *International conference on artificial neural networks*. Springer, Berlin, Heidelberg, 2011.

[13] Du, Bo, et al. "Stacked convolutional denoising auto-encoders for feature representation." *IEEE transactions on cybernetics* 47.4 (2016): 1017-1027.

[14] Fukami, Kai, et al. "Model order reduction with neural networks: Application to laminar and turbulent flows." *arXiv preprint arXiv:2011.10277* (2020).

[15] Mohan, Arvind T., et al. "Embedding hard physical constraints in neural network coarse-graining of 3d turbulence." *arXiv preprint arXiv:2002.00021* (2020).

[16] Srinivasan, Prem A., et al. "Predictions of turbulent shear flows using deep neural networks." *Physical Review Fluids* 4.5 (2019): 054603.

[17] Nakamura, Taichi, et al. "Convolutional neural network and long short-term memory based reduced order surrogate for minimal turbulent channel flow." *Physics of Fluids* 33.2 (2021): 025116.

[18] Cheng, M., et al. "An advanced hybrid deep adversarial autoencoder for parameterized nonlinear fluid flow modelling." *Computer Methods in Applied Mechanics and Engineering* 372 (2020): 113375.

[19] Li, Yi, et al. "A public turbulence database cluster and applications to study Lagrangian evolution of velocity increments in turbulence." *Journal of Turbulence* 9 (2008): N31.

[20] Perlman, Eric, et al. "Data exploration of turbulence simulations using a database cluster." *Proceedings of the 2007 ACM/IEEE conference on Supercomputing*. 2007.

[21] Isola, Phillip, et al. “Image-to-Image Translation with Conditional Adversarial Networks.” *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, doi:10.1109/cvpr.2017.632.

[22] Mehdi Mirza and Simon Osindero, . "Conditional Generative Adversarial Nets". CoRR abs/1411.1784. (2014).

[23] Phillip Isola and Jun-Yan Zhu and Tinghui Zhou and Alexei A. Efros, . "Image-to-Image Translation with Conditional Adversarial Networks". *CoRR* abs/1611.07004. (2016).

[24] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, 2015

[25] Y. Li, E. Perlman, M. Wan, Y. Yang, R. Burns, C. Meneveau, R. Burns, S. Chen, A. Szalay & G. Eyink. "A public turbulence database cluster and applications to study Lagrangian evolution of velocity increments in turbulence". *Journal of Turbulence **9**, No. 31*, 2008.

[26] E. Perlman, R. Burns, Y. Li, and C. Meneveau. "Data Exploration of Turbulence Simulations using a Database Cluster". *Supercomputing SC07*, ACM, IEEE, 2007.

[27] Minping, W., Chen, S., Eyink, G., Meneveau, C., Johnson, P., Perlman, E., Burns, R., Li, Y., Szalay, A., and Hamilton, S., 2012, “Forced Isotropic Turbulence Data Set (Extended).

 © 2021 by ASME

GREY WOLF OPTIMIZATION TUNED TID AND I-TD CONTROLLERS FOR TRAJECTORY TRACKING OF DYNAMICAL AERIAL SYSTEM

Karan Sehgal, Manik Verma
E-Mail Id: karansehgal_2k18ee089@dtu.ac.in, manikverma0200@gmail.com
Delhi Technological University, Delhi, India

Abstract- The mathematical modelling and linearization of the laboratory non-linear aerial dynamical system, the Twin rotor MIMO system (TRMS), are presented in this paper. It can be used to solve real-world problems and to test the efficiency of controllers in control engineering. To achieve stable hovering conditions, a feedback control system with Tilted-Integral Derivative (TID) and Integral-Tilted Derivative (I-TD) controllers is linked within the continuous feedback loop for improving the control signal and varying the operational gains accordingly, taking into account high order nonlinearities and TRMS instability. Grey Wolf Optimization, a reliable and powerful metaheuristic evolutionary algorithm, is used to optimize the designed controllers. Manual disturbances are applied to both the yaw and pitch actuator angle stabilization controllers to validate their accuracy.

Keywords: Dynamical Aerial System, TID, I-TD, Grey Wolf Optimization, TRMS, Controller Design

1. INTRODUCTION

TRMS is a benchmarked experimental laboratory model of an unmanned aerial helicopter system that resembles or is called a prototype model. Pitch and yaw angle motion is controlled by two actuators driven by 12V, 44-Watt D.C. servomotors. They are positioned on both edges of the connecting beam, which is counterbalanced by an arm with adjusting weight at its ends [1]. The unknown dynamical characteristics increase system nonlinearities due to heavy cross-coupling among pitch and yaw rotors [2]. Several conventional and robust feedback controllers are designed to manage such complex systems. PID controllers are illustrated and chosen in dynamical systems due to their simple structural architecture [3]. The fractional-order PID controller is often used to stabilize TRMS to limit the deficiencies of integral and derivative schemes [4]. In addition, as shown in [5], [6], [7], the FOPID controller gains can be tuned using various nature-inspired multi-objective algorithms. Various modern and robust control strategies such as multiloop linear quadratic regulator [8], sliding mode [9], and robust H-infinity control [10] have been used. Another kind of fractional order PID controller known as Tilted Proportional Controller is proposed in [11] in line with these robust and modern controllers. TID controllers are used to controlling a variety of non-linear dynamical systems, including magnetic levitation [12] and two different types of LFC control [13]. The paper provides a unique comparison of TID and I-TD controls on laboratory aerial systems that have been benchmarked (TRMS). In addition, grey wolf optimization [14], [15] is used to tune the built controllers. The present paper is divided into five sections: Section I contains an introduction, section II contains mathematical modelling of TRMS, and section III contains Grey Wolf Optimized controller design and tuning. The precise comparison and graphical results are presented in Section IV, accompanied by conclusions and references.

2. DYNAMICAL MODELLING OF THE BENCHMARKED AERIAL SYSTEM (TWIN ROTOR MIMO SYSTEM)

The TRMS laboratory configuration, as shown in Fig. 2.1, is designed to mimic the characteristics of a helicopter, and has earned a reputation as a benchmarked device for research and development.

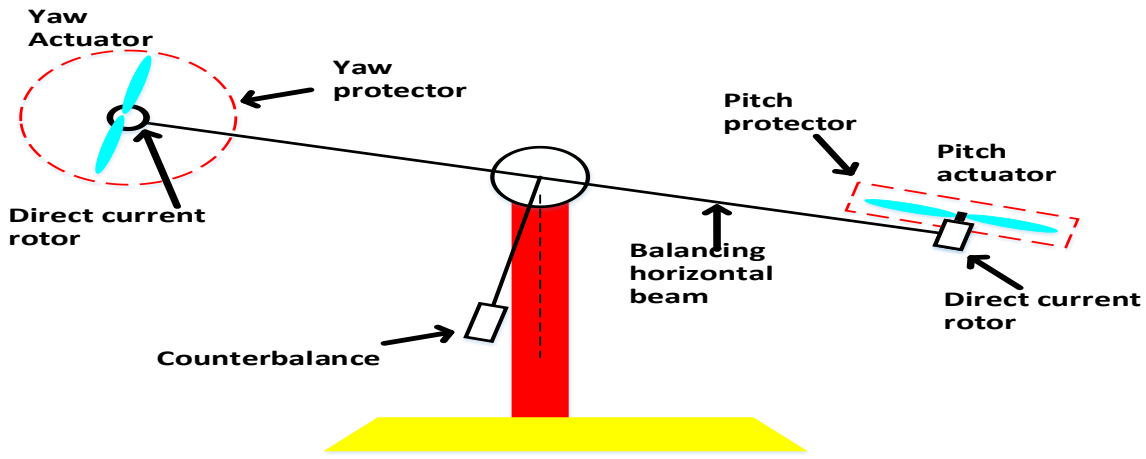


Fig. 2.1 Benchmarked laboratory aerial setup (TRMS)

The modelling equations of the Twin Rotor Aerial System (TRAS) for vertical hovering are given as equations (1) to (5).

@2017, IJTRS All Right Reserved, www.ijtrs.com

$$I_1\ddot{\psi} = W_1 - W_{FG} - W_{B\psi} - W_G \quad (1)$$

Where,

$$W_1 = a_1.\tau_b + b_1\tau_a \text{ are the nonlinear characteristics} \quad (2)$$

$$W_{FG} = W_g.\sin\psi \text{ is gravity momentum} \quad (3)$$

$$W_{B\psi} = B_{1\psi}.\dot{\psi} + B_{2\psi}.\text{sign}(\dot{\psi}) \quad (4)$$

$$W_G = K_{gy}.M_1.\dot{\phi}.\cos\psi \text{ is gyroscopic momentum} \quad (5)$$

The pitch actuator's electrical circuit is given by the s-domain transfer function in (6):

$$\tau_a = \frac{k_1}{T_{11}+T_{10}}.u_1 \quad (6)$$

For horizontal hovering, the modelling equations are given as equations (7)-(9).

$$I_2.\ddot{\phi} = W_2 - W_{B\phi} - W_R \quad (7)$$

$$W_2 = a_2.\tau_2^2 + b_2.\tau_2 \quad (8)$$

$$W_{B\psi} = B_{1\phi}.\dot{\psi} + B_{2\phi}.\text{sign}(\dot{\phi}) \quad (9)$$

Also, the cross-coupling moment M_R is approximated as equation (10).

$$W_{CR} = \frac{k_c(T_0 s+1)}{T_p s+1}.\tau_1 \quad (10)$$

The yaw DC actuator with a circuital loop in the s-domain momentum transfer function is described by equation (11).

$$\tau_b = \frac{k_2}{T_{21}s+T_{20}}.u_2 \quad (11)$$

On solving the equations (1) to (9), we obtain $d(\psi)$ and $d(\phi)$ for main and tail rotors given as equations (12) and (13) respectively.

$$d(\psi) = \frac{1}{I_1}\left[-\int B_1.\psi.d(\psi) + \int\left(0.0163 * \sin 2\psi.\left(d(\phi)\right)^2 - \int\left(W_{fg}.\sin\psi\right)dt\right)\right] \quad (12)$$

$$d(\phi) = \frac{1}{I_2}\left[-\int B_1.\phi.d\phi + \int\left(b_1\frac{1.1u_1}{1.2s+1} + a_1\left(\frac{1.1u_1}{1.2s+1}\right)^2\right)dt\right] \quad (13)$$

After putting the system parameter values and constants in Table-1, a linearized state-space model transfer function of TRMS has been derived as equation (14)

$$T(s) == \begin{bmatrix} T_{11}(s) & T_{12}(s) \\ T_{21}(s) & T_{22}(s) \end{bmatrix} = \begin{bmatrix} \frac{1.247}{(s+0.843)(s^2+0.08624s+4.707)} & 0 \\ \frac{1.472(s+0.2867)}{s(s+5)(s+0.839)(s+0.5)} & \frac{3.4}{s(s+5)(s+1)} \end{bmatrix} \quad (14)$$

The derived transfer functions for the pitch rotor are given as equation (15), and the two poles of the pitch transfer function are in RHS and cross the imaginary axis in the root locus plot in Fig. 2.2, indicating that the system is unstable as the system gain increases.

$$T(s)_p = \frac{1.247}{s^4+1.83s^3+5.617s^2+8.265s+3.563)} \quad (15)$$

Table -2.1 Modelling Constants and Static Gains

TRMS modelling Constants	**Values**
$\mathbf{I_1}$- Momentum of the Main actuator	$7.9 \times 10^{-2}\text{Kgm}^2$
$\mathbf{I_2}$- Momentum of the Tail actuator	$4 \times 10^{-3}\text{Kgm}^2$
$\mathbf{a_1}$- Main actuator Static Gain	0.4134
$\mathbf{b_1}$- Main actuator Static Gain	0.655
$\mathbf{a_2}$- Tail actuator Static Gain	0.67
$\mathbf{b_2}$- Tail actuator Static Gain	0.74
$\mathbf{W_g}$- Gravity Moment	0.399
$\mathbf{B_{1\psi}}$- Frictional Moment of Pitch Propeller	$5 \times 10^{-0.6}$
$\mathbf{B_{2\psi}}$- Frictional Moment of Pitch Propeller	$0.7 \times 10^{-0.6}$

DOI Number: https://doi.org/10.30780/specialissue-ICAASET021/004 **pg. 21**
Paper Id: IJTRS-ICAASET2021-004
@2017, IJTRS All Right Reserved, www.ijtrs.com

$\mathbf{B_{1\varphi}}$- Friction Moment of Yaw Propeller	$2 \times 10^{-0.5}$
$\mathbf{B_{2\varphi}}$- Friction moment of Yaw Propeller	$3.5 \times 10^{-0.5}$
$\mathbf{K_{gy}}$- Gyroscopic Moment	0.053
$\mathbf{k_1}$- DC Constant in the Main actuator	1.45
$\mathbf{k_2}$- DC Constant in the Tail actuator	0.79
$\mathbf{T_{11}}$- Moment of the Main actuator	1.01

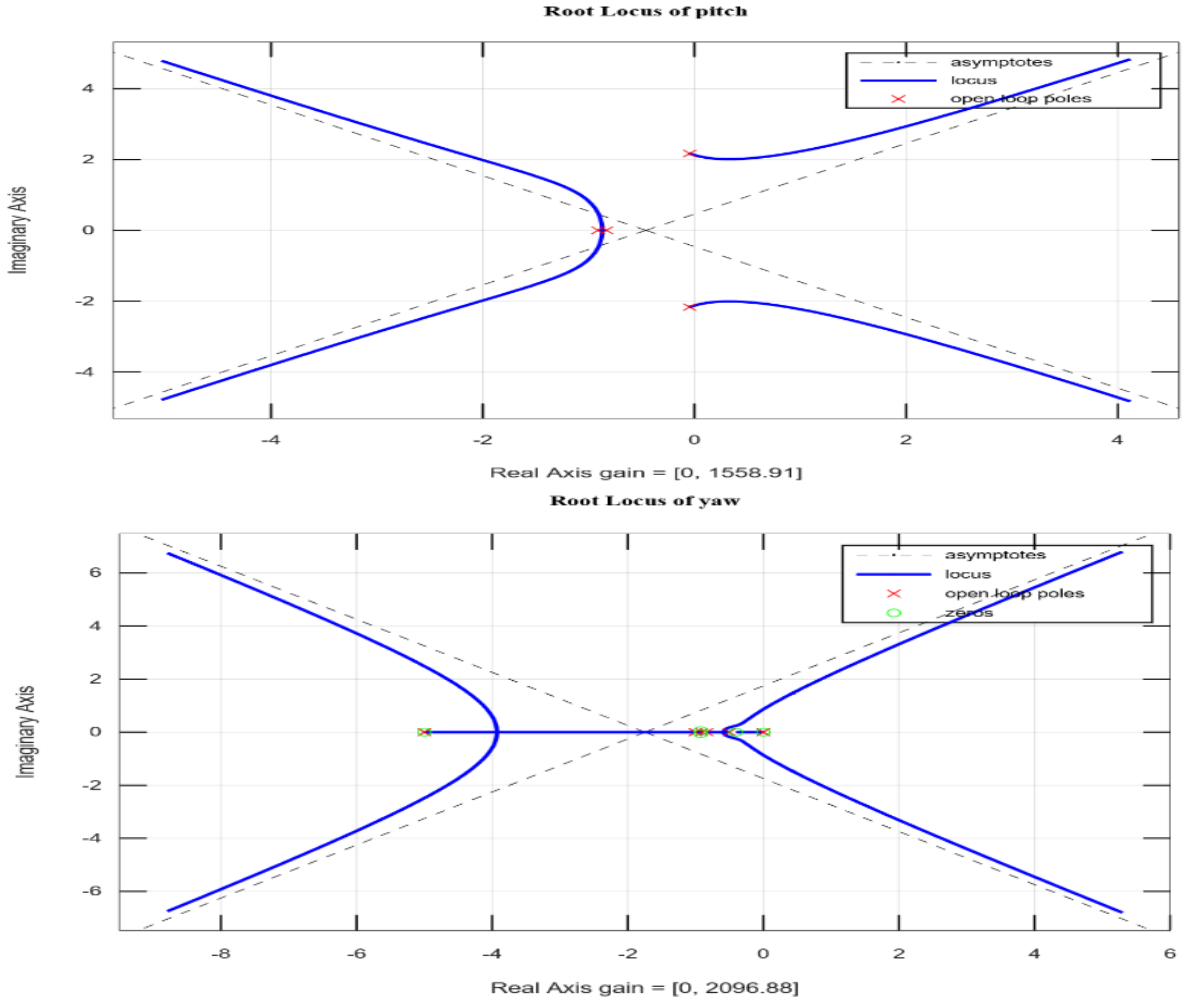


Fig. 2.2 Root Locus for pitch and yaw rotor respectively

The yaw rotor's transfer functions are given as equation (16). In the root locus plot in Fig. 2.2, the two poles of the yaw's transfer function are in RHS and cross the imaginary axis, showing that the system is unstable as the system gain increases.

$$T(s)_y = \frac{4.362s^4+31.66s^3+56.24s^2+36.61s+7.569}{s^8+14.24s^7+74.53s^6+183s^5+239s^4+171.2s^3+63.42s^2+9.465s} \quad (16)$$

3. DESIGNING AND OPTIMIZATION OF CONTROL SCHEMES

The control problem and desired target for testing TID and I-TD controllers are optimal and precise trajectory tracking as well as fast stabilization control of yaw and pitch actuators. The error e(t) between the setpoint and experimental trajectory is fed into both controllers, and the output is c(t), which is a regulated voltage signal for the aerial system's yaw and pitch actuators.

3.1 Tilted Integral Derivative (TID) Controller Designing

The TID controller's control scheme can be thought of as an upgraded fractional variant of the traditional IO-PID controller's control scheme, which has three tuneable gains tilted, integral, and derivative, which are denoted by the letters k_t, k_i, and k_d, respectively. Despite the standard PID controller, a fractional-order s-domain transfer function $s^{1/n}$ replaces the proportional control block, where n is a positive integer, the desired performance can be achieved by selecting the value of n. In the present paper, the value of n has been empirically determined to be 3.

@2017, IJTRS All Right Reserved, www.ijtrs.com

Peak overshoot and settling time have been set at 4% and 8 seconds, respectively, to ensure smooth hovering and take-off of the aerial system in real-time. The TID controller's general output control equation and s-domain Laplace transfer function are given as equations (17) and (18).

$$c(t) = k_t D^{\frac{-1}{n}} e(t) + k_i D^{-1} e(t) + k_d D^1 e(t) \quad (17)$$

$$U(s) = \frac{c(s)}{e(s)} = k_t s^{\frac{-1}{n}} + k_i s^{-1} + k_d s^1 \quad (18)$$

The schematic block diagram of the TID controller applied to TRMS with unity feedback system is shown in Fig.3.1.

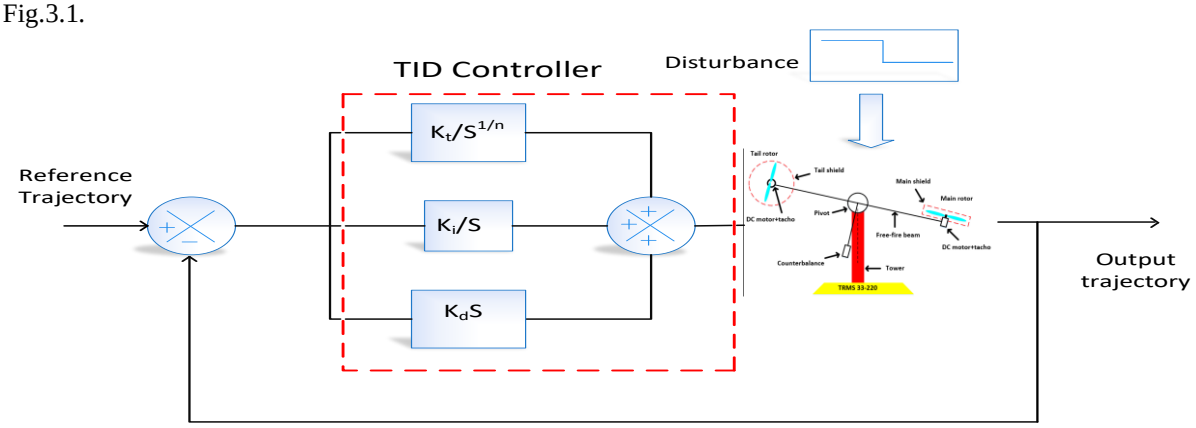


Fig. 3.1 TID controller applied to TRMS

The general characteristics equation is given as equation (19) by considering the aerial system with a TID controller applied with a unity feedback system.

$$1 + T(s)U(s) = 0 \quad (19)$$

The measured real and imaginary parts for pitch and yaw were given in equations (20-21) and (22-23), respectively, after substituting s_{12}=-0.5±0.518i and n=3 in equations (19).

$$Real_{TIDp} = 1 + 1.93k_{tp} - 2.35k_{ip} - 1.59k_{dp} \quad (20)$$

$$Img_{TIDp} = -1.19k_{tp} - 4.21k_{ip} + 6.58k_{dp} \quad (21)$$

$$Real_{TIDy} = 1 + 9.25k_{ty} - 3.58k_{iy} - 5.57k_{dy} \quad (22)$$

$$Img_{TIDy} = 1.69k_{ty} - 1.24k_{iy} - 4.27k_{dy} \quad (23)$$

The final objective function f_{TIDp} and f_{TIDy} for pitch and yaw of the TID, the controller is obtained by substituting the values of $Real_{TIDp}$, Img_{TIDp}, $Real_{TIDy}$, Img_{TIDy} and used to calculate the optimal operational gains for pitch and for yaw actuators, when tuned by grey wolf optimization.

$$f_{TIDp} = |Real_{TID}| + |Img_{TID}| + \tan^{-1}\left|\frac{Real_{TID}}{Img_{TID}}\right| \quad (24)$$

$$f_{TIDp} = \left|1 + 1.93k_{tp} - 2.35k_{ip} - 1.59k_{dp}\right| + \left|-1.19k_{tp} - 4.21k_{ip} + 6.58k_{dp}\right| + \tan^{-1}\left|\frac{1+1.93k_{tp}-2.35k_{ip}-1.59k_{dp}}{-1.19k_{tp}-4.21k_{ip}+6.58k_{dp}}\right| \quad (25)$$

$$f_{TIDy} = \left|Real_{TIDy}\right| + \left|Img_{TIDy}\right| + \tan^{-1}\left|\frac{Real_{TIDy}}{Img_{TIDy}}\right| \quad (26)$$

$$f_{TIDy} = \left|1 + 9.25k_{ty} - 3.58k_{iy} - 5.57k_{dy}\right| + \left|1.69k_{ty} - 1.24k_{iy} - 4.27k_{dy}\right| + \tan^{-1}\left|\frac{1+9.25k_{ty}-3.58k_{iy}-5.57k_{dy}}{1.69k_{ty}-1.24k_{iy}-4.27k_{dy}}\right| \quad (27)$$

3.2 Integral -Tilted Derivative (I-TD) controller Designing Formulas

The I-TD controller for TRMS is designed similarly to the TID controller, with the exception that the tilted and derivative blocks are held in feedback to the reference trajectory, and the integral block's voltage control signal is directly given to the aerial device, as shown in Fig. 3.2.

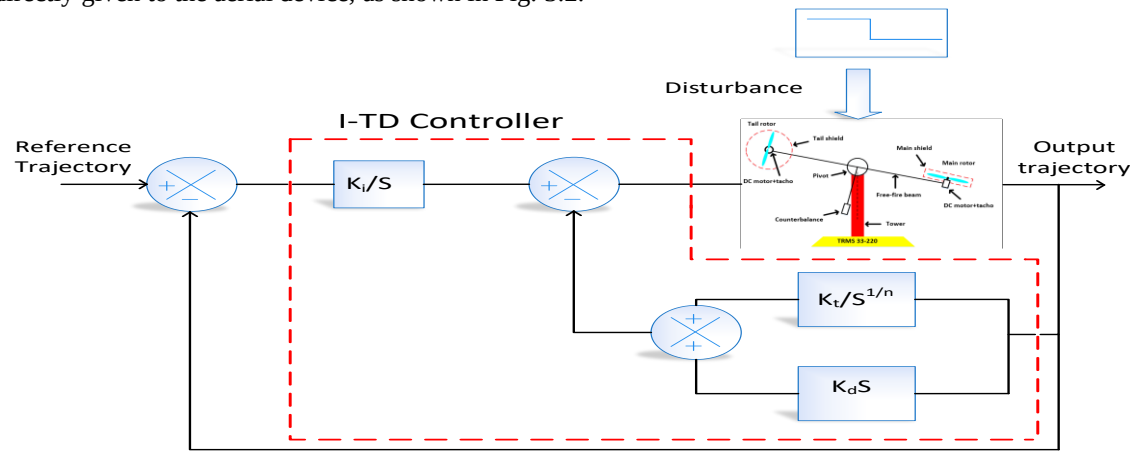


Fig. 3.2 I-TD controller applied to TRMS

@2017, IJTRS All Right Reserved, www.ijtrs.com

Also, the controller output in time and the Laplace domain are given as equations (28) and (29).

$$c(t) = k_i D^{-1} e(t) - (k_t D^{\frac{-1}{n}} e(t) + k_d D^1 e(t)) \quad (28)$$

$$U(s) = \frac{c(s)}{e(s)} = k_i s^{-1} - (k_t s^{\frac{-1}{n}} + k_d s^1) \quad (29)$$

The measured real and imaginary parts for pitch and yaw for I-TD controllers are given in equations (30-31) and (32-33), respectively, by substituting the value of the dominant pole s_{12}=-0.5±0.518i and n=3 in the characteristic equation.

$$Real_{I-TDp} = 1 - 1.93k_{tp} - 2.35k_{ip} + 1.59k_{dp} \quad (30)$$

$$Img_{I-TDp} = 1.19k_{tp} - 4.21k_{ip} - 6.58k_{dp} \quad (31)$$

$$Real_{I-TDy} = 1 - 9.25k_{ty} + 3.58k_{iy} + 5.57k_{dy} \quad (32)$$

$$Img_{I-TDy} = -1.69k_{ty} - 1.24k_{iy} + 4.27k_{dy} \quad (33)$$

The final objective function f_{I-TDp} and f_{I-TDy} for pitch and yaw of the I-TD controllers are obtained by substituting the values of $Real_{I-TDp}$, Img_{I-TDp}, $Real_{I-TDy}$, Img_{I-TDy} and used to calculate the optimal operational gains for pitch and for yaw actuators, when tuned by grey wolf optimization.

$$f_{I-TDp} = |Real_{I-TD}| + |Img_{I-TD}| + \tan^{-1}\left|\frac{Real_{I-TD}}{Img_{I-TD}}\right| \quad (34)$$

$$f_{I-TDp} = \left|1 - 1.93k_{tp} - 2.35k_{ip} + 1.59k_{dp}\right| + \left|1.19k_{tp} - 4.21k_{ip} - 6.58k_{dp}\right| + \tan^{-1}\left|\frac{1-1.93k_{tp}-2.35k_{ip}+1.59k_{dp}}{1.19k_{tp}-4.21k_{ip}-6.58k_{dp}}\right| \quad (35)$$

$$f_{I-TDy} = \left|Real_{I-TDy}\right| + \left|Img_{I-TDy}\right| + \tan^{-1}\left|\frac{Real_{I-TDy}}{Img_{I-TDy}}\right| \quad (36)$$

$$f_{I-TDy} = \left|1 - 9.25k_{ty} + 3.58k_{iy} + 5.57k_{dy}\right| + \left|-1.69k_{ty} - 1.24k_{iy} + 4.27k_{dy}\right| + \tan^{-1}\left|\frac{1-9.25k_{ty}+3.58k_{iy}+5.57k_{dy}}{-1.69k_{ty}-1.24k_{iy}+4.27k_{dy}}\right| \quad (37)$$

3.3 Controllers Tuning by Grey Wolf Optimization (GWO)

Recently, nature-inspired computational algorithms have gained a lot of traction in control engineering because they are based on predatory activities such as searching, hunting, pollinating, and so on. One of these nature-inspired algorithms is Grey Wolf Optimization. It considers hunting and the social hierarchy of the grey wolf. It is used to optimize the controller parameters and minimizing the cost function. As shown in Fig. 3.3, grey wolves are treated as top prey predators and are categorised as alpha, beta, and omega based. The alpha wolf is considered the squad leader and has the authority to make major decisions such as the solution finding area and hunting. Beta wolves help and obey the alpha wolf's hunting orders, and they control the omega wolf, which is the lowest in the hierarchy.

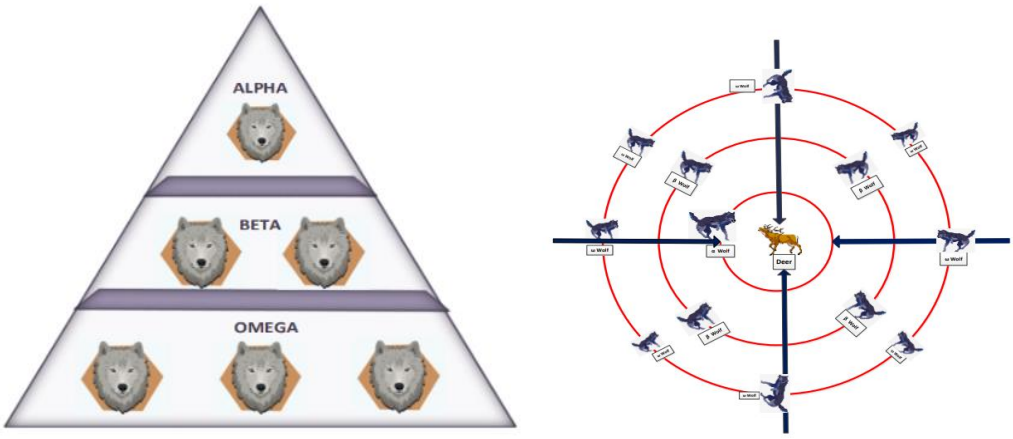


Fig. 3.3 Triangular hierarchy of alpha, beta, and omega wolves and hunting map of grey wolves

Equations are used to model the statistical modelling of encircling the prey (38-47).

$$Dist = |B.Y_s(i) - Y(i)| \quad (38)$$

$$Y(i+1) = Y_s(i) - A.Dist \quad (39)$$

$$A = 2ax_1 - a \quad (40)$$

$$B = 2.x_2 \quad (41)$$

Here A, B, and Y_s, Y denote vector coefficients and the position of hunting set points as well as the location of the wolf. x_1, x_2 are randomized numbers ranging from 0 to 1. Equations (42) to (47) show how to recognise alpha, beta, and omega wolves.

$$Dist_\alpha = |B_1.Y_\alpha(i) - Y(i)| \quad (42)$$

$$Y_{11} = Y_\alpha(i) - A_1.Dist_\alpha \quad (43)$$

$$Dist_\beta = |B_2.Y_\beta(i) - Y(i)| \quad (44)$$

$$Y_{12} = Y_\beta(i) - A_2.Dist_\beta \quad (45)$$

$$Dist_\omega = |B_3.Y_\omega(i) - Y(i)| \quad (46)$$

@2017, IJTRS All Right Reserved, www.ijtrs.com

$$Y_{13} = Y_{\omega}(i) - A_3.Dist_{\omega} \quad (47)$$

The modelled minimising cost function of TRMS implanted through GWO via equation (18) for TID Controller for pitch and yaw actuators is reflected in equations (48) and (49), respectively.

$$Dist_{\alpha/\omega/\beta} = \min|f_{TIDp}| \quad (48)$$

$$Dist_{\alpha/\omega/\beta} = \min|f_{TIDy}| \quad (49)$$

The minimising cost function for the I-TD controller is given as equations (50) and (51) for pitch and yaw controllers, respectively.

$$Dist_{\alpha/\omega/\beta} = \min|f_{I-TDp}| \quad (50)$$

$$Dist_{\alpha/\omega/\beta} = \min|f_{I-TDy}| \quad (51)$$

From equations (43), (45), and (47), the best hunter wolf giving optimal solution is given as equation (52).

$$Y(i+1) = (Y_{11} + Y_{12} + Y_{13})/3 \quad (52)$$

Table-3.1 Pseudocode of GWO

Step Performed	Algorithm
Initialization of Optimization	Dimension of the particular complication Limitation of the particular complication Size of Population Optimizing Variable Limit condition (max – loops compiled or required permit error)
START	Regions of all of α, β, ω wolves
FINDING	While not stop condition, Obtain the updated cost-minimizing function Upgrade the wolf regions Restrict the area of the coordinates Update α, β, ω and also Upgrade the limiting condition Stop

After optimization by GWO, the optimized values of operational gains for TID and conventional I-TD controllers are provided in Table-3.2 and Table-3,3, respectively.

Table 3.2 I-TD Controller values

ROTOR	Control Variables	GWO Tuned values
PITCH	k_{tp}	4.39
	k_{ip}	6.32
	k_{dp}	18.87
YAW	k_{ty}	14.98
	k_{iy}	12.33
	k_{dy}	25.23

Table-3.3. TID Controller values

ROTOR	Control Variables	GWO Tuned values
PITCH	k_{tp}	5.45
	k_{ip}	8.23
	k_{dp}	17.95
YAW	k_{ty}	16.82
	k_{iy}	17.42
	k_{dy}	29.52

4. RESULTS AND VALIDATION OF CONTROLLER

Both TID and I-TD grey wolf tuned controllers have been experimentally tested and validated in real-time with optimally tuned operational gains provided in Table-3.1 and Table-3.2. Fig.4.1 and Fig.4.2 show the main and tail rotor response for the I-TD controller over the desired pitch angle considered as a reference trajectory.

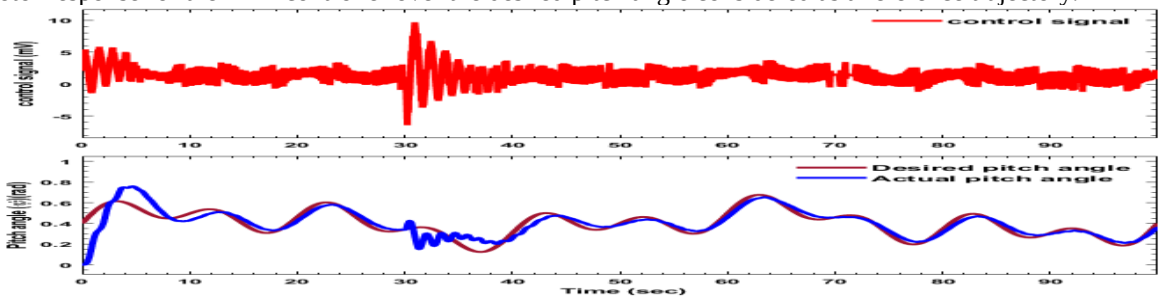


Fig. 4.1 I-TD controller pitch response

@2017, IJTRS All Right Reserved, www.ijtrs.com

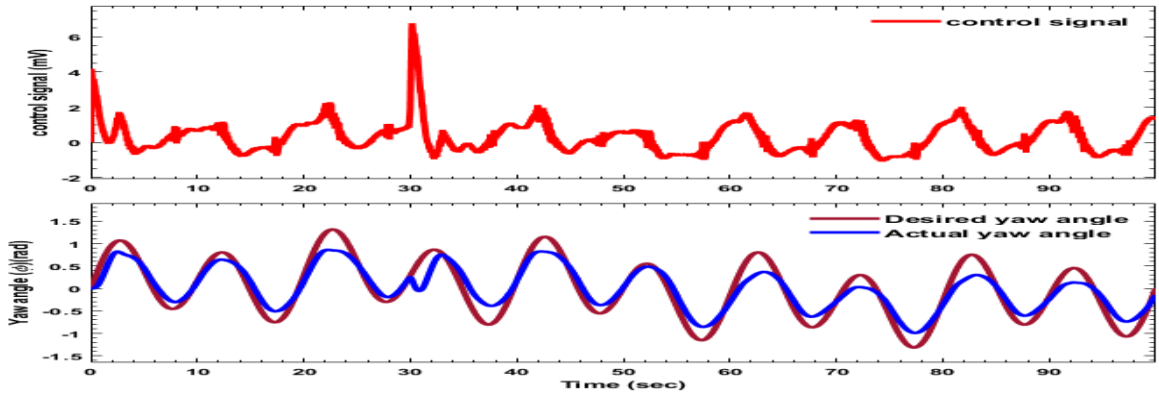


Fig. 4.2 I-TD controller yaw response

Fig. 4.3 and Fig. 4.4 show the main and tail rotor response for the TID controller, respectively. Also, the TID controller's control signals are displayed to observe controller performance in the time domain.

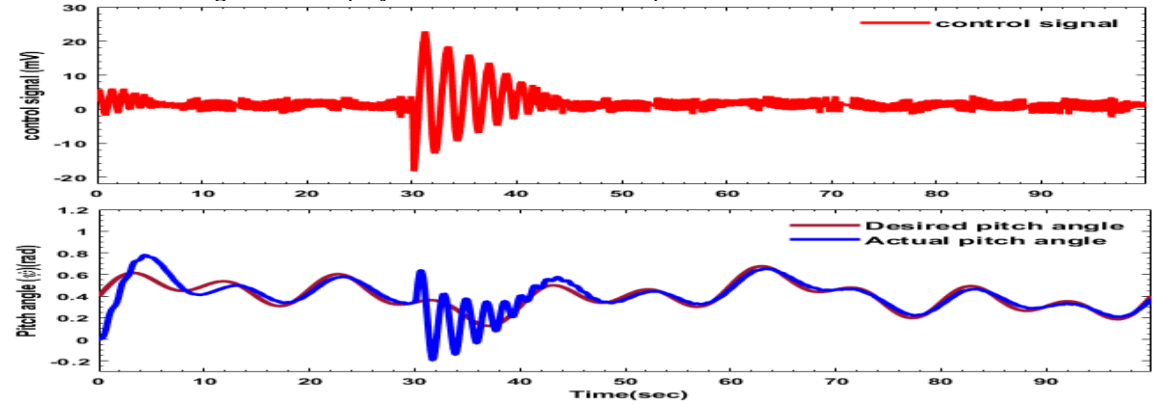


Fig. 4.3 TID controller pitch response

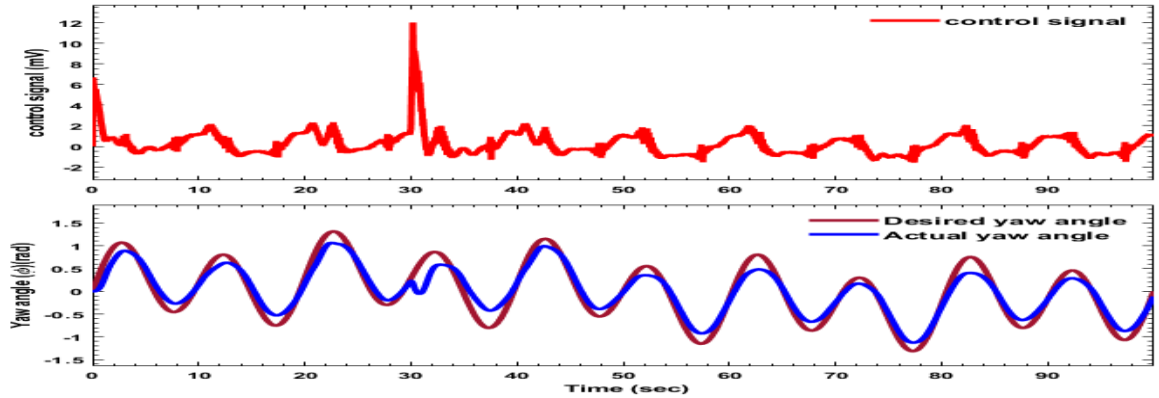


Fig. 4.4 TID controller yaw response

The I-TD controller outperforms the TID controller in terms of trajectory tracking for given reference waves, according to the controller's responses. Tables 4.1 and 4.2 show the responses of engineered controllers in terms of time to stabilise after a 30th-second manual disturbance. These given manual disturbances are of a random magnitude and resemble real-time atmospheric disturbances on the unmanned aerial system.

Table-4.2 TID Controller Values

Rotor	Disturbance at (Sec)	Stabilization at (Sec)	Settling time (Sec)
Pitch	30^{th}	44^{th}	14
Yaw	30^{th}	33^{th}	3

Table-4.3 I-TD Controller values

Rotor	Disturbance at (Sec)	Stabilization at (Sec)	Settling time (Sec)
Pitch	30^{th}	40^{th}	10
Yaw	30^{th}	32^{th}	2

Table-4.4 Performance of TID & I-TD w.r.t PID

Rotor	TID (% faster than PID)	I-TD (% faster than PID)
Pitch	44	60
Yaw	25	50

@2017, IJTRS All Right Reserved, www.ijtrs.com

Table-4.4 shows the percentage of time that TID and I-TD controllers outperform the regular PID tuned with grey wolf optimization. The I-TD controller stabilises the system response faster than the TID controller for vertical motion by pitch rotor. The I-TD controller's increased stability is due to a continuous closed feedback loop of Tilted and Derivative blocks, which increases the control signal.

CONCLUSION

TID and I-TD controllers are compared in this paper for the first time. The benchmarked aerial dynamical system's stabilization is controlled by two TID and I-TD controllers for pitch/main and yaw/tail actuators. A bio-inspired grey wolf optimization algorithm is used to improve the efficiency of both the designed controllers. Both controllers were given manual disruptions every 30 seconds, and the time it took them to stabilize and return to the reference trajectory was measured experimentally. The mathematical comparison and trajectory monitoring graphs show that the I-TD controller takes less time to stabilize after a disturbance and performs better in terms of trajectory stabilization and tracking control.

REFERENCES

[1] Feedback Instruments, “Twin Rotor Mimo System,” 2016.
[2] P. China, D. Dutta, and M. R. Kumar 2020 Modelling of Twin Rotor MIMO System using Differential Evolution with Success-based Parameter Adaptation Proc. 1st Int. Conf. on Power, Control and Computing Technologies (Raipur) pp 223-28.
[3] S. K. Valluru, R. Kumar and R. Kumar 2019 Design and Implementation of L-PID and IO-PID Controllers for Twin Rotor MIMO System Proc. Int. Conf. on Power Electronics, Control, and Automation (New Delhi)
[4] S. K. Valluru, R. Kumar and Ayush 2019 Trajectory Tracking Control of TRMS using FO-PID and FOI-PD Controllers: An Experiment Proc. 3rd Int. Conf. on Recent Developments in Control, Automation & Power Engineering (Noida) pp 319-22
[5] S. K. Valluru, R. Kumar and R. Kumar 2020 Design and Real Time implementation of fmincon, MOGA tuned IO-PID and FO-PIλDμ Controllers for Stabilization of TRMS Procedia Computer Science vol 171 pp 1241-50
[6] S. K. Valluru and M. Singh 2017 Investigation of NARMA L-2 and Artificial Bee Colony Tuned PID Controllers for Bench Scaled Nonlinear Dynamical System Int. J. Control Theory Appl. vol 10 no. 6 pp 363–74
[7] H. Prabha and R. Kumar 2020 Real Time Experimental control and Design of FOPIλ and L-PID Controllers tuned by Invasive weed optimization for Trajectory Control of TRAS Proc. Int. Conf. on Power Electronics & IoT Applications in Renewable Energy and its Control (Mathura) pp 148-52
[8] S.Valluru, M. Singh, Ayush, and A. Dharavath 2019 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 vol 989, ed Choudhury S., Mishra R., Mishra R., Kumar A. (Singapore: Springer Nature) pp 599–608.
[9] P. U. Shaikh and S. Karvekar 2016 Implementation of Sliding Mode Controller for Twin Rotor Multi Input Multi Output System without state observer Proc. IEEE 1st Int. Conf. on Power Electronics, Intelligent Control and Energy Systems (Delhi) pp 1-6
[10] Ayush, H. Prabha and R. Kumar 2020 Robust H∞ Control Approach for Trajectory Tracking of Twin Rotor MIMO System Proc. Int. Conf. on Emerging Trends in Communication, Control and Computing pp 1-4
[11] Lurie, B.J. 1994 Three-parameter tunable tilt-integral-derivative (TID) controller U.S. Patent
[12] D. Sain, S. K. Swain and S. K. Mishra 2016 TID and I-TD controller design for magnetic levitation system using genetic algorithm Perspectives in Science vol 8 pp 370-73
[13] P. N. Topno and S. Chanana 2015 Tilt Integral Derivative control for two-area load frequency control problem Proc. 2nd International Conference on Recent Advances in Engineering & Computational Sciences (Chandigarh) pp 1-6
[14] Mirjalili S, Mirjalili SM, Lewis A 2014 Grey wolf optimizer Advances in Engg. Software 69:46–61.
[15] H. Prabha, Ayush, R. Kumar, A.L. Meena. 2021 Real-Time Stabilization Control of Helicopter Prototype by IO-IPD and L-PID Controllers Tuned Using Gray Wolf Optimization Method Computational Methods and Data Engineering. Advances in Intelligent Systems and Computing vol 1227, ed Singh V., Asari V., Kumar S., Patel R. (Singapore: Springer).
[16] Astrom.K.J. and Hagglund.T, “Benchmark Systems for PID Control,” in IFAC Proceedings on Digital Control: Past, Present and Future of PlD Control, 2000, vol. 32, no. 4, pp. 165–166.

@2017, IJTRS All Right Reserved, www.ijtrs.com

Corrected Proof

Check for updates

Water Science & Technology

IWA PUBLISHING

© 2021 The Authors

Water Science & Technology Vol 00 No 0, 1 doi: 10.2166/wst.2021.241

Heat and water flux modeling in an earth dam

Sushant Kumar*, A. K. Sahu and Munendra Kumar

Department of Civil Engineering, Delhi Technological University, Delhi, India
*Corresponding author. E-mail: sushant_phd2k17@dtu.ac.in

ABSTRACT

This study aims to identify the water flux in an earth dam using heat flux due to convection. Sixteen earth dam models were constructed in a hydraulic flume by varying geometrical and flow input parameters to identify heat and water flux. Homogeneous as well as earth dams with clay cores were built in a hydraulic flume. Temperature measurements were done to calculate heat flux in the experimental model. A finite element model of the earth dam using Seep/w was developed to obtain water flux, while temp/w was used to obtain heat flux. These results were used as input in Temp/w and Seep/w in Geostudio 2020. Significant reduction of the heat and water flux was seen while comparing the homogeneous models with central impervious core models. An increase in the heat and water flux was observed on increasing the downstream filter's length, longitudinal slope, and vice versa with the upstream slope and the thickness of the clay core. Comparing fluxes in a homogeneous dam model (model 1) with the clay core model (model 9) with top width 2.4 m and bottom width 18 m in model 9, both water flux and heat flux were reduced by 78.46%. While comparing it with model 10, with bottom core width of 18 m and top core width of 1.9 m, both water flux and heat flux reduced by 77.72%. Heat flux measurements were found to be a valuable alternative to detecting water flux and seepage in an earth dam at a reduced cost.

Key words: convection, earth dam, heat flux, porous medium, temperature, water flux

HIGHLIGHTS

- This study identifies the heat due to the convective heat transfer in the respective earth dam model and water flux due to seepage in a porous media such as earth dams by varying geometrical and flow input parameters.
- A finite element model using Seep/w was used to determine the water flux, and Temp/w was used to obtain heat flux.
- It was observed that near walls of a hydraulic flume, soil pores of the earth dam were of larger volume, resulting in an increased local water flow.
- Increased heat and water fluxes were observed when the downstream filter length and longitudinal slope were increased, and vice versa with the upstream slope and thickness of the clay core.
- Heat flux obtained was a practical alternative in detecting seepage and water flux in an earth dam.

1. INTRODUCTION

Earthen dams are simple and cost-effective structures made of soil based on silicon that stand on their self-weight to prevent sliding and overturning. Earth dams are prone to seepage, and inadequate seepage may cause stability problems as it leads to a reduction in the shear strength of the dam, which could lead to its failure.

Radzicki & Bonelli (2010) applied an Impulse Response Function Thermal Analysis model to detect leakage in an earth hydraulic structure – an earth dam and dikes. Chen *et al.* (2009) developed a Finite element method-based model of fully coupled multiphase flow, thermal, and stress/deformation in a porous geological media. A three-dimensional software name 'THYME3D' was also developed. Johansson & Sjödahl (2004) inspected the downstream toe of about ten embankment dams to detect seepage with data obtained by measuring temperature. Yousefi *et al.* (2013) studied leakage in earth-fill dams using seepage flow and thermal simulation. Convective diffusion equation and mass balance were discretized using the finite element analysis in saturated and unsaturated zones. Results showed that temperature variation was more effective for inspection of leakage than the piezometric seepage levels. Alekseevich & Sergeevich (2017) developed a thermal-seepage regime numerical model based on the finite element method for dam foundations operated in permafrost conditions. Several factors that affect the temperature measurement were the thermal conductivity of soil, microclimate,

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (http://creativecommons.org/licenses/by/4.0/).

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Corrected Proof

and the vegetation in the area. It was found that the amount of water rising can be estimated by computing the amount of heat energy transferred to the surface.

Abu-Hamdeh (2014) studied the effect of water content and bulk density on the specific heat and volumetric heat capacity in a laboratory on sand and loam. The calorimeter method was used to determine the specific heat of both types of soils. It was found that the specific heat increases with increasing moisture content of the soil. In contrast, the volumetric heat capacity increased with increasing the soil density and its moisture content. On similar moisture content and soil density, loamy soil had higher specific heat and volumetric heat capacity than the sandy soil. Amanifard *et al.* (2007) investigated the effect of an electrical double layer (EDL) near the solid/liquid interface on a three-dimensional heat transfer characteristic. Numerical investigation on the pressure drops of water flow through a rectangular microchannel was also carried out. It was observed that the liquid flow in rectangular microchannels was influenced significantly by the EDL, especially in the high electric potentials. Del Piero (2020) proposed a heat conduction theory in rigid heat conductors based on mechanical concepts and compared it with traditional thermodynamic theories. It was found that the proposed mechanical approach provided the same equation as the thermodynamic approach but in a simpler way. Badruddin *et al.* (2020) reviewed heat transfer in porous media in various geometrical shapes like a vertical plate, cylindrical shape, cavity, etc. Heat transfer in natural convection, mixed convection, thermal equilibrium, and thermal non-equilibrium was also explained for the review's purpose. Cuong *et al.* (2017) reviewed temperature measurement to estimate seepage in an embankment dam. The downstream toe was highlighted as the most important location to detect the seepage using fiber optic sensors. Short-term temperature analysis was good for estimating localized leakage, which helps minimize the cost of repair works. It was also reviewed that from the temperature measurements, seepage velocity can be evaluated.

Buntebarth (2020) determined the thermal conductivity and the thermal diffusivity of dry granular material by heat transfer between the source and the investigation media. An empirical relation was established between the thermal conductivity and the diffusivity of the media. Misra *et al.* (1995) presented a review on factors affecting soil thermal conductivity. They determined soil thermal conductivity by empirical and semi-empirical methods of five soil types: gravel, sand, silt, clay, and peat. A theoretical thermal conductivity model of granular soil material was developed in an almost dry phase. Abu-Hamdeh (2003) investigated the effect of water content and bulk density on the specific heat, volumetric heat capacity, and sandy and clay soil's thermal diffusivity. Laboratory experiments were performed using the calorimetric method. The results also observed that the clay soil generally had higher specific heat and volumetric heat capacity than sandy soil. Chuvilin & Bukhanov (2019) conducted an experimental study to determine the thermal conductivity of frozen soil at gas pressure below equilibrium using a KD-2 needle probe which caused a little impact on the soil samples in the study area.

Kurz *et al.* (2017) presented experimental laboratory results to estimate the thermal conductivity of frozen and unfrozen soil samples of clay, silt, and peat subjected to seasonal freezing and thawing at the study area. Thermal conductivity values obtained from empirical methods were compared with the value obtained from a thermal probe in the laboratory. Kodešová *et al.* (2013) measured thermal conductivity and heat capacity using a KD2 PRO device with 13 TR-1 and SH-1 sensors. The results measurement observed that the highest thermal conductivities were measured in soils on quartz and substrates.

2. SOIL PROCUREMENT AND ITS TESTING

The soil was procured from a depth of 2 m below the ground level of the ground of Delhi Technological University, Delhi, India (28.7501° N, 77.1177° E). The soil's geotechnical properties, used to construct the dam in hydraulic flume, were tested in the laboratory. The test results for geotechnical properties were used to develop the dam model, such as natural moisture content, sieve analysis, liquid limit, plastic limit, permeability, standard compaction test, consolidation test, and direct shear test. X-ray diffraction (XRD) test was carried out using Bragg-Brentano with a Bruker D8 advance machine to find out the crystalline materials in the soil used in the construction of the shell and core of the earth dam models in the hydraulics laboratory.

3. THERMAL PROPERTIES

3.1. Thermal conductivity

In the past, soil thermal properties were used to predict water, heat, and solute transport in soils (Abu-Hamdeh 2003). Among other essential thermal properties, soil thermal conductivity has a crucial role in temperature modeling in a porous medium. It depends on moisture content, mineral composition, temperature, and texture.

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Corrected Proof

However, for low-density clay soils, the critical moisture content coincides with the plastic limit. The thermal conductivity of coarse-textured, angular grained soils is higher than that of fine-textured soils (Misra *et al.* 1995). Various authors have reported the thermal conductivity of soil at different temperatures and mineral compositions. For silty sand having quartz content of 64%, water content of 15%, and dry density of 1.77 g/cm^3, reported thermal conductivity of 1.04 W/mK in the unfrozen state and 1.61 W/mK in the frozen state at −6 °C. (Chuvilin & Bukhanov 2019). The thermal conductivity of sand collected from two places had a value of 1.9, 1.25 at 20% water content. The thermal conductivity of saturated sand was 3.12 W/mK (Buntebarth 2020). The average water thermal conductivity value W/mK was 0.56 at 17 °C (Chuvilin & Bukhanov 2019). The thermal conductivity of water was reported to be 0.594 W/mK (Kosky *et al.* 2012).

3.2. Specific heat and volumetric heat capacity

'Specific heat is the amount of heat required to change the temperature of a mass unit of a substance by one degree. In other words, the specific heat capacity of a substance is the heat capacity of a sample of the substance divided by the sample's mass. It is used to calculate the energy change associated with a temperature change.' Many factors influence soil heat capacity, among which water content and soil density are the key factors that influence it. In sandy soil, specific heat is associated with water content for known bulk density. Water in the liquid state has one of the highest specific heats, comparing with common substances. The specific heat of soil is determined using a calorimeter by measuring its heat capacity and dividing its mass (Kosky *et al.* 2012). The sum of volumetric heat capacities of each soil component multiplied by their fraction gives soil volumetric heat capacity. 'It stands for stored internal energy ability of a given soil volume while undergoing a given temperature change.' The analysis of coupled heat-water transfer in soil plays a vital role in the temperature and heat transfer within an earth dam. Various literature has mentioned specific heat and volumetric heat capacity, among which few are presented. Dry sand had specific heat of 932–958 J/kgK measured using a calorimeter (Ižvolt & Dobeš 2014). While specific heat of the sand, quartz, is 830 J/Kg.°C, dry soil is 800 J/Kg.°C and wet soil is 1,480 J/Kg °C, respectively. (Engineering toolbox 2003, Specific heat of some common substances). Kodešová *et al.* (2013) reported a soil's specific heat of 0.73 kJ/KgK, and a volumetric heat capacity of 1.9 MJ/m^3/K, at the dry density of 1.34 g/cm^3. Specific heat of the water was 4.18 kJ/kgK and volumetric heat capacity was 4.18 MJ/m^3/K. For clay of dry density 1.46 g/cm^3, specific heat was 0.86 kJ/kgK and volumetric heat capacity was 2.30 MJ/m^3/K. While specific heat of water was 4,190 kJ/kgK. And at 30 °C, the specific heat of the water was recorded as 4.1175 kJ/kgK or 74.181 J/(mol k) (Engineering toolbox 2004, water - specific heat).

4. HEAT TRANSFER IN SOIL

'Heat transfer refers to thermal energy that passes between objects by either thermal conductivity, thermal convection, in which a fluid moves between regions of different temperature, or thermal radiation, in which energy is transmitted by electromagnetic radiation.' Temperature difference induces a heat flux. This induced heat flux flows from the hotter medium to the colder; it requires a difference in the temperature and a medium through which heat flows. Heat flux also depends on the thermal transfer coefficient of soil used in constructing the earth dam and its foundation. In an earth dam, heat can flow through soil-to-soil particles (conduction) and fluids' motion in the pores (convection). The air temperature surrounding the earth dam and water in the reservoir upstream of the earth dam serves as the dam's principal thermal loading. Other sources that may affect the earth dam's temperature variation are geothermal, a frozen process in colder regions, humidity changes around an earth dam, and wind influence. Tokoro *et al.* (2016) proposed a thermal conductivity model and empirical equations for soil. The thermal conductivity of soil and its moisture content had a non-linear relationship. Ižvolt & Dobeš (2014) monitored the impact on the measured value of three different building materials' specific heat and thermal conductivity due to the calorimetry test procedure's modification. Alim *et al.* (2017) used Seep/w to calculate seepage discharge for a homogeneous earth dam resting on an impervious base without any filter or clay core. They proposed an equation for seepage through a homogeneous earth dam resting on an impervious base.

4.1. Heat conduction

'It is defined as the amount of heat passing in unit time through a unit cross-sectional area of the soil under a unit temperature gradient.' Viscosity and water density are temperature-dependent phenomena in the earth dam,

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Corrected Proof

which affect the heat conduction in the soil particles present in an earth dam's body. Heat conduction can occur due to soil's material property present in the earth dam body and foundation. Heat conduction is affected by the saturated/unsaturated region in the earth dam, as below the phreatic line, natural or free convection also occurs. Above the phreatic line, conduction predominates; that is, in the upper zone of the dam, there is pure heat conduction and no heat advection as negligible water flow is present above the phreatic line; hence, slow heat transport from the dam surface inward to the dam occurs. While in the central zone, both convection in x-direction and conduction in y-direction takes place.

4.2. Convection in soil

Convective heat transfer occurs in an earth dam whenever there is a difference in temperature in the earth dam body and the water upstream. Natural convection is produced by motion caused due to density difference in a fluid due to temperature difference, called convection currents. It continues as long as temperature difference exists in an earth dam and upstream water. Temperature acts as a natural tracer to find the seepage flow and water flux to localize leakage in embankment dams. In an earth dam, heat flux was used to detect leakage regions and determine seepage velocity using temperature observation. When seepage water goes through an earth dam's pores, it exchanges heat with the adjacent soil medium (Cuong *et al.* 2017). Heat flux varies in an earth dam due to convective heat transfer, due to the energy transported by seepage flow. Increased fluid speed enhances high heat transfer in convection, and with high heat transfer, a high-pressure drop can be observed.

5. METHODOLOGY

The physical model was scaled down so that it was possible to perform experiments in the laboratory using a hydraulic flume. The length scale factor for the experimental model was 1/100. This scale factor was applied to the length, width, and height of the dam, and the discharge obtained. The effect of the longitudinal slope was studied by keeping these dimensions: upstream slope 38°, downstream slope 45°, bottom width 52 m; the height of the dam was kept at 20 m, while the longitudinal slope was varied at 0.00%, 6.07%, 2.28%, 4.00% slope for the models 1,6,7,8 respectively. The effect of the central impervious core (clay) on water and heat flux was studied in models 9 and 10 by comparing with the homogeneous model (model 1). The dimensions of models 9 and 10 were similar to model 1 but with an impervious core of bottom width 18 m each and top width of 2.4 m and 1.9 m, respectively. Table 1 shows the different dimensions of the homogeneous earth dam models with their dimensions and upstream head. Table 2 shows an earth dam's models with different horizontal filter lengths and a filter thickness of 3 m.

5.1. Experimental setup

An earthen dam of known dimensions was scaled down for experimental work; 16 earth dam models were constructed in a hydraulic flume of dimensions 300 cm length, 30.5 cm width, and 40 cm height. A Proctor penetrometer was used to control the compaction in layers of the earth dam during construction. The effect of scaling was considered as it affects the outcome of the result. A homogenous as well as an earth dam with central clay core was constructed. The temperature variation inside the dam was measured to calculate the heat flux across the earth dam in each model due to the convective water flow. Hydraulic flume was titled to change

Table 1 | Earth dam models' dimensions and upstream head of homogeneous section

Model number	Upstream water head (m)	Upstream slope (degrees)	Downstream slope (degrees)	Bottom width of the dam (m)	Height of dam (m)
1	13.5	38	45	52	20
2	13.5	38	59	52	20
3	13.5	50	34	52	20
4	13.5	50	34	56	20
5	13.5	50	34	56	15
11	13	38	45	52	20
12	13.5	45	45	52	20
13	13.5	50	45	52	20

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Corrected Proof

Table 2 | Earth dam models with different filter lengths

Model number	Upstream slope (degree)	Downstream slope (degree)	Bottom width of dam (m)	Height of dam (m)	Length of the filter from the downstream end (m)	% length ratio of filter to the total length (m)
1	38	45	52	20	0.0	00.00
14	38	45	52	20	15.5	29.80
15	38	45	52	20	23.0	44.23
16	38	45	52	20	13.0	25.00

the longitudinal slope to study its effect on water flux and heat flux. Fluid viscosity was adjusted by adding sugar and then measuring the rotational rheometer, a viscosity measurement instrument, in the physics laboratory to study its effect on water flux and heat flux.

Proper compaction in an earth dam leads to increased stiffness and strength, minimizing the settlements and preventing the dam's sliding failure. A Proctor penetrometer was used to achieve proper compaction in the construction of the 16 models of the earth dam. The soil was placed in five layers of 4 cm each at optimum moisture content with 90–95% variance using a Proctor penetrometer while keeping each layer's penetration rate at 1.3 cm/s. An average of three penetration resistance were compared with the calibration chart of moisture-penetration resistance and moisture-density curve relationship. A longitudinal filter was provided, and its length was varied for studying its effect on water flux and heat flux. Models 14–16 were constructed with a downstream filter with varying lengths, as mentioned in Table 2. In the present study, the water flux was obtained from the seepage discharge calculation; while heat flux in models was calculated using temperature readings obtained by a digital thermometer in a hydraulic flume.

5.2. Numerical modeling

Geostudio 2020 software is one of the robust programs based on the finite element technique and can simulate and analyze water flux using Seep/w and heat flux using Temp/w. A finite element mesh of size 0.1 m was chosen after several trial computations by changing the mesh size. The mesh size was determined so that it doesn't affect the computation time to a large extent and yields a good result.

6. RESULTS AND DISCUSSION

- The basic properties of soil were tested, and their results are shown in Table 3.

Figure 1 showing results of intensity vs. angle (2Θ) for soil used in the construction of homogeneous earth dam models as well as in the shell of the earth dam in non-homogeneous earth dam models, while Figure 2 shows XRD results of intensity vs. angle (2Θ) for soil used in the core of non-homogenous earth dam. Silicon oxide

Table 3 | Basic properties of soil

S.No.	Name of the test		Result
1	Natural moisture content		1.49%
2	Specific gravity		2.57
3	Grain size distribution	Soil classification	Silty sand
		D_{60}	1.61
		D_{30}	0.36
		D_{10}	0.09
		C_u	17.19
		C_c	0.87
4	Atterberg's limit	Liquid limit	26.24%
		Plastic limit	12.69%
5	Compaction:	Dry unit weight of the soil	18.05 kN/m^3
		Optimum moisture content	13.10%
6	Permeability of soil (shell)		6.78×10^{-6} m/s

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Corrected Proof

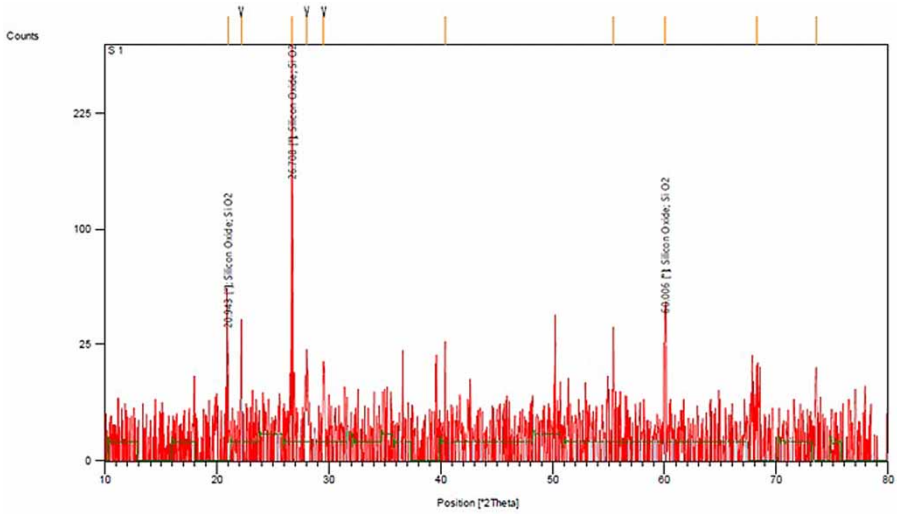


Figure 1 | XRD of soil used in the shell of earth dam models.

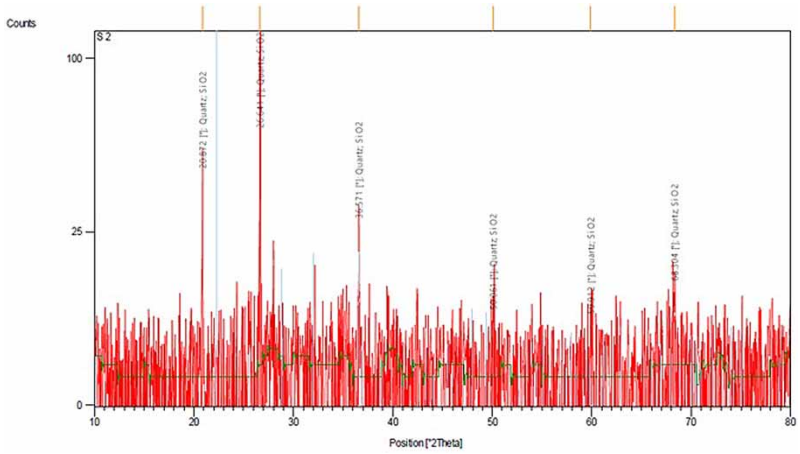


Figure 2 | XRD of soil used in the core of the earth dam models.

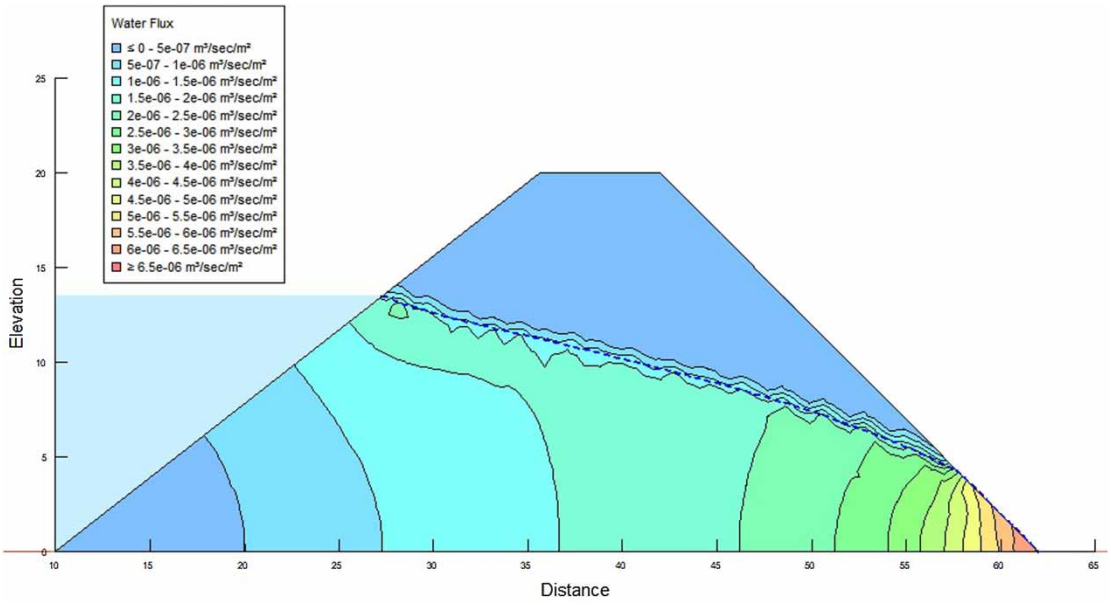


Figure 3 | Water flux in model 1.

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Corrected Proof

(SiO_2) was the primary chemical compound in the soil (shell) as the angle (2Θ) was at 26.715 with an intensity of 100% and having the empirical formula of O_2Si. The h, k, and l were 0,1, and 1 respectively, and spacing between planes (d) equals 3.33420 A°. In contrast, quartz was the primary chemical compound found in the soil (core), having the chemical formula SiO_2 and empirical formula O_2Si. The lattice's crystal system was hexagonal with an angle (2Θ) of 26.64 at an intensity of 100%. h, k, l, of the lattice was 0,1,1 respectively, and spacing between planes (d) equals 3.34353 A°.

6.1. Contours of water flux and heat flux

The contours of water flux in $m^3/sec/m^2$ are shown in Figures 3–18 for models 1 to 16, while heat flux is shown in Figures 19–34. The variation of water flux and heat flux for the location is shown. The earth dam model's phreatic

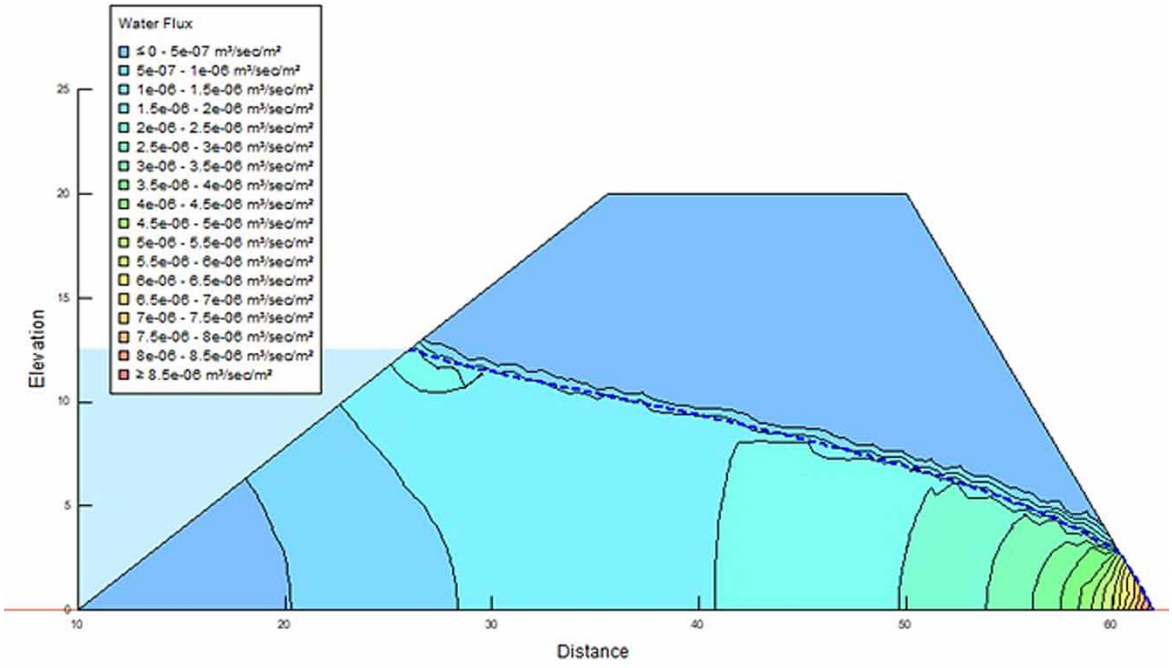


Figure 4 | Water flux in model 2.

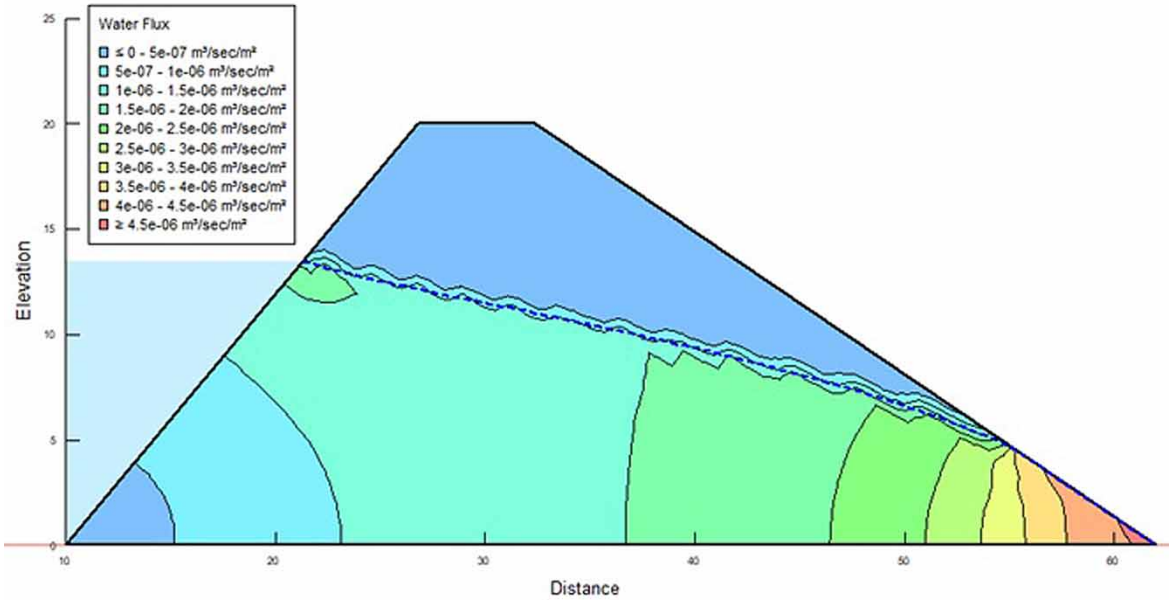


Figure 5 | Water flux in model 3.

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Corrected Proof

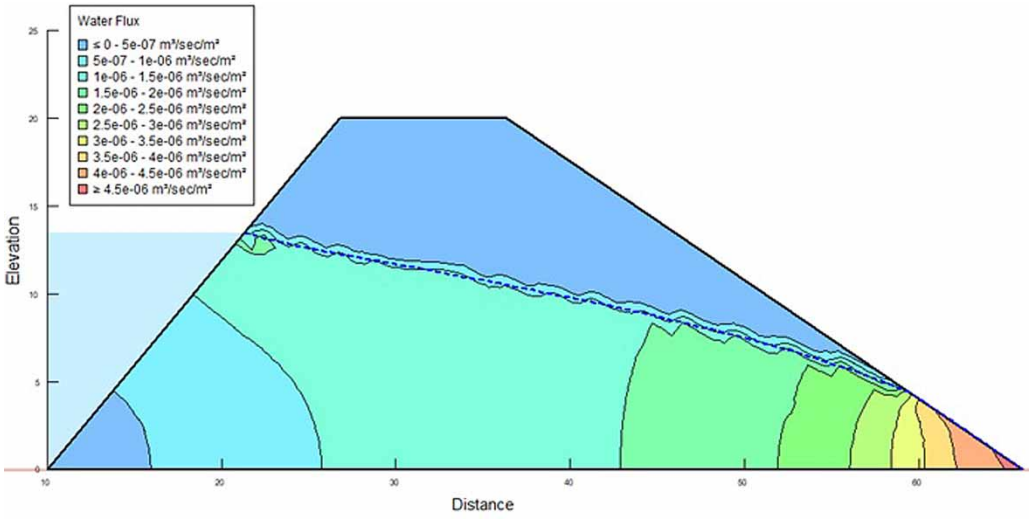


Figure 6 | Water flux in model 4.

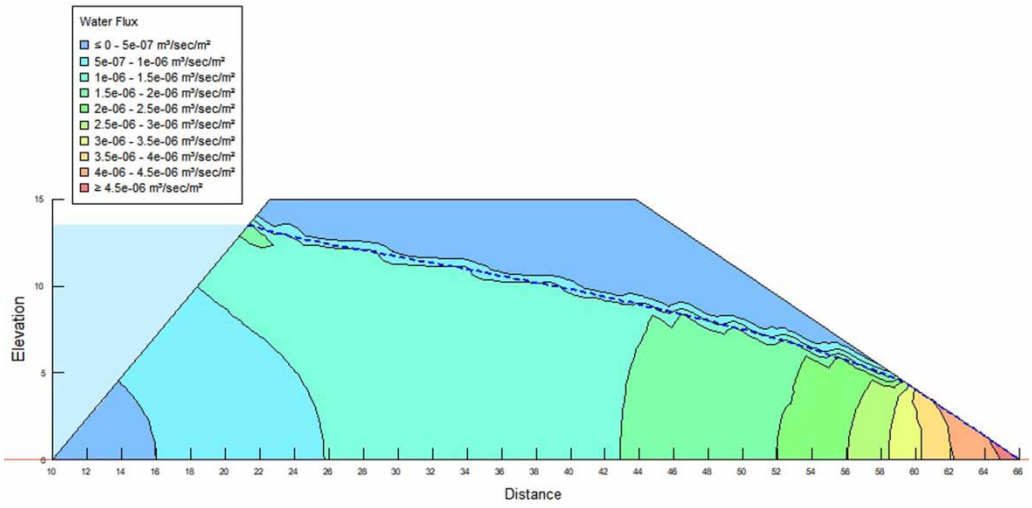


Figure 7 | Water flux in model 5.

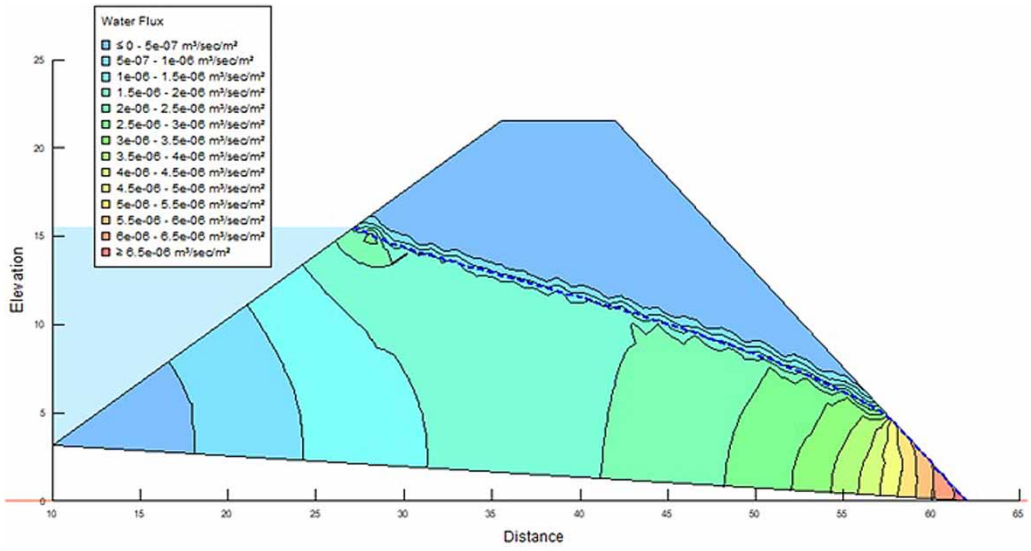


Figure 8 | Water flux in model 6.

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Corrected Proof

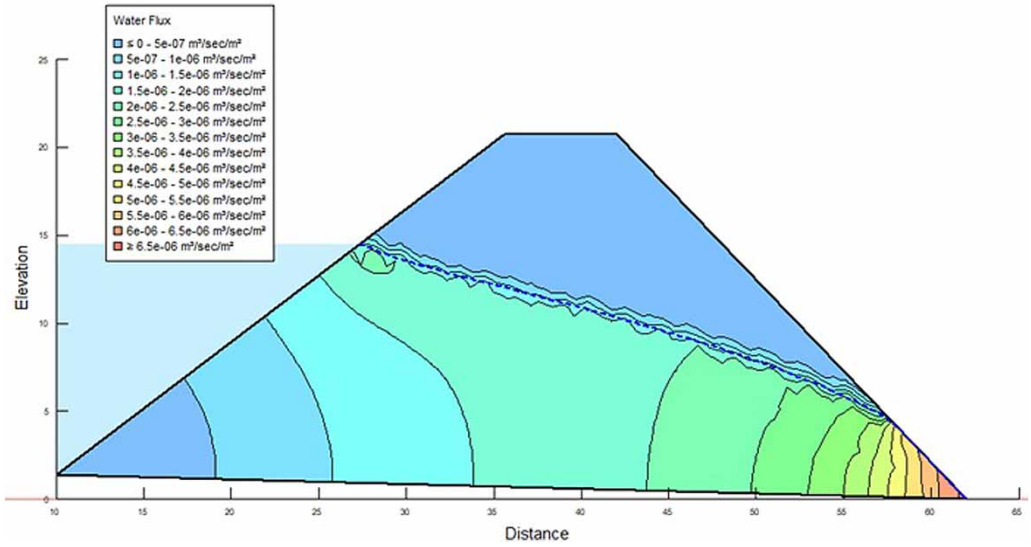


Figure 9 | Water flux in model 7.

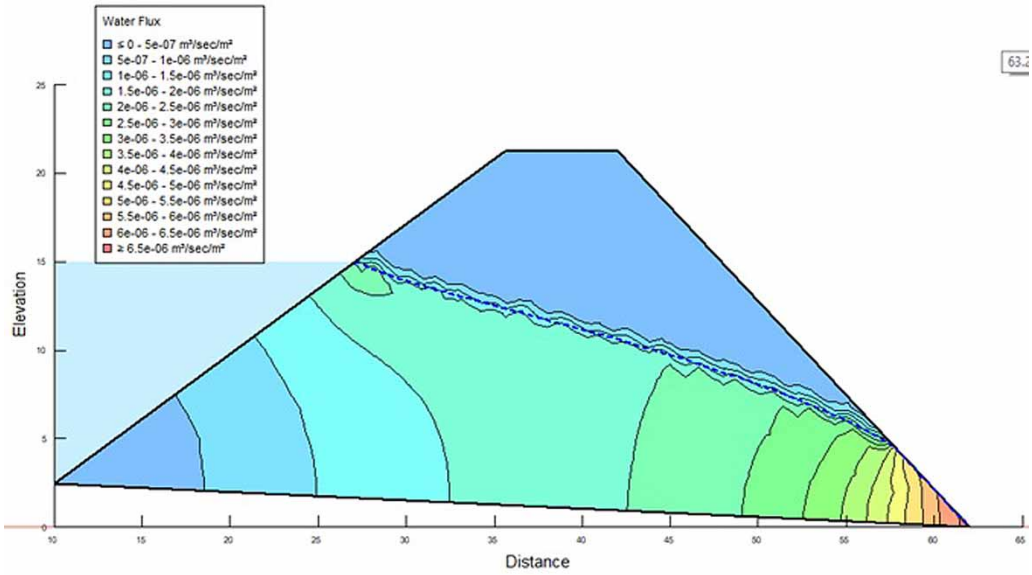


Figure 10 | Water flux in model 8.

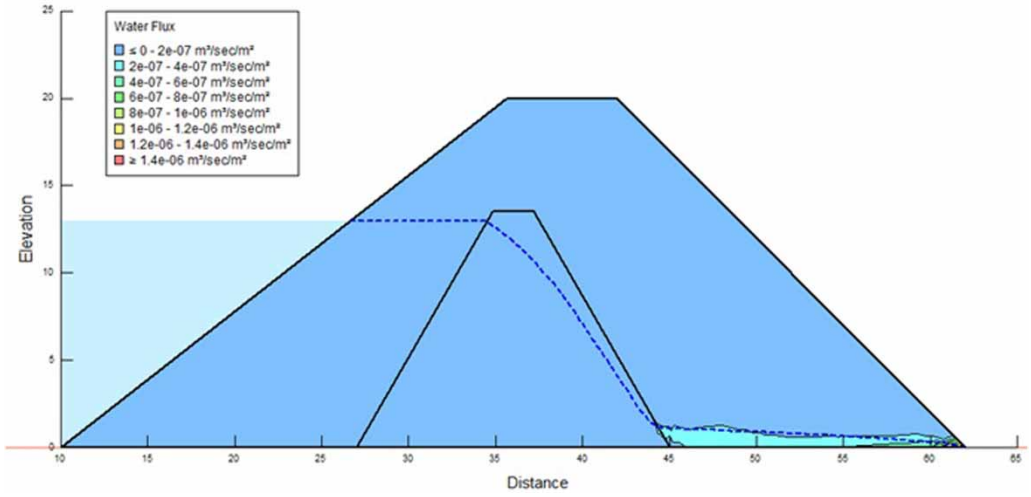


Figure 11 | Water flux in model 9.

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Corrected Proof

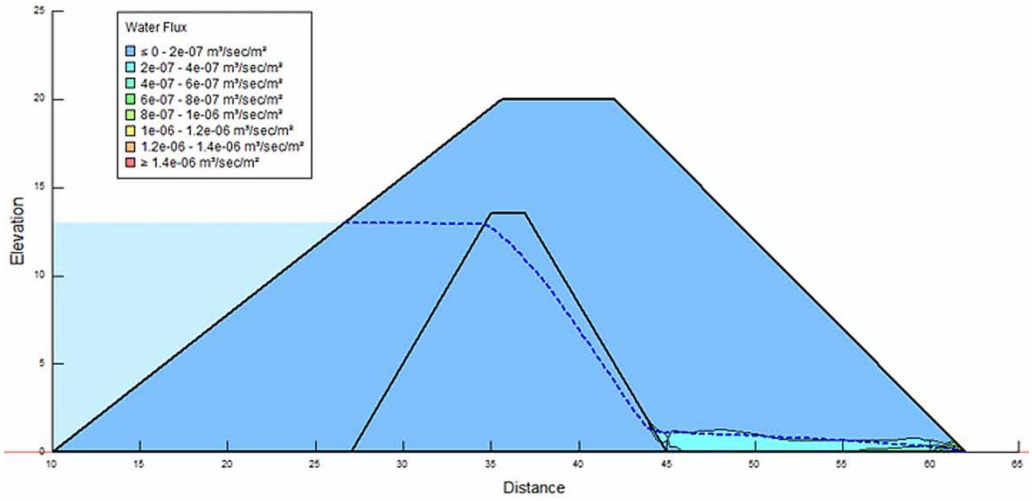


Figure 12 | Water flux in model 10.

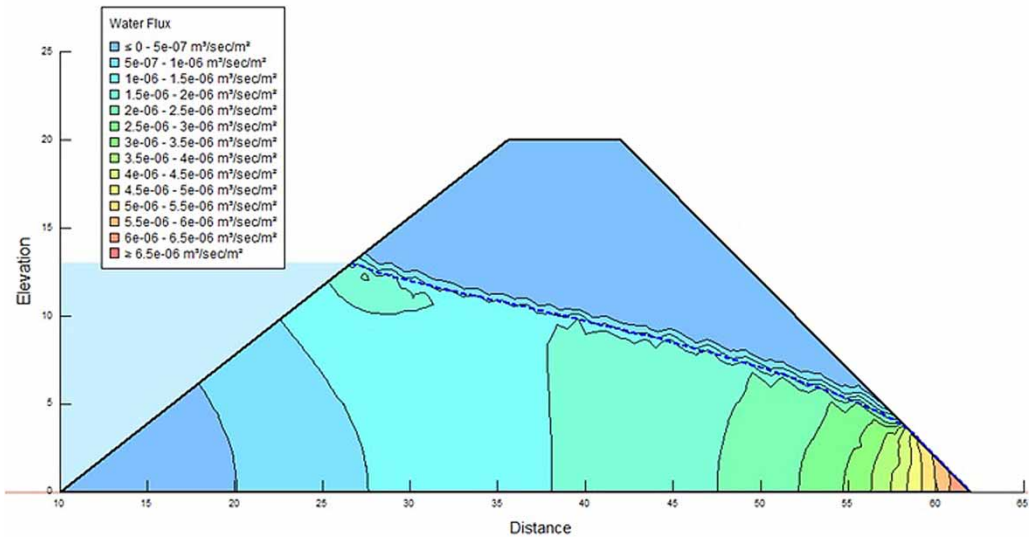


Figure 13 | Water flux in model 11.

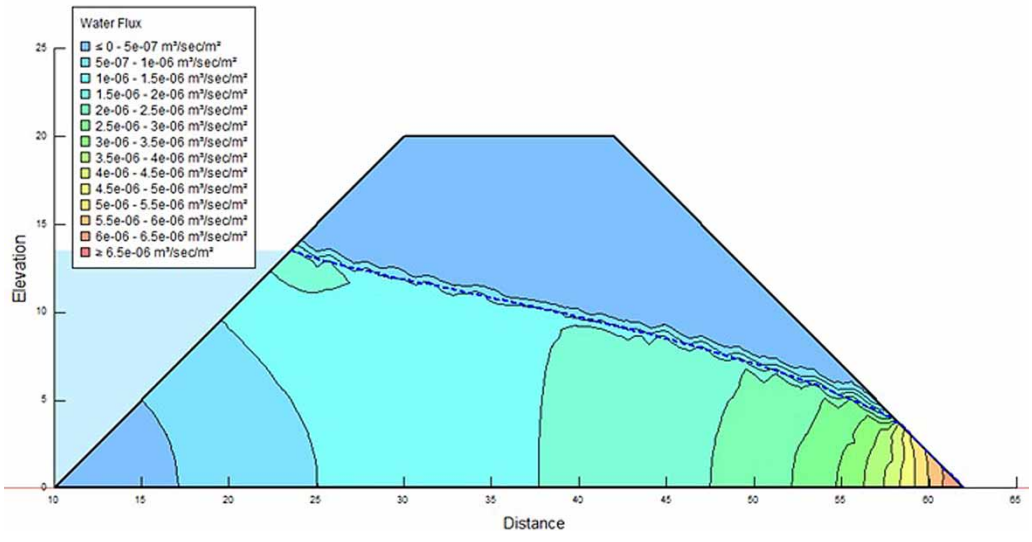


Figure 14 | Water flux in model 12.

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Corrected Proof

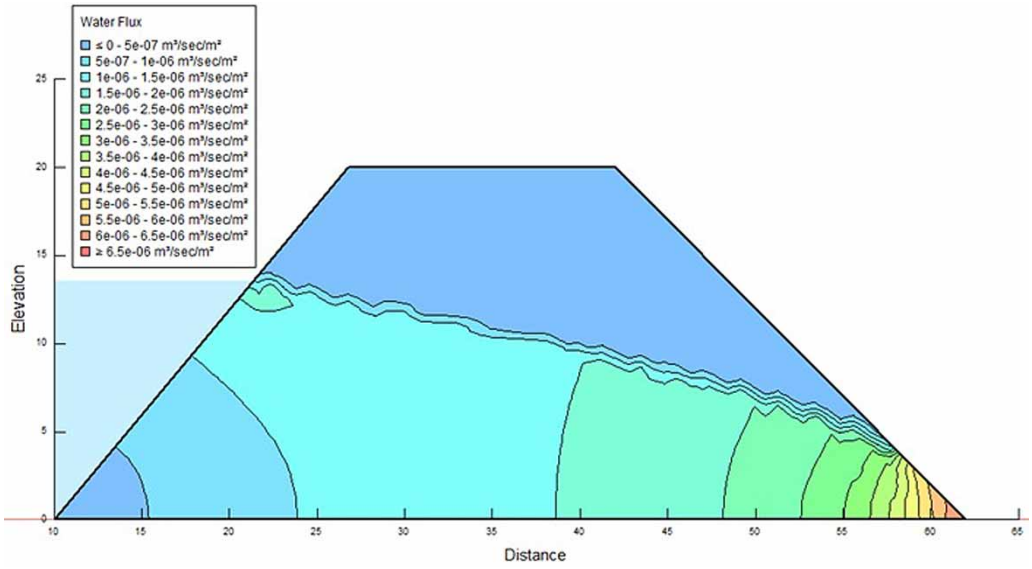


Figure 15 | Water flux in model 13.

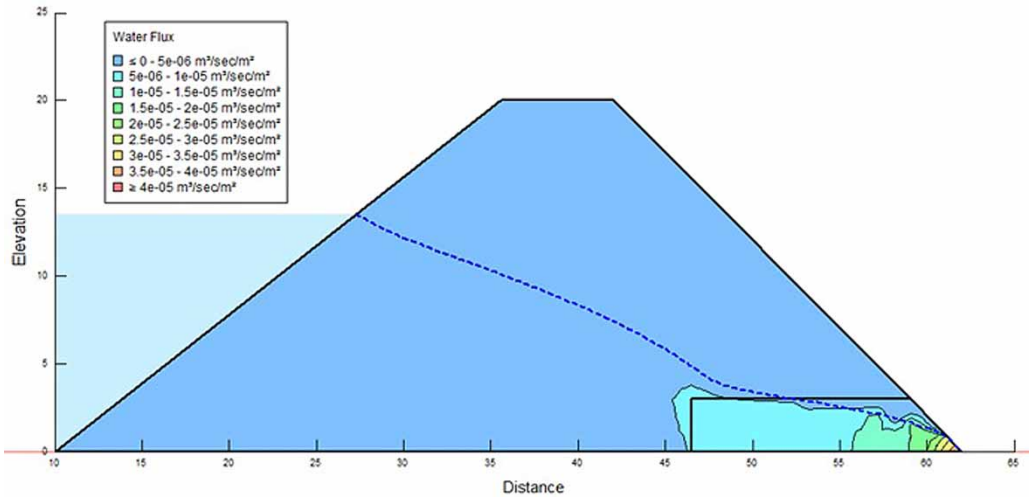


Figure 16 | Water flux in model 14.

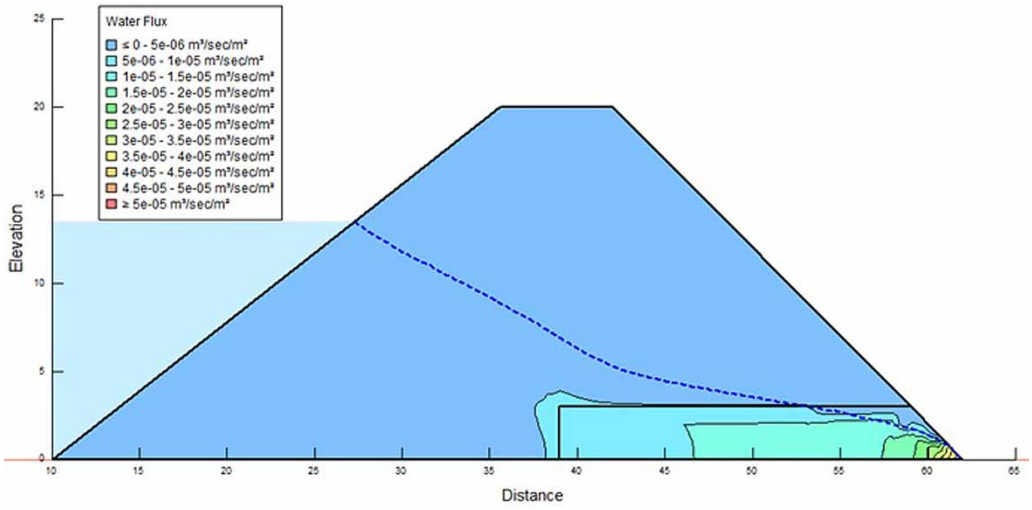


Figure 17 | Water flux in model 15.

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Corrected Proof

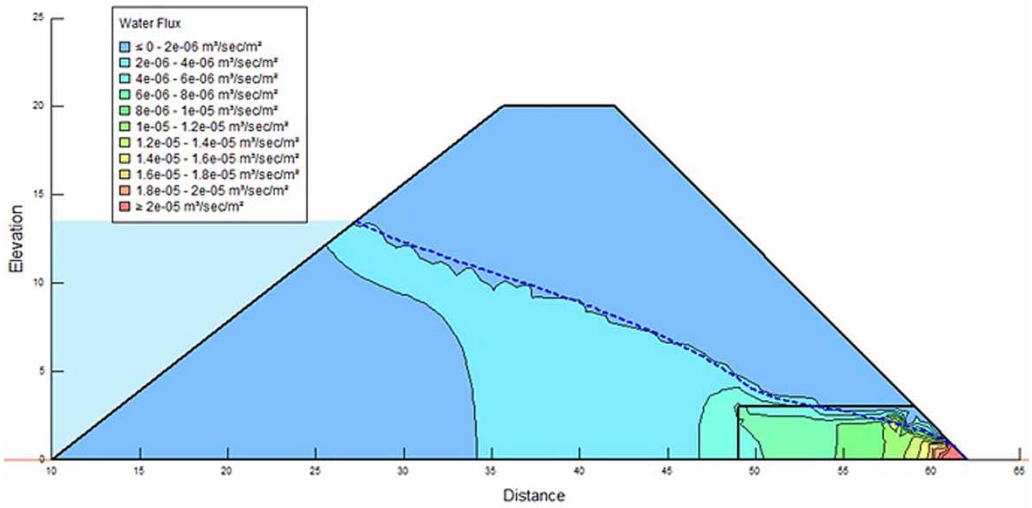


Figure 18 | Water flux in model 16.

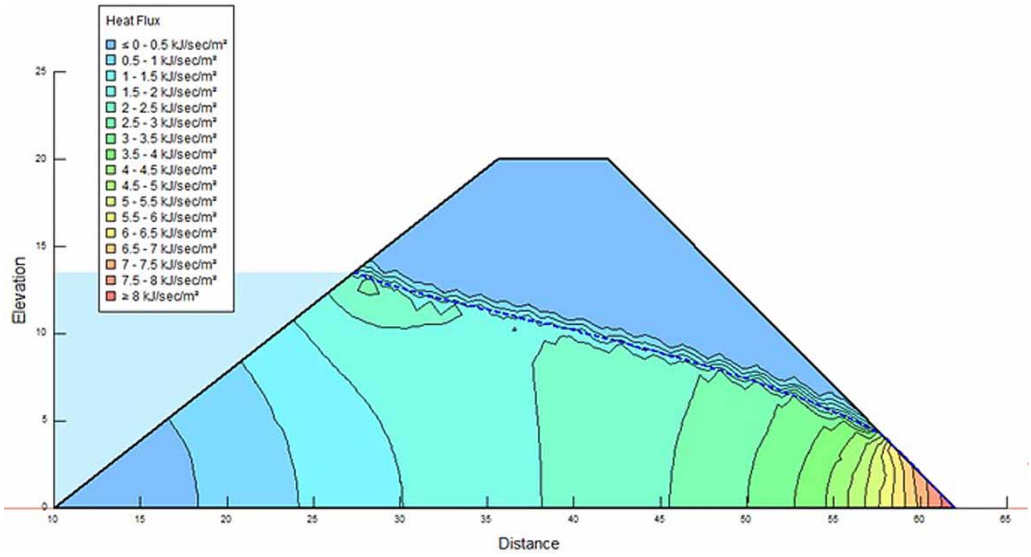


Figure 19 | Heat flux in model 1.

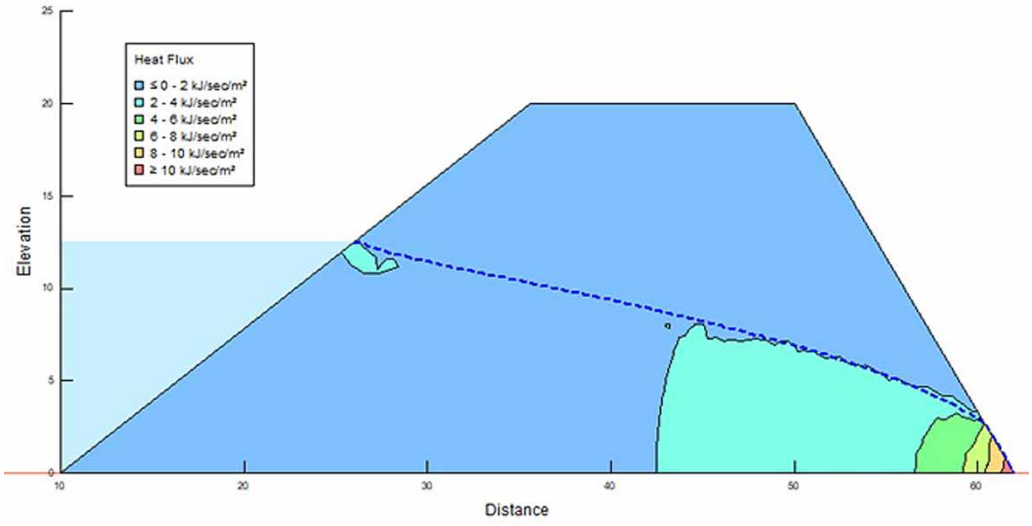


Figure 20 | Heat flux in model 2.

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Corrected Proof

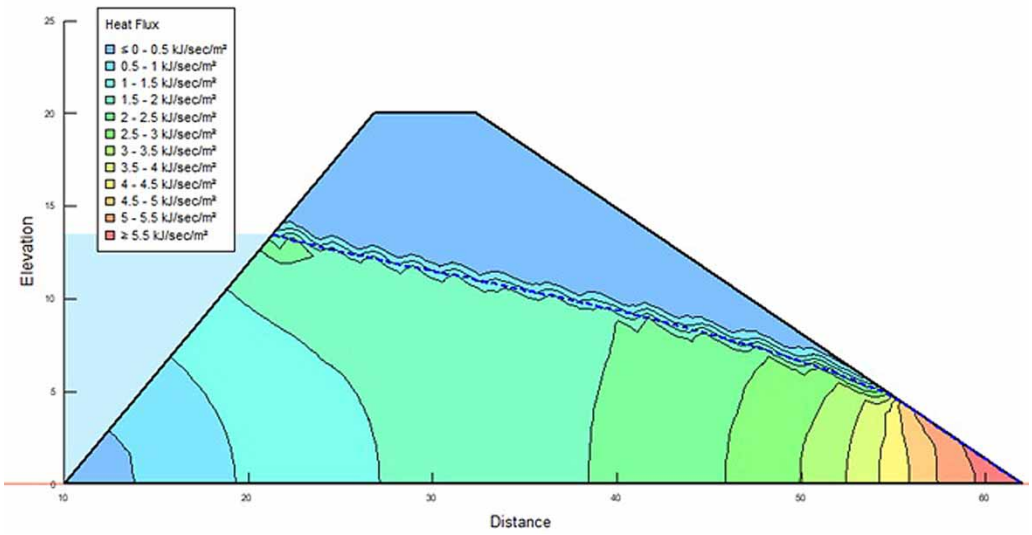


Figure 21 | Heat flux in model 3.

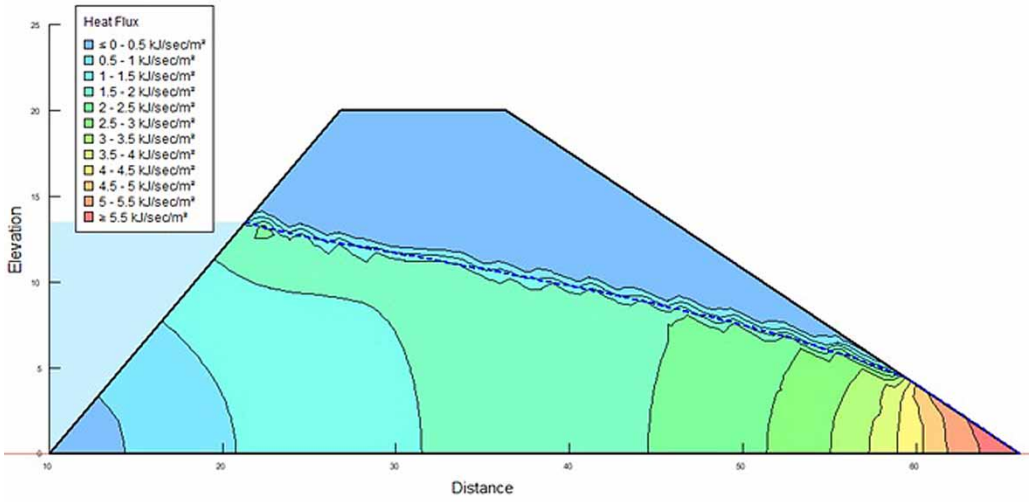


Figure 22 | Heat flux in model 4.

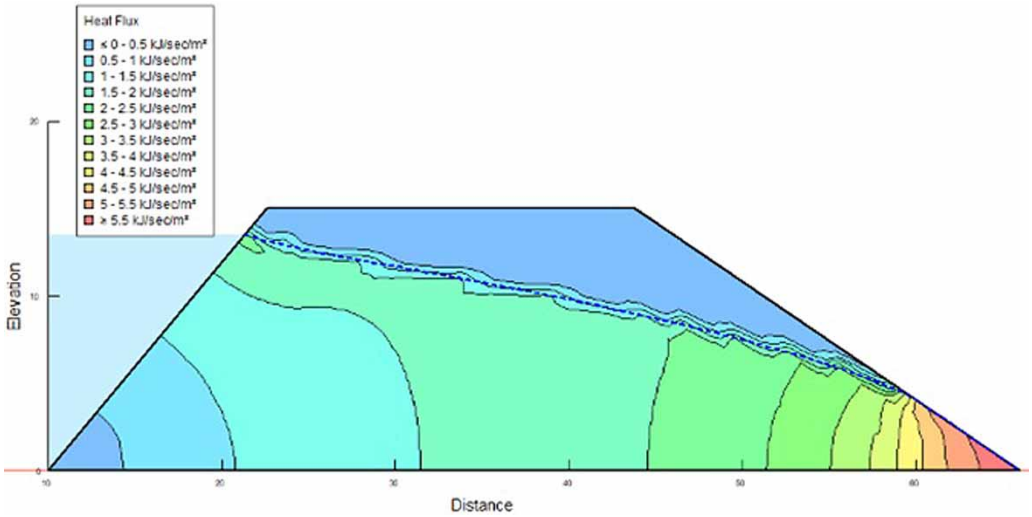


Figure 23 | Heat flux in model 5.

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Corrected Proof

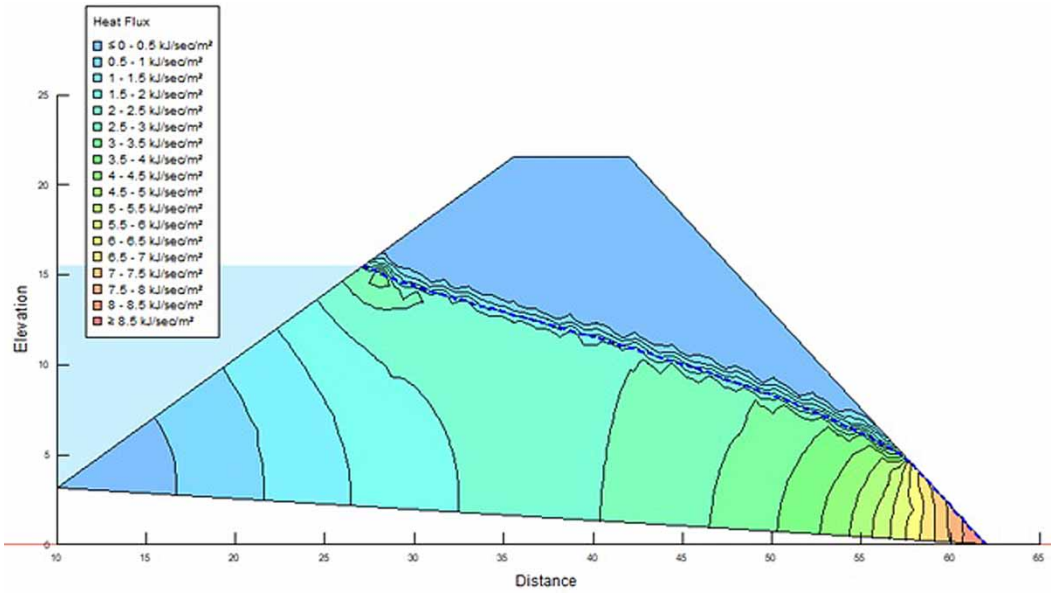


Figure 24 | Heat flux in model 6.

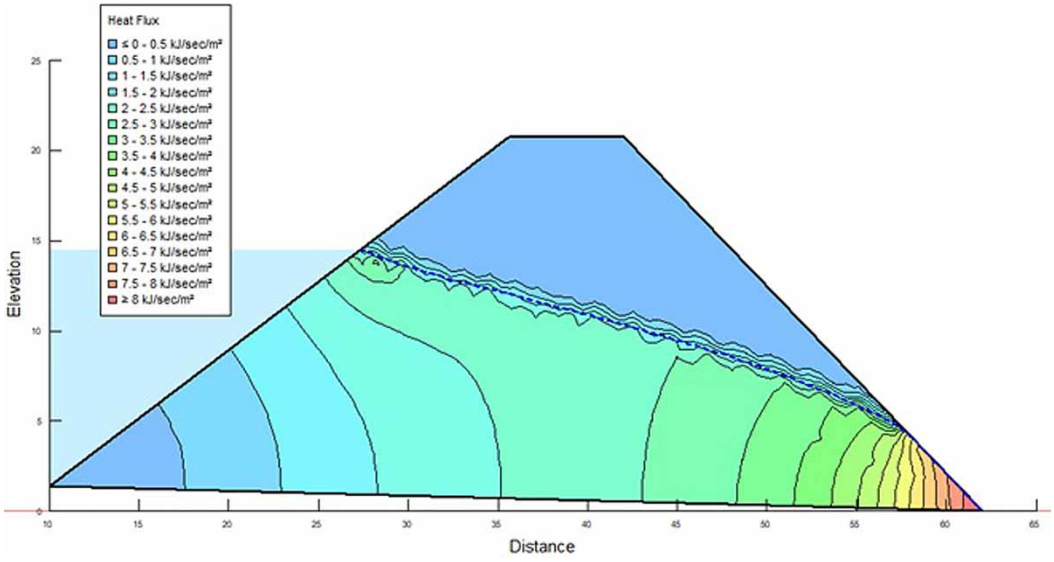


Figure 25 | Heat flux in model 7.

line marked with a dotted line in its respective figure is also shown, which keeps the flow of heat and water below it.

- Figure showing the graph of heat flux is Figure 35, with its x-axis showing the number of earth dam models, and the y-axis denoting heat flux in kJ/sec/m^2. Figure 36 represents a water flux graph with its x-axis as the number of earth dam models, and the y-axis denotes water flux in m^3/sec/m^2. While Figure 37 shows the flux on the y-axis to compare heat flux and water flux with the respective model shown on the x-axis.
- On increasing the downstream slope of an earth dam from 45 degrees (in model 1) to 59 degrees (in model 2), the water flux and heat flux increased by 34.81% and 34.37%. While increasing the downstream slope from 34 degrees (in model 3) to 45 degrees (in model 13), the water and heat flux increased by 42.36%.
- On increasing the upstream slope from 38 degrees (in model 1) to 45 degrees (in model 12); that is, 18.42%, the water flux and heat flux decrease by 0.34% and 0.99%, respectively. While increasing the slope from 38 degrees (in model 1) to 50 degrees (in model 13); that is, 31.57%, water flux and heat flux both decreased by 0.51%.

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Corrected Proof

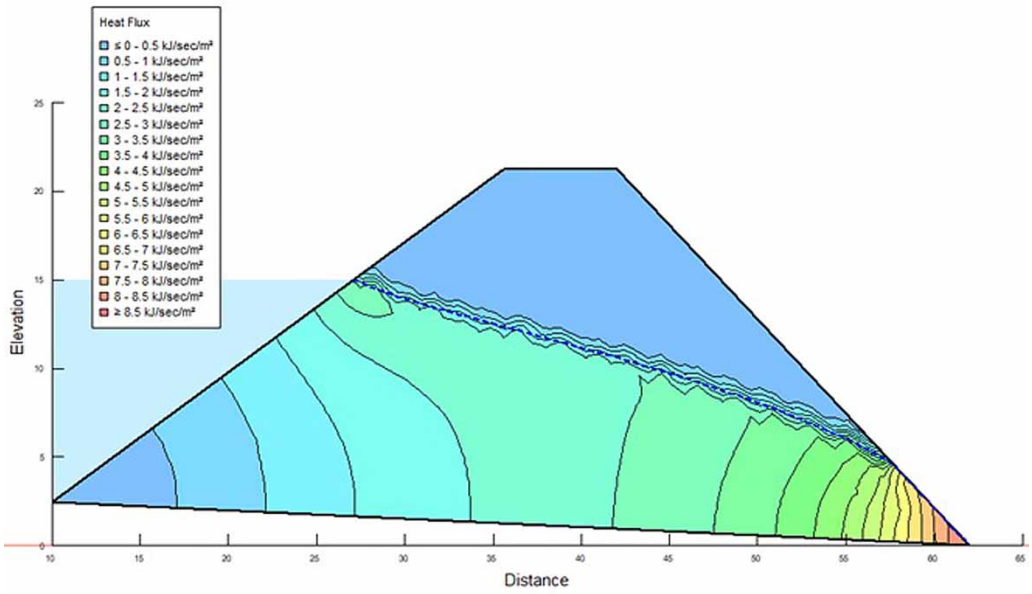


Figure 26 | Heat flux in model 8.

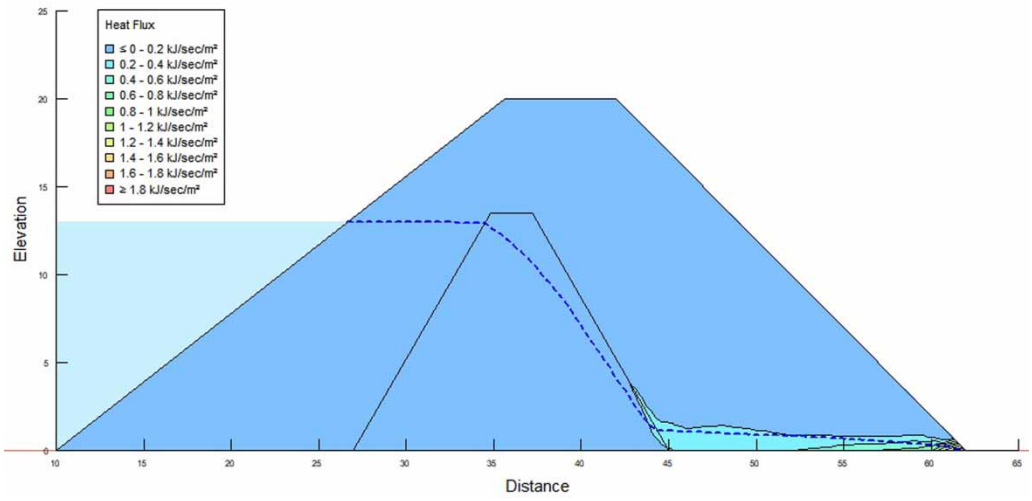


Figure 27 | Heat flux in model 9.

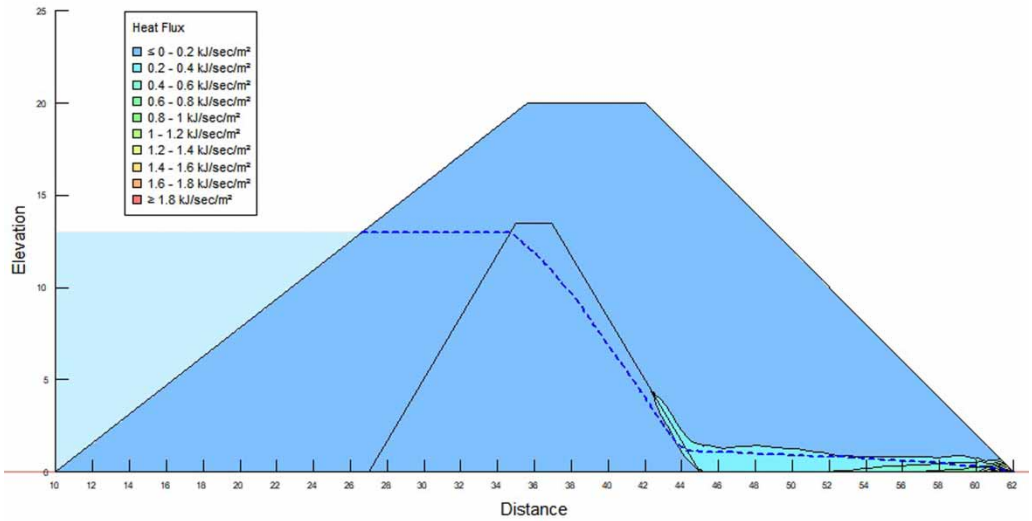


Figure 28 | Heat flux in model 10.

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Corrected Proof

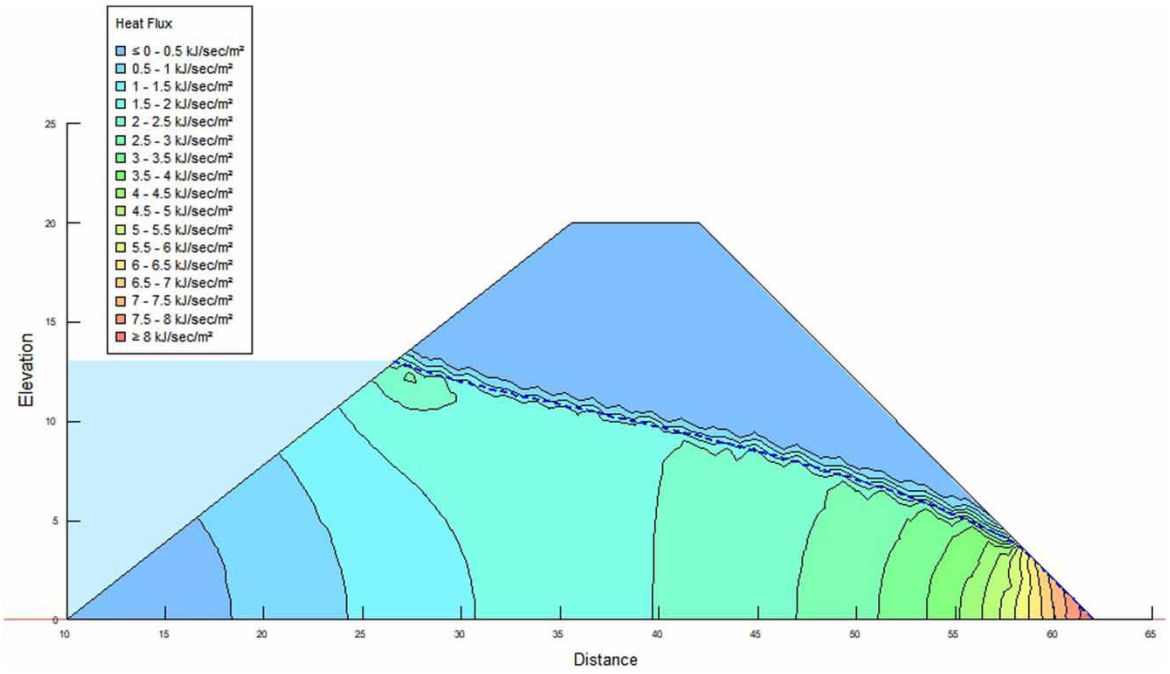


Figure 29 | Heat flux in model 11.

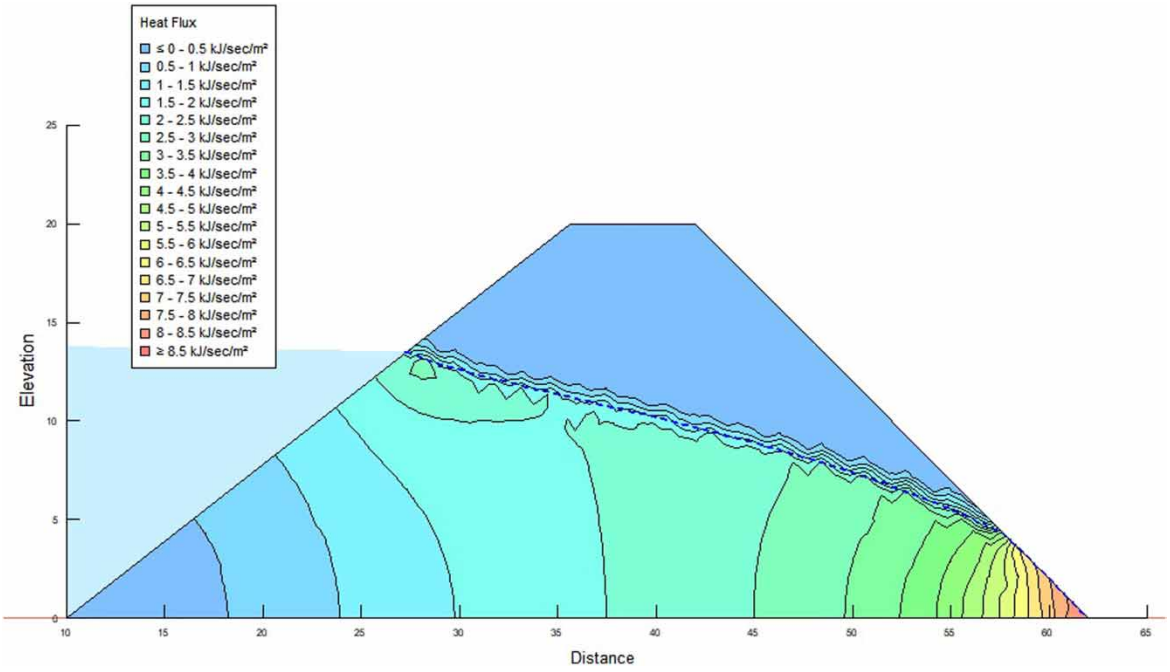


Figure 30 | Heat flux in model 12.

- On increasing the dam's length from 52 m (in model 3) to 56 m (in model 4); that is, by 7.69%, the water flux and heat flux both had an insignificant reduction of 0.02%.
- On increasing the earth dam's height for the same upstream head by 33.33%; that is, 15 m (in model 5) to 20 m (in model 4), the water flux and heat flux remain unchanged.
- On increasing upstream head of water from 13 m (in model 11) to 13.5 m (in model 1); that is, by 3.85%, the water flux and heat flux increased by 0.34%.

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Corrected Proof

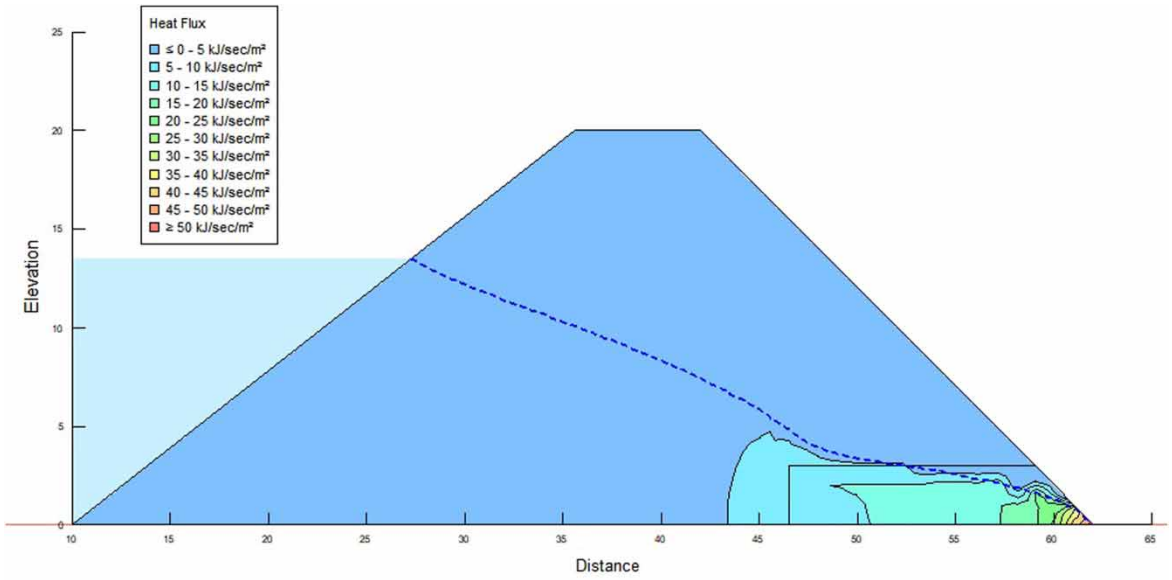


Figure 31 | Heat flux in model 13.

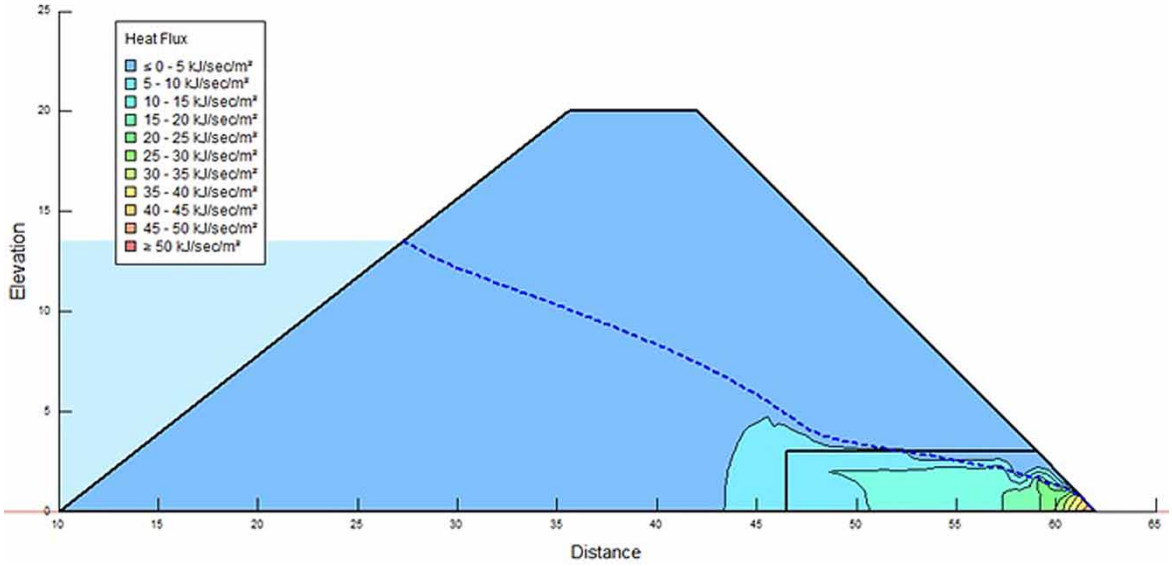


Figure 32 | Heat flux in model 14.

- On increasing the longitudinal slope by 2.28% (in model 1 to model 7), water flux and heat flux both increased by 1.33%, and increasing longitudinal gradient by 4% (in model 1 and 8), water flux and heat flux increased by 2.01% and 2.06% respectively. While increasing slope by 6.07% (from model 1 to model 6), both water flux and heat flux increased by 2.55%.
- With a decrease of the top width of clay core by 20.83% from model 9 to model 10, both the water and heat flux increase by 3.44%.
- With the downstream filter of length 13 m in model 16, 15.5 m in model 14, and 23 m in model 15, water flux increased by 414.23%, 533.38%, and 666.51%, respectively, and heat flux increased by 414.23%, 533.38%, and 666.75% respectively.

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Corrected Proof

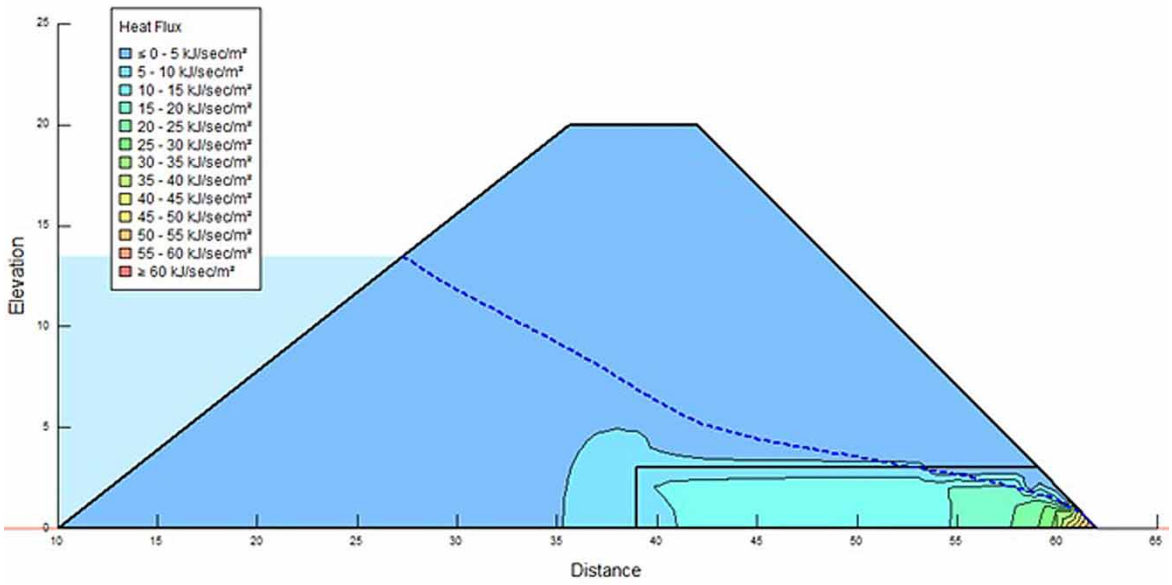


Figure 33 | Heat flux in model 15.

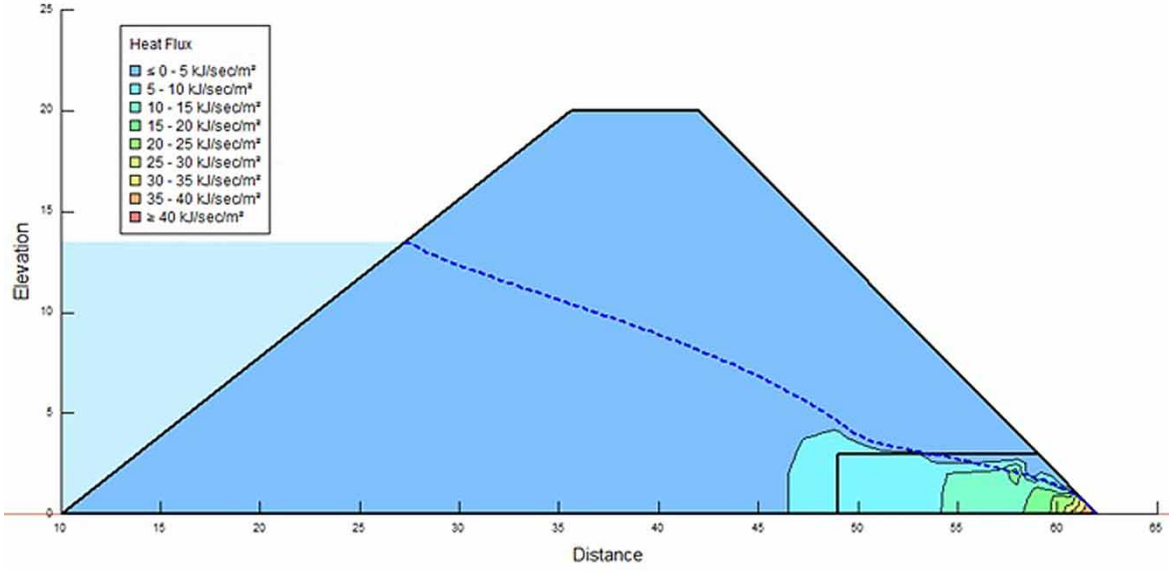


Figure 34 | Heat flux in model 16.

7. CONCLUSION

Sixteen earth dam models constructed in a hydraulic flume were built to study the water flux due to seepage and heat flux due to the convective heat transfer in the respective earth dam models. Temperature measurements were done in the earth dam model placed in a hydraulic flume using digital thermometers to determine the heat flux and validate it with the simulation result of Temp/w in Geostudio 2020. Thus it concluded by observed near walls of a hydraulic flume, soil pores of the dam were of larger volume, resulting in an increased local water flow. Regional flow variation due to wall channeling led to enhanced convection heat transfer, hence increasing the temperature. By increasing the earth dam's length, the heat and water flux had an insignificantly low reduction, and hence flux remained almost the same. With the increase in the dam's height with the same upstream water head, the heat and water flux remained unchanged. Heat flux obtained was found to be a helpful alternative in detecting seepage and water flux in an earth dam. As temperature sensors are cheaper than the

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Corrected Proof

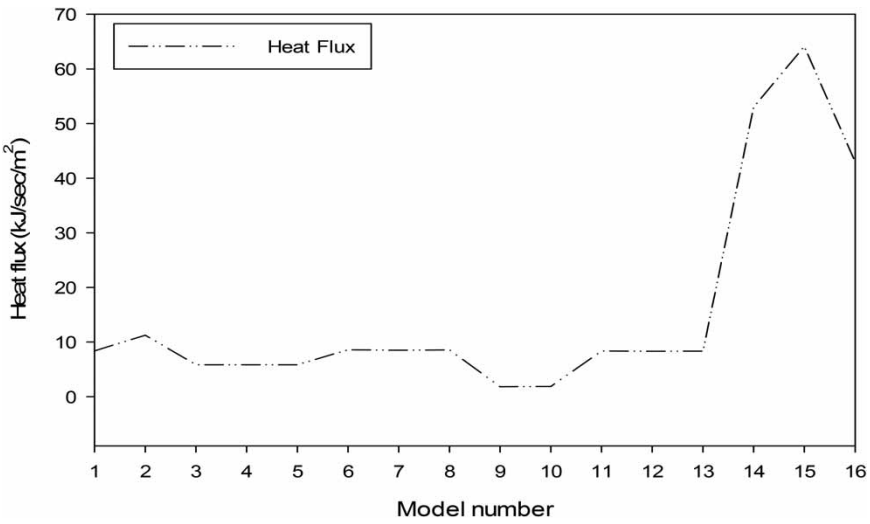


Figure 35 | Heat flux in different models of an earth dam.

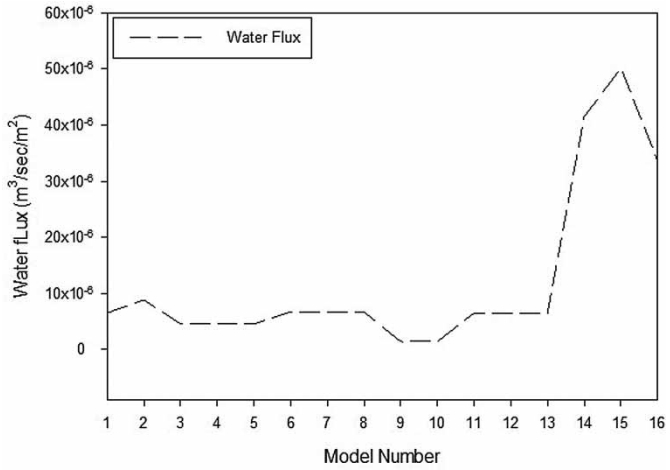


Figure 36 | Water flux in different models of an earth dam.

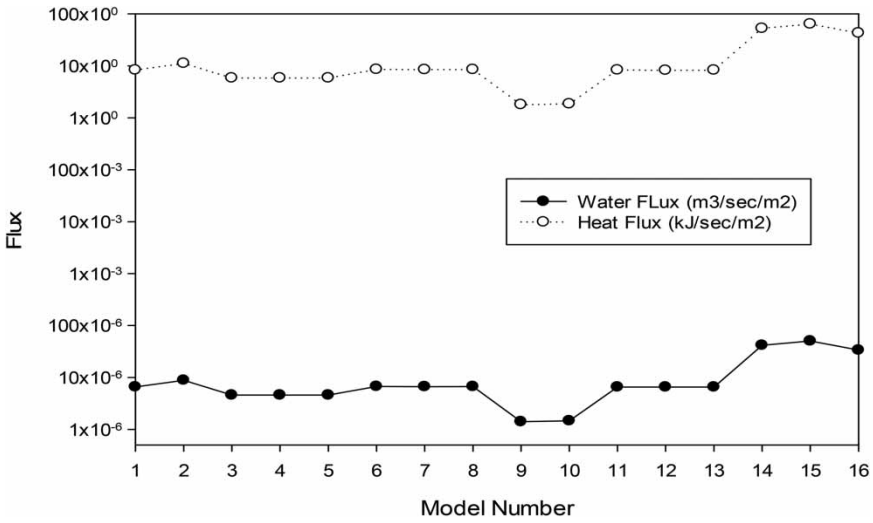


Figure 37 | Water and heat flux comparison in different models of an earth dam.

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Corrected Proof

discharge and water flow measuring devices, it helps to detect water at a reduced cost compared to conventional devices and hence improves the benefit-cost ratio of the project.

ACKNOWLEDGEMENTS

This work is supported by the Delhi Technological University India. Moreover, Sandeep Mishra's help for the XRD analysis and Lab head, Physics Department of the Delhi Technological University, India, is for the viscosity test.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

Abu-Hamdeh, N. 2003 Thermal properties of soils as affected by density and water content. *Journal of Biosystems Engineering* **86** (1), 97–102.

Abu-Hamdeh, N. 2014 Specific heat and volumetric heat capacity of granular materials as affected by moisture and density. *Journal of Applied Mechanics and Material* **575**, 103–107.

Alekseevich, A. N. & Sergeevich, A. A. 2017 Numerical modelling on tailings dam thermal-seepage regime considering phase transitions. *Journal of Modelling and Simulation in Engineering* **2017**, 7245413.

Alim, M. A., Ahmed, F. & Islam, M. S. 2017 Seepage analysis of Mahananda earthen embankment at Chapai Nawabganj in Bangladesh. *Journal of American Journal of Engineering and Technology Management* **2** (1), 1–6.

Amanifard, N., Borji, M. & Haghi, A. K. 2007 Heat transfer in porous media. *Journal of Brazilian Journal of Chemical Engineering* **24** (2), 223–232.

Badruddin, I. A., Azeem, Yunus Khan, T. M. & Baig, M. A. A. 2020 Heat transfer in porous media: a mini Review. *Materials Today: Proceedings* **24**, 1318–1321.

Buntebarth, G. 2020 Thermal properties of sand and mineral flours. *Journal of SN Applied Sciences* **2**, 396.

Chen, Y., Zhou, C. & Jing, L. 2009 Modeling coupled THM processes of geological porous media with multiphase flow: theory and validation against laboratory and field scale experiments. *Journal of Computers and Geotechnics* **36**, 1308–1329.

Chuvilin, E. & Bukhanov, B. 2019 Thermal conductivity of frozen sediments containing self-preserved pore gas hydrates at atmospheric pressure: an experimental study. *Journal of Geosciences* **9**, 65.

Cuong, B. Q., Thai, N. C., Yihong, Z. & Chunju, Z. 2017 Estimating seepage in embankment dams based on temperature measurement: a review paper. *Journal of International Journal of Engineering Research and Technology (IJERT)* **6** (1), 106–113.

Del Piero, G. 2020 A mechanical model for heat conduction. *Journal of Continuum Mechanics and Thermodynamics* **32**, 1159–1172.

Engineering ToolBox 2003 *Specific Heat of Some Common Substances*. Available from: https://www.engineeringtoolbox.com/specific-heat-capacity-d_391.html [1/7/2021].

Engineering ToolBox 2004 *Water- Specific Heat*. Available from: https://www.engineeringtoolbox.com/specific-heat-capacity-water-d_660.html [10/28/2020].

Ižvolt, L. & Dobeš, P. 2014 Test procedure for the value of specific heat capacity and thermal conductivity coefficient, XXIII R-S-P seminar, Theoretical Foundation of Civil Engineering (23 RSP) (TFoCE 2014). *Journal of Procedia Engineering* **91**, 453–458.

Johansson, S. & Sjödahl, P. 2004 Downstream seepage detection using temperature measurements and visual inspection monitoring experience from Røsvatn field test dam and large embankment dams in Sweden. In *Proc. Intl. Seminar on Stability and Breaching of Embankment Dams*, p. 21.

Kodešová, R., Vlasáková, M., Fér, M., Teplá, D., Jakšík, O., Neuberger, P. & Adamovský, R. 2013 Thermal properties of representative soils of the Czech Republic. *Journal of Soil and Water Research* **8** (4), 141–150.

Kosky, P., Balmer, R., Keat, W. & Wise, G. 2012 *Exploring Engineering*, 3rd edn. Elsevier, New York, NY, pp. 259–264.

Kurz, D., Alfaro, M. & Graham, J. 2017 Thermal conductivities of frozen and unfrozen soils at three project sites in northern Manitoba. *Journal of Cold Regions Science and Technology* **140**, 30–38.

Misra, A., Becker, B. R. & Fricke, B. A. 1995 A theoretical model of the thermal conductivity of idealized soil. *HVAC&R Research* **1** (1), 81–96.

Radzicki, K. & Bonelli, S. 2010 Thermal Seepage monitoring in the earth dams with impulse response function analysis model. In: *8th ICOLD European Club Symposium*, Innsbruck, Austria, pp. 624–629.

Tokoro, T., Ishikawa, T., Shirai, S. & Nakamura, T. 2016 Estimation methods for thermal conductivity of sandy soil with electrical characteristics. *Journal of Soil and Foundation, The Japanese Geotechnical Society* **56** (5), 927–936.

Yousefi, S., Ghiassi, R., Noorzad, A., Ghaemian, M. & Kharaghani, S. 2013 Application of temperature simulation for seepage inspection in earth-fill dams. *Journal of Gradevinar* **65** (9), 825–832.

First received 8 April 2021; accepted in revised form 13 June 2021. Available online 23 June 2021

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2021.241/908469/wst2021241.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP - 2021)

HTGAN : An architecture for Hindi Text based Image Synthesis

Dr. Anil Singh Parihar
Dept. of Computer Engineering
Delhi Technological University
New Delhi, India
anil@dtu.ac.in

Aditya Kaushik
Dept. of Computer Engineering
Delhi Technological University
New Delhi, India
adityakaushik99@gmail.com

Aditya Vikram Choudhary
Dept. of Computer Engineering
Delhi Technological University
New Delhi, India
aditya14483@gmail.com

Amit Kumar Singh
Dept. of Computer Engineering
Delhi Technological University
New Delhi, India
amitkumarsingh2750@gmail.com

2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP) | 978-1-6654-3277-1/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCCSP52374.2021.9465527

***Abstract*—Synthesis of high-quality images that are semantically consistent with their text descriptions has emerged as an essential and onerous problem in the areas of CV and NLP. Various significant steps have been taken in the task of generation of images using English textual descriptions, with Multimodal GANs being at the forefront of all these efforts. In this work we attempt to extend the existing English text to image generation (T2I) techniques to the novel task of Hindi T2I using language translation models. To achieve this, the input Hindi sentences were translated to English using a transformer based Neural Machine Translation module the output of which was then fed to the Generative Adversarial Networks based Image Generation Module. The outcomes of this approach have been evaluated using well established metrics such as Inception Score and BLEU Score which yield scores that indicate the generation of lifelike images which semantically align with the input text descriptions.**

***Keywords*—Text to Image, Neural Machine Translation, Transformers, Generative Adversarial Networks.**

I. Introduction

Various Innovative practices have been put to use for the task of generating photo-realistic images from textual description. T2I refers to the generation of images using some textual inputs. The textual inputs are descriptive captions in Hindi. Since we require images that are realistic in nature and that match with the textual descriptions at the same time this task is very complex in nature. So the completion of this task requires a novel and unique amalgamation of NLP and CV techniques.

Standalone Image Generation in itself has been a relatively new domain of research with the earliest breakthrough by Goodfellow et al. [1] in 2014, while Conditional English text-to-image generation is even more underexplored which became an area of active research in 2016 with GAN-INT-CLS [2] and still has a huge room for improvement in terms of the generalization capabilities of the techniques and the quality of images generated. Hindi T2I presents itself as an area of research that has been untouched given the paucity of datasets in this field and the inherent structural and grammatical differences between a resourceful language like English and a low-resource language like Hindi which in turn inhibits the ability to scale the contemporary techniques in T2I to a low-resource setting like Hindi T2I. The methodology being proposed in this paper is a first step in the direction of solving the unique and intriguing challenge of Hindi T2I and shows significantly promising results in a low-resource setting with sufficient generalization capabilities.

II. Related Work

Lifelike T2I has been a task that has been a topic of interest in the field of both Computer Vision as well as Natural Language Processing. T2I tasks have primarily been tackled using Generative Models such as Generative Adversarial Networks and Variational Autoencoders. In [3], Kingma & Welling introduced the Auto-encoding Variational Bayes Algorithm which paved the way for utilisation of Variational Autoencoders in Image Generation tasks with the use of probabilistic methods. Rezende et al. [4] by introducing backpropagation of stochastic variables for optimisation of variational lower bound likelihoods further upgraded the performance. Gregor et al., in DRAW [5], combined attention mechanisms with iteratively image-enhancing VAE networks which conditioned the image generation process on spherical gaussians to further improve this process. With PixelRNN [6], Oord et al. modelled the pixels as discrete variables and by utilising Row-level LSTMs and Diagonal-level Bi-LSTMs for iterative pixel generation conditioned on previous rows and diagonally adjacent pixels, further enhanced the image synthesis methodology.

In 2014, Mirza & Osindero introduced ConditionalGAN [7] which facilitated the generation of images which were conditioned on latent inputs apart from the input noise vector, for the first time. LAPGAN [8] further improved upon this by developing a laplacian-pyramid-like hierarchy of Conditional GANs which were conditioned on output images of the previous step. Mansimov et al. built upon the groundwork laid by [5], [7] and [8] with their work AlignDRAW [9] which sought to synthesize lifelike images from their text captions by

978-1-6654-3277-1/21/$31.00 ©2021 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:33:40 UTC from IEEE Xplore. Restrictions apply.

integrating a Bidirectional-Attentional RNN Language Model with a modified version of DRAW which conditioned the image generation on the outputs of the language model based on the image captions.

GAN-INT-CLS [2] introduced by Reed et al. in 2016 was the first to utilise text based conditioning in GANs. In addition to this, Reed et al. also introduced a matching-aware discriminator which was capable of identifying mismatched pairs of text captions and images as well as manifold interpolation which allowed for utilising points lying in between conditioning manifold representations for two different text descriptions as inputs for image generation leading to a more enriched training process and more realistic image generation as a result. In 2016, Reed et al. introduced GAWWN [10] which further extended the latent conditioning manifold by utilising both text and image features to generate a synthetic image from these two by creating a "Structured Joint Embedding of visual descriptions and images", which allowed the generator output to be conditioned on both the textual descriptions as well as the original image. Nguyen et al., in 2017, developed Plug and Play Generative Networks [11] by introducing Conditioning Networks by using learned priors developed with Denoising Autoencoders. Zhang et al. in StackGAN [12] and StackGAN++ [13] built upon the idea of Denton et al. in [8] by stacking multiple stages of GANs together with each level producing a higher quality image than the previous stage while iteratively improving upon the previous output as well. Additionally, Zhang et al. also introduced Conditioning Augmentation in [12] for generating additional training samples as well as Joint Approximation of Conditional and Unconditional Probability Distributions in [13] which allowed for the discriminator loss to account for how lifelike the generated images are and how semantically consistent they are with their text descriptions. AttnGAN [14] by Xu et al. introduced the use of attention mechanisms in the structure of multiple GANs for image generation which allowed the GANs to focus on specific sections of text when generating pixels pertaining to those sections along with Deep Attentional Multimodal Similarity Model which assessed the generated images for their semantic similarity with the input text descriptions and hence adding an additional parameter to the generator loss.

Qiao et al. in [15] further improved upon the work of Xu et al. by introducing Global-Local Attention which facilitated the GANs to apply both word-level as well as global attention in the process of generating a section of the image pertaining to those particular words or sentences. Additionally, Qiao et al. introduced Semantic Text Regeneration and Alignment Module (STREAM) which facilitated the reconstruction of text descriptions from the images so generated and the measure of dissimilarity between the text so generated and original text description was taken into account in the loss of the generator module which is along the lines of the element of reconstruction loss utilised in CycleGAN [16] and DualGAN [17] both of which regenerate input image distribution from the output image distribution by joint training of bidirectional GANs.

Our work also entails the requirement for a robust translation system. Machine Translation is a task that has been addressed using both statistical methods as well as neural networks. Yamada & Knight [18] , in 2001, developed a syntax-based SMT system by applying probabilistic operations on the parse tree of the input language and utilising expectation-maximization algorithms for estimation of model parameters. Marcu & Wong [19] introduced a phrase-based translation system to learn phrase-word equivalency in bilingual corpora using a joint probability model, which further improved upon the performance of [18]. However, in recent times neural machine translation techniques have proven to be better capable of handling the task of machine translation with greater generalization capabilities over parallel corpora. Various neural networks such as LSTM [20] and GRU [21] have been put to application for machine translation tasks and have shown great efficacy thus giving rise to the popularity of NMT. This has also been aided by the increased availability of parallel corpora for varied language pairs.

Bahdanau et al. [22], in 2014, introduced attentional decoder networks for machine translation tasks which utilised a bidirectional RNN encoder network for transforming the source language inputs which are then provided as inputs to the attentional decoder network. In [23] , Vaswani et al. introduced Transformer Networks for machine translation tasks which introduced self-attention, positional encodings and multi-head attention which led to significant performance gains in almost all Natural Language tasks including machine translation. Most current state-of-the-art models in the domain of machine translation are transformer networks based upon the work of Vaswani et al.

III. HTGAN

As shown in *Figure* 1, the architecture of HTGAN for the Hindi T2I task comprises of a combination of a Translation Module and an Image Generation Module, which involves a transformer model for translation followed by a Conditional GAN based on DM-GAN [24], put forward by Zhu et al., for image generation. The Hindi image description for which the image has to be generated is first translated to an English description by the translation module which is then used as an input to the Image Generation Module which generates the image based on the aforementioned translation output.

A. Translation Module

The translation module consists of a transformer model based on [23] trained on the Hindi-English Parallel Corpus in the Hindi Visual Genome Dataset. The transformer consists of encoder-decoder stacks which comprise of the following:

1) Multi-Head Attention: The transformer utilises a scaled version of dot-product attention within the encoder and decoder units. For a tuple of (*Query, Key, Value*) the attentional function can be stated as follows :

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:33:40 UTC from IEEE Xplore. Restrictions apply.

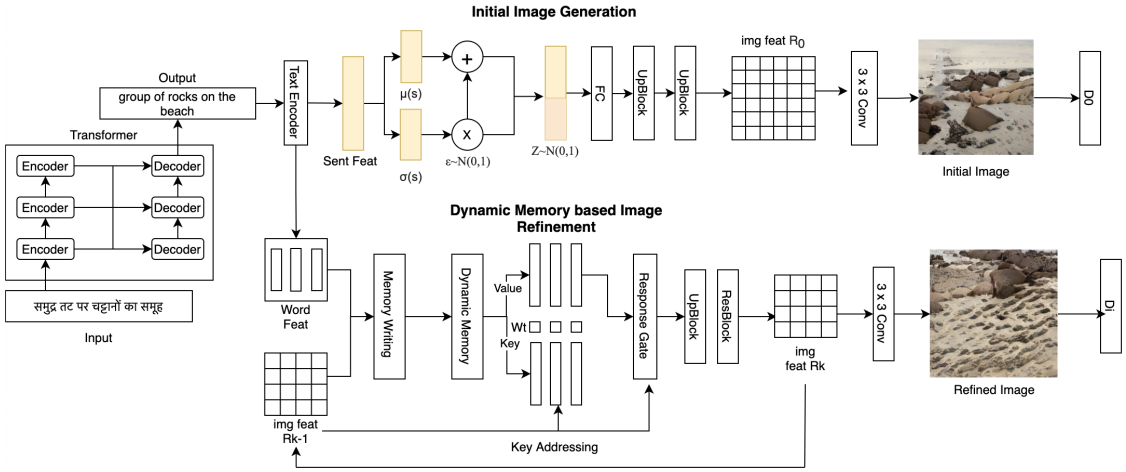


Fig. 1. Model Architecture for HTGAN

$$Att(Qry,Key,Val) = softmax\left(\frac{Qry\mathrm{Key}^T}{sqrt\left(dim_{Key}\right)}\right) * Val \tag{1}$$

where dim_{key} represents the dimension of the Key vector the square root of which acts as the scaling factor to avoid vanishing gradients in the softmax layer. Multi-head attention generates *p* different linear projections of the queries, keys and values, on which attention is applied in parallel fashion. The *p* attentional encodings so generated are then concatenated and projected through another linear layer.

$$MHA(Qry, Key, Val) = concat\left(AH_1, AH_2, .., AH_P\right) * W^O_{LP} \tag{2}$$

$$AH_i = Att\left(QryW^{Qry}_{LP}, KeyW^{Key}_{LP}, ValW^{Val}_{LP}\right) \tag{3}$$

where *MHA* represents the Multi-Head Attention, *AH* represents the Attention Heads and W_{LP}'s represent the weights for the linear projection layers.

2) Self-Attention: Both the encoder and decoder layers utilise slightly different variations of self-attention. In the encoder stack, Self-Attention refers to the process of generating the (*Query*, *Key*, *Value*) tuples for the *i*th encoder layer on the basis of the output of the *i*-1th encoder layer which allows the *i*th encoder layer to attend to all the sequence positions in the *i*-1th layer's output. Although the self-attention in the decoder stack also follows the same principle as stated above, the key difference in the self-attention used in the decoder layers is that the weights corresponding to words that have not yet been generated need to be masked out by setting these weights to -∞.

3) Encoder-Decoder Attention: The decoder layers also consist of encoder-decoder attentional layers, wherein the Key and Value inputs for the *i*th decoder layer come from the output of the *i*th encoder in order to facilitate the decoder to attend on the entirety of the input sequence while the Query input is taken from the output of the *i*-1th decoder layer.

4) Positional Encodings: The input text embeddings are combined with positional encodings which represent information about the absolute and relative positions of sequence tokens in relation to each other. The dimensionality of these encodings are the same as that of the input embeddings so that the two can be added with greater ease.

The encoder-decoder stacks consist of *N* layers where each layer consists of the components as stated above. The encoder stack is responsible for generating encodings from the input Hindi text which are then used as inputs for the decoder stack in order to generate the translated English text.

B. Image Generation Module

The Image Generation Module consists of a Conditional GAN model based on DM-GAN by Zhu et al. This module generates images in an iterative fashion which begins with the generation of an initial image conditioned on the sentence features as well as the latent noise input. The image so generated is iteratively refined using a dynamic-memory based refinement process.

1) Initial Image Generation: In the initial stage, the image is generated based on a latent noise input as well as the textual features extracted by the text encoder viz. the sentence level-features *s* and word-level features *W*.

$$R_0 = G_0(Z, s) \tag{4}$$

R_0 represents the image features so generated, while *z* represents latent noise input selected from a Gaussian Distribution. The image x_0 so generated represents a rough outline of the image to be generated.

2) Dynamic-Memory based Refinement: The image initially generated (x_0) in the first stage is iteratively enhanced through the use of a Key-Value Dynamic Memory Network in conjunction with the various stages of the GAN. The Dynamic Memory component comprises of four different stages as follows:

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:33:40 UTC from IEEE Xplore. Restrictions apply.

- *Gated Memory Writing* : In this stage of the network, the image features generated from the previous stage GAN are merged with the word-level textual features *W*. This is used to calculate the significance of each word in generating image features, g_i^w which in turn is combined with the aforementioned image and textual features. This result is stored in the memory for future use.

$$g_i^w = \sigma\left(\gamma * w_i + \delta * \Sigma r_i / N\right) \tag{5}$$

$$m_i = Conv_{wf}\left(w_i\right) * g_i^w + Conv_{if}\left(\Sigma r_i /_N\right) * \left(1 - g_i^w\right) \tag{6}$$

  In Equation 5, γ represents the text-feature coefficient matrix whereas δ represents the image-feature coefficient matrix. On the basis of g_i^w, the memory features are generated as described in Equation 6, where $Conv_{wf}$ and $Conv_{if}$ represent the convolutional operations for the word and image features so as to embed those into the same feature space as that of the memory features. r_i represents the *i*th image feature.
- *Key Addressing* : For the purpose of refinement of the images generated at the *i*-1th stage of the GAN, at the *i*th stage relevant memory features need to be retrieved to allow this stage to focus on generating fine-grain details for the important image features missing in the *i*-1th stage. The key to be retrieved is evaluated using a similarity based weight which is as follows for the *i*th memory slot and *j*th image feature

$$n_{ij} = \exp\left(\Omega_K\left(m_i\right)^T r_j\right) / \sum \exp\left(\Omega_K\left(m_i\right)^T r_j\right) \tag{7}$$

  Here, Ω KEY represents the key access method in the dynamic memory key-value store while m_i represents the *i*th memory slot and r_j represents the *j*th image feature.
- *Value Retrieval* : On the basis of the similarity measure that is obtained in the process of key addressing, the output memory feature is evaluated as follows

$$OMF_j = \Sigma n_{ij} \Omega_V\left(m_i\right) \tag{8}$$

  where Ω VALUE represents the value access method in the dynamic memory key-value store.
- *Gated Memory Response* : Using the output memory feature calculated during the Value Retrieval step the image generated at the *i*-1th stage is enhanced by concatenating the image features from the *i*-1th stage with the output memory features and then applying sigmoid activation function on the concatenation.

$$Gate_{Response} = \sigma\left(W concat\left(OMF_i, r_i\right) + b\right) \tag{9}$$

$$r_i^{new} = OMF_i * Gate_{Response} + r_i * \left(1 - Gate_{Response}\right) \tag{10}$$

  Here, $Gate_{Response}$ represents the response gate which contributes towards the image enhancement, σ represents the sigmoid activation function and *W* and *b* are the weight and bias parameters for the activation function.

3) Objective functions: The generator networks are trained to optimise the following objective function:

$$L = \Sigma L_{gen_i} + \alpha * L_{CA} + \beta * L_{DAMSM} \tag{11}$$

Here α and β represent the weight coefficients for Conditional Augmentation Loss and DAMSM Loss respectively.

The adversarial loss for the *i*th generator is defined as follows:

$$L_{gen_i} = -\left(EV_{x \sim p_{gen_i}} \log dis_i(x)\right)/2 - \left(EV_{x \sim p_{gen_i}} \log dis_i(x, s)\right)/2 \tag{12}$$

where the first term represents the visual realism loss while the second term represents the semantic consistency loss.

L_{CA} corresponds to the loss for the conditional augmentation step, wherein random inputs are sampled from a probability distribution generated on the basis of the mean and covariance of the input image distribution to reduce overfitting, which is measured by evaluating the Kullback-Leibler Divergence between the probability distribution so generated and a standard Gaussian Distribution. L_{DAMSM} represents the loss for the DAMSM module which measures the semantic similarity between the input text and the output image generated.

Loss for the discriminator network for the ith stage is given as :

$$L_{dis_i} = -\left(EV_{x \sim p_{input}} \log dis_i(x)\right)/2 + \left(EV_{x \sim p_{gen_i}} \log\left(1 - dis_i(x)\right)\right)/2 - \left(EV_{x \sim p_{gen_i}} \log\left(1 - dis_i(x, s)\right)\right)/2 \tag{13}$$

Along the lines of L_{gen_i} , the first and second terms of L_{dis_i} represents the visual realism loss while the third term represents the semantic consistency loss for the discriminator.

C. Implementation Details

For the translation module, a transformer model was trained. Best results were achieved when the training of the transformer was done for 50 epochs with an Embedding Dimension of 512 and Vocabulary Size of 10004 words. The number of Encoding and Decoding layers (*N*) chosen were both 6.

For the image generation module, a pre-trained instance of DM-GAN was used which was trained for 200 epochs on the MS-COCO dataset. A variable learning rate was used for both the generator and the discriminator with a batch size of 16. The Embedding Dimension was taken to be 256 with each image having 5 different captions.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:33:40 UTC from IEEE Xplore. Restrictions apply.

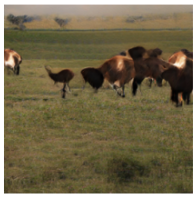
Original Hindi Description	*पेड़ में हरे पत्ते*	*पिज़्ज़ा बड़ा है*	*मैदान में गायें*	*उथले पानी में चट्टानों*
Translated English Description	*green leaves in tree*	*the pizza is large*	*cows in the field*	*rocks on the water*
Original Image (Genome Dataset)				
Generated Image				

Fig. 2. Images generated by HT-GAN for sample Hindi text descriptions

IV. Quantitative Analysis

A. Datasets used for Experimentation

This paper uses the Hindi Visual Genome Dataset [25] which is a multimodal dataset consisting of image-text description pairs. The text descriptions are available in both Hindi and English languages and hence the dataset provides an optimal platform to analyze the models in the field of English-Hindi multimodal research. The model consists of 29K text-image pairs in the training dataset and 1.6K pairs in the test dataset. Our language transformer was trained on this dataset and the image generated through the translated sentences were analyzed on the test set. The DM-GAN model was pretrained on the MS-COCO [26] dataset which is a benchmark dataset in the area of T2I. This is a generic dataset which contains 80 distinct categories of images and hence helps to improve the generalization capabilities of the GAN. The COCO dataset contains in total around 150K training and test samples with each image having 5 different captions which helps in interpolation and contributes to a smooth output distribution.

B. Evaluation metrics

1) Bilingual Evaluation Understudy Score: BLEU Score [27] is a well known metric used to judge the performance of machine translation models. It compares the text generated from the translation model with the original reference translation. The comparison is based on calculating the number of n-grams that are present in both the translated sentence as well as the reference sentence. BLEU score lies in the range [0.0, 1.0] with a score of 1.0 indicating a perfect match between the translated and original sentences.For our model, due to the small lengths of the captions, comparison is done upto 4-grams with equal weights of 0.25 given to all the 4 n-gram comparisons.

TABLE I
Results obtained on Evaluation Metrics

Evaluation Metric	**Value**
Inception Score (DM-GAN - English T2I)	32.43 ± 0.58
Inception Score (HTGAN - Hindi T2I)	19.31 ± 1.43
BLEU Score (HTGAN - Translation Module)	0.653

2) Inception Score: Inception Score [28] is a well-recognized metric used to evaluate the performance of GANs. Inception score evaluates the quality of model generated images. This is done using the Inception V3 model which is a widely used model for the task of image recognition trained on the ImageNet [29] dataset. The Inception model [30] evaluates the generated images by calculating the conditional probabilities of the generated images belonging to different classes. The probabilities hence calculated are used to account for both - how visually realistic the image is and how much diversity is present in the generated image. Higher Inception Score is an indicator of better performance. There is also a drawback of this method - Since the Inception model is pre-trained on the ImageNet dataset, it is limited by the fact that it can classify images to only the categories that are present in the ImageNet dataset.

C. Results

For our Translation module, the BLEU score obtained was **0.653** which indicates an excellent ability of the model to capture the contextual nuances involved in the low-resource Hindi

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:33:40 UTC from IEEE Xplore. Restrictions apply.

to English translation task. This performance can be attributed to the multi-headed attention mechanism and encoder-decoder structure used in transformers which refines their ability to focus on the underlying context better than most other models leading to a better ability to map the context from one language to another.

The Inception Score obtained on the images generated by the Image Generation Module conditioned on the English sentences generated after translating the input Hindi sentences was **19.31 ± 1.43** while for the English T2I task, DM-GAN gives an Inception Score of **32.43 ± 0.58**. This difference in the value of Inception Score can be attributed to the higher resource availability in the field of the English T2I task in comparison with the novel Hindi T2I task. This shows that the images generated showed high visual realism and diversity. The use of the Dynamic Memory Key-Value stores in the Image Generation Module allow the GAN stages to iteratively enhance on the images from the previous stage by fetching the suitable image features which need to be refined in that stage where in the response image features are conditioned on both the textual inputs as well as the image features from the previous stage which has led to a significantly competent image generation performance in such a low resource setting.

V. Concluding Remarks

The architecture proposed in the paper comprising the Translation and Image Generation Modules achieves significantly satisfactory results on the novel Hindi text-to-image synthesis task. The translation module generates high quality Hindi to English translations owing to the various attentional mechanisms employed by it. In the Image Generation Module, the SotA DM-GAN generates lifelike images with high visual realism and semantic consistency based on the translated inputs provided. In our current work, the two modules have been separately trained. End-to-end training of these two modules is yet to be explored and can form the basis for future propositions in this novel task.

With constant improvements in the field of T2I as well as Natural Language Processing and development of better performance metrics, the pipeline for Hindi T2I can be further streamlined and optimised for better performance in the future.

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[2] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative adversarial text to image synthesis." in *Proceedings of the 33 rd International Conference on Machine Learning*, 2016, pp. 1060–1069.

[3] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *Proceedings of the International Conference on Learning Representations*, 2014.

[4] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," in *Proceedings of the 31st International Conference on International Conference on Machine Learning-Volume 32*, 2014, pp. II–1278.

[5] K. Gregor, I. Danihelka, A. Graves, D. Rezende, and D. Wierstra, "Draw: A recurrent neural network for image generation," in *International Conference on Machine Learning*, 2015, pp. 1462–1471.

[6] S. Reed, A. van den Oord, N. Kalchbrenner, V. Bapst, M. Botvinick, and N. De Freitas, "Generating interpretable images with controllable structure," Tech. Rep., 2016.

[7] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[8] E. L. Denton, S. Chintala, R. Fergus *et al.*, "Deep generative image models using a laplacian pyramid of adversarial networks," in *Advances in neural information processing systems*, 2015, pp. 1486–1494.

[9] E. Mansimov, E. Parisotto, J. L. Ba, and R. Salakhutdinov, "Generating images from captions with attention," in *Proceedings of the International Conference on Learning Representations*, 2016.

[10] S. Reed, Z. Akata, H. Lee, and B. Schiele, "Learning deep representations of fine-grained visual descriptions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 49–58.

[11] A. Nguyen, J. Clune, Y. Bengio, A. Dosovitskiy, and J. Yosinski, "Plug & play generative networks: Conditional iterative generation of images in latent space," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4467–4477.

[12] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas, "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 5907–5915.

[13] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. Metaxas, "Stackgan++: Realistic image synthesis with stacked generative adversarial networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 8, pp. 1947–1962, 2018.

[14] T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, and X. He, "Attngan: Fine-grained text to image generation with attentional generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1316–1324.

[15] T. Qiao, J. Zhang, D. Xu, and D. Tao, "Mirrorgan: Learning text-to-image generation by redescription," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1505–1514.

[16] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[17] Z. Yi, H. Zhang, P. Tan, and M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2849–2857.

[18] K. Yamada and K. Knight, "A syntax-based statistical translation model," in *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, 2001, pp. 523–530.

[19] D. Marcu and D. Wong, "A phrase-based, joint probability model for statistical machine translation," in *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, 2002, pp. 133–139.

[20] J. Schmidhuber, "Long short-term memory," *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, 1997.

[21] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *NIPS 2014 Workshop on Deep Learning, December 2014*, 2014.

[22] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc., 2017.

[24] M. Zhu, P. Pan, W. Chen, and Y. Yang, "Dm-gan: Dynamic memory generative adversarial networks for text-to-image synthesis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5802–5810.

[25] S. Parida, O. Bojar, and S. R. Dash, "Hindi visual genome: A dataset for multi-modal english to hindi machine translation," *Computación y Sistemas*, vol. 23, no. 4, 2019.

[26] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[27] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," in *Proceedings of the 40th*

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:33:40 UTC from IEEE Xplore. Restrictions apply.

annual meeting of the Association for Computational Linguistics, 2002, pp. 311–318.

[28] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," *Advances in neural information processing systems*, vol. 29, pp. 2234–2242, 2016.

[29] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[30] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:33:40 UTC from IEEE Xplore. Restrictions apply.

2021 2nd International Conference for Emerging Technology (INCET)
Belgaum, India. May 21-23, 2021

Implementation of a Shadow Filter using CFCC

Prakhar Prakash
Department of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
prakharsinha343@gmail.com

Yashodhan Srivastava
Department of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
yashodhansrivastava9@gmail.com

Shubham Kumar
Department of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
kumarshubham525@gmail.com

Alok Kumar Singh
Department of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
aksingh@dce.ac.in

2021 2nd International Conference for Emerging Technology (INCET) | 978-1-7281-7029-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/INCET51464.2021.9456361

***Abstract*: In this paper a new shadow filter implementation has been designed using Current Follower Current Conveyor (CFCC) block and grounded elements. Shadow filter allows for electronically tunable filter parameters by simply varying the gain of an external amplifier in the feedback path and without disturbing the core filter circuit. The shadow filter utilizes a CFCC block and three X+Copy ($X+_C$) CFCC blocks and its parameters can be independently tuned with the gain of the external amplifier. The gain of this external amplifier depends on the ratio of two grounded resistors. The workability of the proposed filter has been demonstrated in PSPICE using 180nm TSMC CMOS technology node.**

Keywords: Current mode circuits, Shadow filters, CFCC

I. INTRODUCTION

With advancements in technology, the demand for high-speed, power-efficient electronic circuits continues to rise [1]. In such a scenario, researchers are always on the lookout to develop and test new circuits that would serve the purpose. Since the development of current-mode circuits, they have continued to be the first choice of researchers, as they have many advantages over conventional voltage-mode circuits. Current-mode circuits offer better bandwidth performance, greater linearity, greater dynamic range, reduced circuitry, lesser chip area and lower power consumption. Another remarkable advantage that current mode circuits offer is that, summing operation can be realised simply by a circuit node, eliminating the need for a dedicated summing circuit, as in the case of voltage-mode circuits. One of the most popular current-mode active elements is the current conveyor block, first proposed by Sedra and Smith in 1968 [2]. The original current conveyor circuit has been modified since and used in a variety of situations. Cascading current follower and current conveyor blocks, we arrive at the Current Follower Current Conveyor (CFCC) block, a five terminal device that has current mirroring in its first and third stages and voltage mirroring at the second stage [1]. This property of the CFCC block allows us to design filter circuits for any required frequency response with the help of only grounded elements [3-5].

Shadow filter circuits find applications in the field of telecommunications wherein additional trimming is required for the filter parameters to adjust for the non-ideal effects without disturbing the original filter circuit. They are also used in radio frequency (RF) transmitters and receivers. They offer a great deal of flexibility as we can vary any one of the filter parameters orthogonally, by simply varying the gain of the external amplifier [6-8]. The external amplifier should be designed in such a manner that the gain depends only on the variations of grounded elements.

II. CFCC BLOCK

The input current at port P is copied to port Z with its direction reversed w.r.t. the input current. The potential present at port Z appears at the port I. The current present at port I is then copied into the output ports X+ and X- [3-5]. The current direction at X+ is the same as that of current at port I but has a reversed sense of direction at port X- with respect to current at port I. The Current Follower Current Conveyor (CFCC) block has been created by cascading current follower and current conveyor blocks as shown in Fig. 2. with the output of the current follower circuit being attached to the input of the current conveyor block [1]. The resulting block has both the output and input in current mode and can be used in many configurations owing to its flexible nature of voltage and current mirroring at its ports.

TABLE I. MOSFET W/L RATIOS FOR CFCC BLOCK WITH ADDITIONAL OUTPUT PORT $X+_C$.

MOSFET	W/L ratio
M1, M2, M29, M30	25/0.25
M3, M4, M31, M32	50/0.25
M5-M28	2.5/0.25
M33-M44	2.5/0.25

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:33:25 UTC from IEEE Xplore. Restrictions apply.

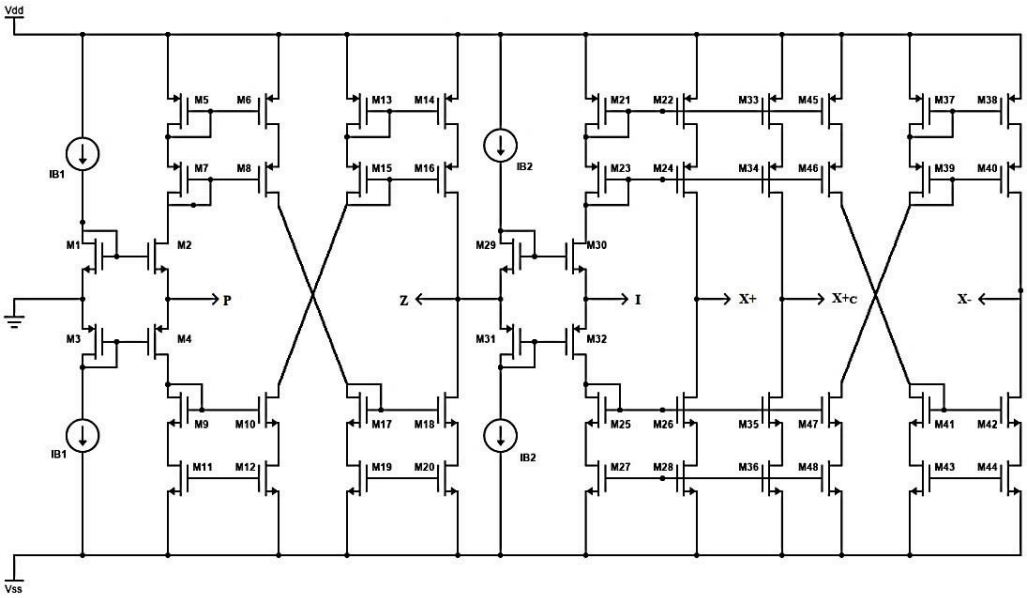


Fig. 1. MOS layout for CFCC block with additional output port $X+_C$.

Based upon the requirements, an additional output port $X+_C$ was added which generates a copy of the current at the X+ terminal. The introduction of this additional port helped in realizing the required filter using a minimum number of MOSFETs, thereby reducing chip area and power consumption.

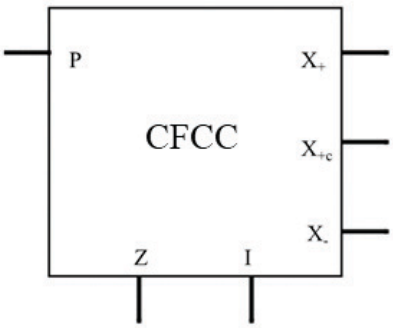


Fig. 2. CFCC Block with additional $X+_C$ port

The port relations for the standard CFCC block have been shown in the matrix:

$$\begin{bmatrix} V_p \\ I_z \\ V_i \\ I_{x+} \\ I_{x-} \\ I_{zc} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} I_p \\ V_z \\ I_i \\ V_{x+} \\ V_{x-} \\ V_{zc} \end{bmatrix} \qquad (I)$$

III. KHN FILTER USING CFCC

KHN filter can generate lowpass, high pass and bandpass output simultaneously from a single input [16]. It is a multi-feedback filter circuit. Using the CFCC with $X+_C$ port along with the conventional CFCC block, we have designed a KHN filter shown in Fig.3. The values of passive elements R1, R2, R3, R4, C1 and C2 have been calculated for a response with central frequency of 1MHz.

The relationship between output Currents I_{HP}. I_{LP}. I_{BP} and the input current I_{in} are given by the equations below:

$$I_{HP} = \left[\frac{As^2}{s^2 + \frac{A}{R_1C_1}s + \frac{A}{R_1R_2C_1C_2}} \right] I_{in} \qquad (1a)$$

$$I_{BP} = \left[\frac{\frac{A}{R_1C_1}s}{s^2 + \frac{A}{R_1C_1}s + \frac{A}{R_1R_2C_1C_2}} \right] I_{in} \qquad (1b)$$

$$I_{LP} = \left[\frac{\frac{A}{R_1R_2C_1C_2}}{s^2 + \frac{A}{R_1C_1}s + \frac{A}{R_1R_2C_1C_2}} \right] I_{in} \qquad (1c)$$

where $A = \frac{R_3}{R_4}$

IV. SHADOW FILTER USING CFCC

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:33:25 UTC from IEEE Xplore. Restrictions apply.

Introduced by Y. Lakys and A. Fabre, shadow filter is a second-order filter with independently electronically-tunable filter parameters [6]. In their paper, they demonstrated how the characteristic frequency of a second order filter could be independently tuned by introducing an external amplifier in the positive feedback path of the filter circuit [6-8].

By introducing a gain-tuneable amplifier in the feedback path as shown below in Fig. 4, we are able to create a filter whose center-frequency (f_0) and quality factor (Q) can be tuned electronically with the variation of only two grounded resistor elements [9-11]. The internal structure of the filter circuit remains the same and the bandwidth also remains unaffected throughout the operation [12-15].

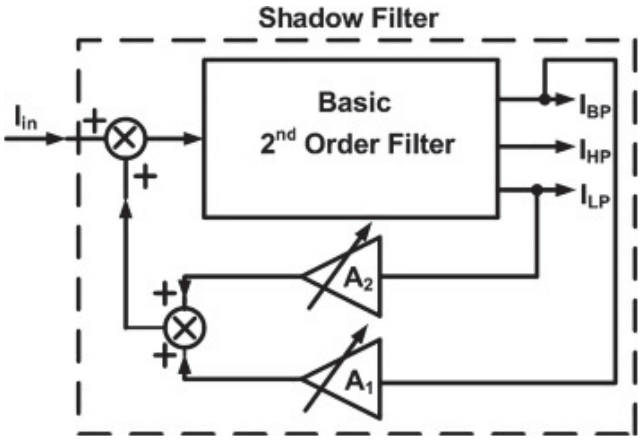


Fig. 3. Second Order Shadow Filter

Modified center-frequency for shadow filter,

$$\omega_0' = \sqrt{\frac{A(1-A')}{R_1R_2C_1C_2}} = \omega_0\sqrt{1-A'}$$

Modified quality-factor for shadow filter,

$$Q' = \frac{\omega_0\sqrt{1-A'}}{\frac{A}{R_1C_1}} = Q\sqrt{1-A'}$$

Bandwidth remains constant as before,

$$\Delta\omega' = \frac{\omega_0'}{Q'} = \frac{A}{R_1C_1} = \Delta\omega$$

Where $A = \frac{R_3}{R_4}$,

$$\omega_0 = \sqrt{\frac{A}{R_1R_2C_1C_2}},$$

and $Q = \sqrt{\frac{R_1C_1}{AR_2C_2}}$

The gain of the external amplifier in the feedback path dependent on the ratio of grounded resistances, $A' = \frac{R_5}{R_6}$.

V. SIMULATION RESULTS

The implementation of the shadow filter circuit in Fig. 3 has been corroborated with PSpice simulations. The results obtained for KHN filter with 1MHz of chosen center-frequency have been shown below in fig. 5(a). The results obtained for Shadow Filter by varying the feedback gain, A' have been also shown in Fig. 5(b) - 4(d).

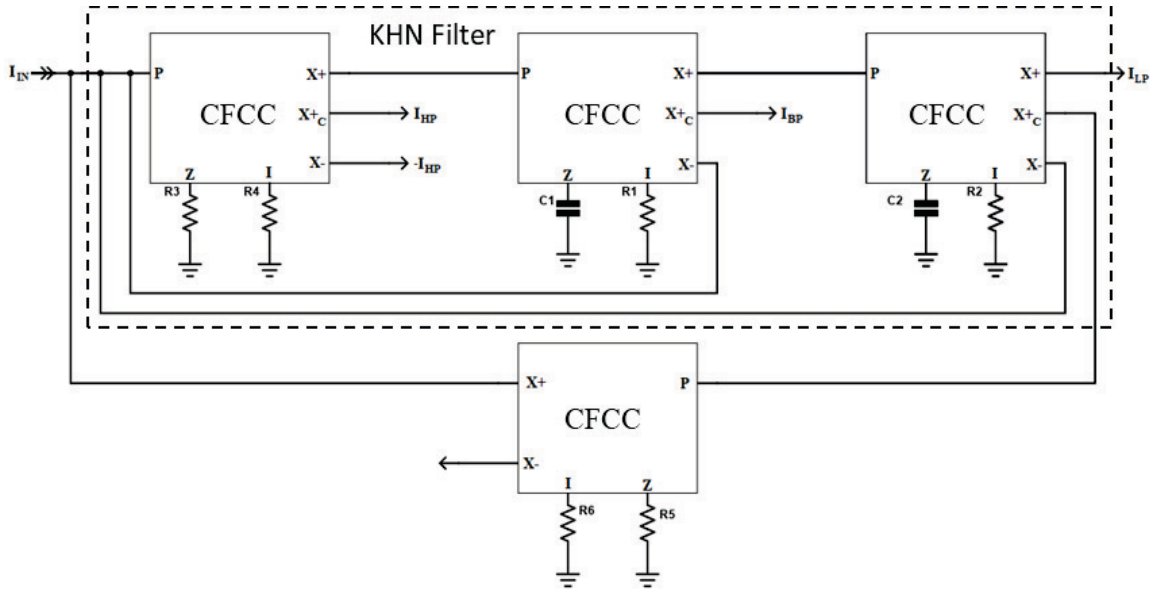


Fig. 1. Shadow Filter Using CFCC $X+_C$ and CFCC blocks

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:33:25 UTC from IEEE Xplore. Restrictions apply.

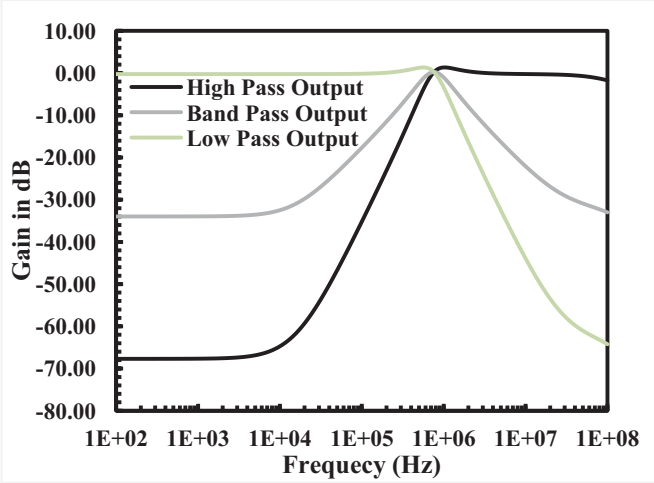


Fig. 5. (a) KHN Filter Response

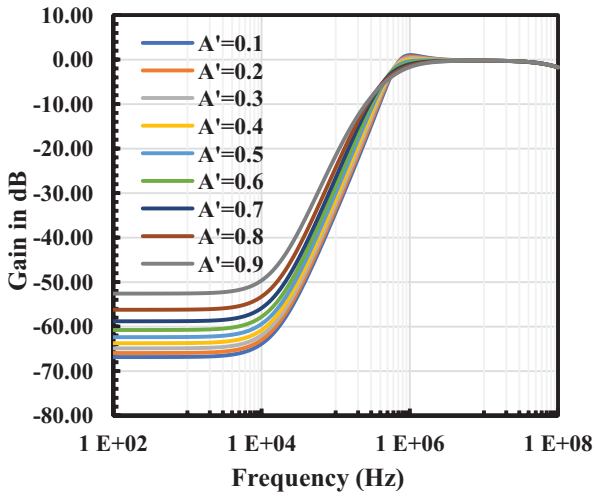


Fig. 5. (b) High Pass Shadow Filter Response

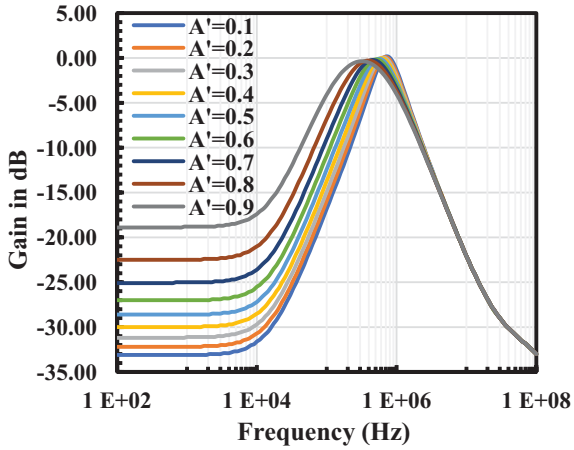


Fig. 5. (c) Band Pass Shadow Filter Response

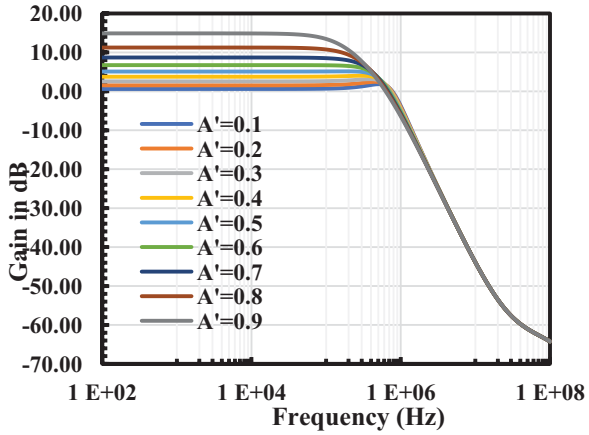


Fig. 5. (d) Low Pass Shadow Filter Response

VI. CONCLUSION

CFCC block was simulated in OrCAD PSpice software using 180nm technology model file. Shadow filter was then implemented using four CFCC blocks and grounded elements only. The gain of amplifier in the feedback path was made variable with the help of two grounded resistors and subsequent variation in center frequency and quality factor has been shown. The shadow filter was simulated for multiple values of gain of the external amplifier and the results obtained for the same have been incorporated.

REFERENCES

[1] D. Biolek, R. Senani, V. Biolkova and Z. Kolka, “Active Elements for Analog Signal Processing: Classification, Review, and New Proposals,” Radioengineering. 17, 2008.

[2] A. Sedra and K. Smith, “A second-generation current-conveyor and its applications,” IEEE Trans., vol. CT-17, pp. 132-134, 1970.

[3] A. K. Singh, P. Kumar and R. Senani, “Electronically tunable grounded/floating inductance simulators using Z-copy CFCCC,” Turkish Journal of Electrical Engineering & Computer Sciences. 26, pp. 1041-1055, 2018.

[4] A. K. Singh, P. Kumar and R. Senani, “New Grounded Immittance Simulators Employing A Single CFCC,” The Journal of Engineering, 2017.

[5] A. K. Singh, P. Kumar and R. Senani, “Current-mode Quadrature Oscillator Using CFCC,” American Journal of Electrical and Electronic Engineering, Vol. 6, No. 3, 2018, pp. 85-89.

[6] Y. Lakys and A. Fabre, “Shadow filters - new family of second-order filters,” Electronics Letters 46, 2010, pp. 276-277.

[7] Y. Lakys and A. Fabre, “Shadow filters generalization to nth-class,” Electron. Lett., 2010, 46, pp. 985–986.

[8] S. C. Roy, “Shadow Filters - A New Family of Electronically Tunable Filters,” IETE Journal of Education. 51, pp. 75-78, 2014.

[9] R. Anurag, R. Pandey, N. Pandey, M. Singh and M. Jain, "OTRA based shadow filters," 2015 Annual IEEE India Conference (INDICON), New Delhi, India, pp. 1-4, 2015.

[10] S. Buakaew, W. Narksarp and C. Wongtaychatham, "Fully Active and Minimal Shadow Bandpass Filter," 2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST), Phuket, Thailand, 2018, pp. 1-4.

[11] M. Abuelma'atti and N. Almutairi, “New CFOA-based shadow banpass filter,” pp. 1-3, 2016.

[12] F. Khateb, W. Jaikla, T. Kulej, M. Kumngern and D. Kubánek, “Shadow filters based on DDCC,” IET Circuits Devices Syst., 11, pp. 631-637, 2017.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:33:25 UTC from IEEE Xplore. Restrictions apply.

[13] P. Moonmuang, T. Pukkalanun, and W. Tangsrirat, “Voltage Differencing Gain Amplifier-Based Shadow Filter: A Comparison Study,” pp. 1-4, 2020.

[14] S. Buakaew, W. Narksarp and C. Wongtaychatham, "High Quality-Factor Shadow Bandpass Filters with Orthogonality to the Characteristic Frequency," 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Phuket, Thailand, 2020, pp. 372-375.

[15] A. Upadhyay, "Shadow Filter for Orthogonal Modification of Center Frequency and Bandwidth with Constant Q Factor," 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, Cambridge, UK, 2014, pp. 591-594.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:33:25 UTC from IEEE Xplore. Restrictions apply.

arXiv:2107.01516v1 [cs.IR] 4 Jul 2021

Improved Representation Learning for Session-based Recommendation

Sai Mitheran[1], Abhinav Java[2], Surya Kant Sahu[3], and Arshad Shaikh[3]

[1] National Institute of Technology, Tiruchirappalli
[2] Delhi Technological University
[3] The Learning Machines

Abstract. Session-based recommendation systems suggest relevant items to users by modeling user behavior and preferences using short-term anonymous sessions. Existing methods leverage Graph Neural Networks (GNNs) that propagate and aggregate information from neighboring nodes i.e., local message passing. Such graph-based architectures have representational limits, as a single sub-graph is susceptible to overfit the sequential dependencies instead of accounting for complex transitions between items in different sessions. We propose using a Transformer in combination with a target attentive GNN, which allows richer Representation Learning. Our experimental results and ablation show that our proposed method outperforms the existing methods on real-world benchmark datasets.

Keywords: Session-based recommendation · Graph neural networks · Transformers

1 Introduction

Traditional recommendation systems use user-item interactions across multiple sessions to model user preferences. However, in session-based recommendation (SBR) systems, the users are anonymous; hence inter-session data cannot be used. The goal here is to predict the items with which the user is likely to interact, given previous item interactions within a single session. In general, the number of user-item interactions is limited, and as a result, modeling user intent is a challenging task. Nevertheless, session-based recommendation is gaining momentum due to increasing privacy concerns. Recent advancements in session-based recommendation systems have focused on modeling the user-item interaction as a directed graph and hence leverages graph-based architectures and related multi-level feature extraction techniques.

In this work, we present TAGNN++ that incorporates Transformers as universal function approximators to enhance capturing complex transitions that address the limitations of GNNs in learning rich representations. We model item interactions using a GNN and both global and local user interaction with a Transformer. We show that our model outperforms the existing state-of-the-art techniques on the Diginetica benchmark[4] and for Mean Reciprocal Ranking

[4] http://cikm2016.cs.iupui.edu/cikm-cup

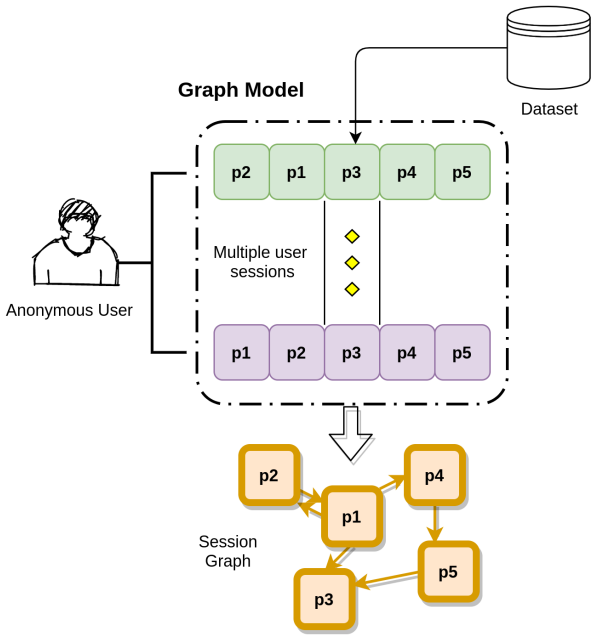


Fig. 1. Session-based Recommendation Systems

(MRR) on the Yoochoose benchmark[5]. We also explore Adaptive Gradient Clipping [1] for Transformer-based architectures specific to our task and present an ablation study to analyze our model performance.

2 Related works

2.1 Deep Learning based SBRs

GNNs offer an intuitive approach to session-based recommendation since each session can be mapped into a graph's chain. Each node of the graph represents an item, and each edge represents the interactions. The natural compatibility between data modeled in such a manner and GNNs allow this method to perform well and was introduced in [15]. Despite the convenient representation of sessions offered by GNNs, it lacks the ability to model long-range dependencies and complicated interactions as substantiated in Graph-Contextualised Self-Attention model [16]. Qiu et al. proposed using a weighted graph attention layer for focusing on the essential parts of the item embeddings [9]. Extending the idea of using attention with GNNs, Yu et al. proposed using both item embeddings using a

[5] http://2015.recsyschallenge.com/challenge.html

GNN and target embedding using an attention mechanism to achieve significant performance gain [17]. However, a GNN applied on a single sub-graph is susceptible to overfit on sequential dependencies instead of accounting for complex transitions between items in different sessions. Using a dual-channel GNN capable of complex item transition modeling addresses this issue [20].

2.2 Use of Adaptive Gradient Clipping

Clipping the gradient is a commonly used approach to improving gradient descent [19], but manual selection of the clipping threshold increases the number of hyperparameters. However, the hyperparameter for gradient clipping needs to be tuned carefully, as it is susceptible to the loss function and architecture [11]. The chosen threshold is vital as if it is set too high, then the gradient norm will always be smaller than that, and clipping is never applied. If too low, then the network's step size may be too small and cause convergence issues leading to unstable learning. We can clip gradients based on the unit-wise ratio of gradient norms to parameter norms [1]. This helps ensure stable training across different batch sizes, allowing larger learning rates to ensure quick convergence, and is effective to tackle poorly conditioned loss landscapes. We incorporate this to our model and verify that performance is marginally increased for Transformer architectures specific to our task.

3 Method

3.1 Problem Formulation and Preliminaries

The SBR problem contains three types of entities : Users, Items, and User-item interactions. Here, we formally introduce the SBR problem and associated notations. Let $S = [S_0, S_1, S_2, ..., S_{N-1}]$ be the set of N sessions and $V = [V_0, V_1, V_2, ..., V_{M-1}]$ be the set of M unique items in the dataset. Let $S_k = [x_0, x_1, x_2, ..., x_t]$ be the item sequence within a session S_k, where $x_t \in V$ is the item clicked by a user at timestep $t = 0, 1, 2....$ The goal is to estimate parameter set θ such that $p(x_{t+1}|x_0, x_1, x_2, ..., x_t; \theta)$ is the maximum-likelihood estimator of x_{t+1}, where $k \in [0, N-1]$.

TAGNN++ leverages two key components, namely, Graph Neural Networks and Self-Attention. We use the same background strategy for constructing the session graph from the given data as proposed in [17] and [15]. Each session is stored in memory as an adjacency matrix that represents a directed graph. In this graph $\mathbb{G}$, the i^{th} node in the node set $\mathbb{V}$ represents an item such that $n_{s,i} \in \mathbb{V}$. Each edge connecting nodes i and j denotes subsequent item selections by the user in the given session s such that $(n_{s,i}, n_{s,j}) \in \mathbb{E}$ [17]. Lastly, each node has associated with it, an incoming weight and an outgoing weight. We effectively capture item transitions and connections using a gated GNN [7], by learning both item and session embeddings as in [17].

3.2 Proposed method - TAGNN++

We observe that a simple attention model is unable to capture both local and global context [12], therefore pointing out the need for a more robust way to represent sequences. Furthermore, GNNs build a representation of data by message passing or neighborhood aggregation, again using only local information. Garg et al. show that even simple graph structures are indistinguishable by GNNs relying only on local information, making it hard for a GNN to compute several graph properties [4]. Additionally, they perform a thorough study of generalization bounds for message passing illustrates the need for better function approximators.

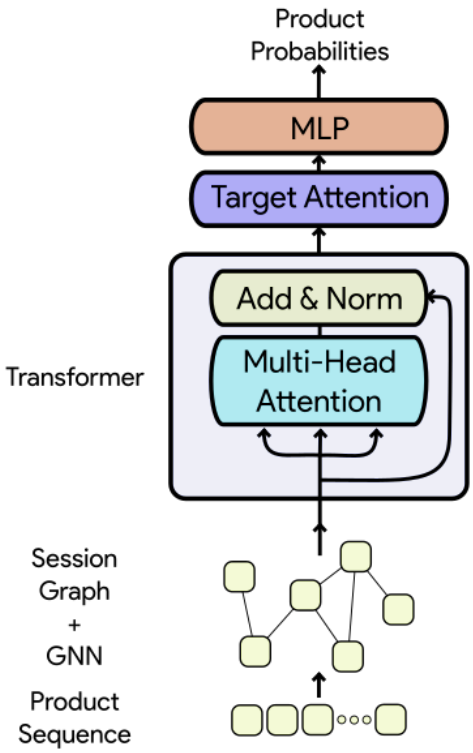


Fig. 2. Proposed Architecture : TAGNN++

To that end, we propose a multi-headed Transformer-based design[6] for target-aware predictions. This allows us to leverage both GNNs and powerful attention models; in essence, item interaction in the form of a graph and a user's long

[6] The code is available at `https://github.com/The-Learning-Machines/SBR`

and short-term interaction through the Transformer. We hypothesize that our design is a *superior function approximator* for the task of next-item prediction in session-based recommendation and back it up with empirical evidence and careful ablative study. Recently, Transformers have been successfully used in Natural Language Processing and Computer Vision by parallelizing the self-attention [12] mechanism over multiple heads achieving state of the art results. [18] also gives a compact support based proof of why Transformers have high representation capacity and are *universal approximators of permutation equivariant sequence to sequence functions.*

At the l^{th} layer the following update takes place in a single head:

$$h_i^{l+1} = \sum_{\forall j \in S} \sigma(\frac{Q^l h_i^l \cdot (K^l h_j^l)^T}{\sqrt{d_k}}) V^l \tag{1}$$

where h_j^l is the hidden state of the j^{th} item in sequence S at layer l such that $h_j^l \forall j \in S$. Q^l, K^l, V^l are Queries, Keys and Values respectively, and σ denotes the softmax operation. Q^l, K^l, V^l are learnable parameters that are updated parallelly as opposed to sequentially and d_k is a hyperparameter indicating the dimension of the linear layer in the Transformer. In our approach, we allow the Transformer to concatenate several representations or transformations of the interaction between target and sequence hidden state. This is done using multiple such attention heads and concatenating the outputs before normalizing across the layer and passing to a Feed-forward module. At the feature level, it is possible that the values in the attention matrix cover a large domain, which is not desirable since the network finds it hard to learn the optimal parameter across different scales quickly. This is where Layer Normalization comes into play by normalizing across the feature space.

Finally, the token-wise Feed-forward Layer (FFL) transforms the normalized context vector to the output sequence. The composition of these FFLs implicitly implement a scalar quantization map such that each input is mapped to the output [18]. We stress that the proposed augmentations to the TAGNN architecture make it more robust to different kinds of data streams. The various operations in the Transformer that include Self-attention, Layer Norm, and the Feed-forward Layer with skip connections play a vital role in enabling Transformers to be universal approximators of sequence-to-sequence functions.

Dataset	**Clicks**	**Train Sessions**	**Test Sessions**	**Total Items**	**Avg. Length**
Diginetica	982961	719470	68977	43074	5.13
Yoochoose 1/64	557248	375043	55898	16339	6.11

Table 1. Dataset Stats

4 Setup and Experimental Results

Evaluation Metrics: We use two metrics from previous studies, i.e., Hit Rate@N, and MRR@N, where we choose N = 20. Hit Rate calculates the number of "hits" in an N-sized list of ranked items, where a "hit" refers to something that the user has clicked on, purchased, or "saved for later", based on context. Mean reciprocal rank (MRR) is a metric used to judge a system where the order/positions of the retrieved items are important.

Datasets: For testing our hypothesis, we use the widely used real-world datasets, Yoochoose[2] and Diginetica[1]. Table **??** provides information on the dataset contents. We use only the recent 1/64 fractions of the Yoochoose dataset, denoted as Yoochoose 1/64, which comprises various sessions that specify the clicks by a given user. It is a collection of records in a file containing a Session ID, Timestamp, Item ID, and Category. This dataset was released at RecSys 2015. Diginetica contains anonymized search and browsing logs, product data, anonymized transactions, and an extensive collection of product images. This dataset was released at CIKM 2016.

Preprocessing: For simplicity, we apply the same preprocessing as [17], [15], [8]. We drop all unit length sessions and remove items that appear less than five times for both datasets as same as previous studies. For generating training and test sets, sessions of last days are used as the test set for Yoochoose, and sessions of last weeks as the test set for Diginetica. For an existing session, we generate a series of input session sequences and corresponding labels. We filter out items from the test set which do not appear in the training set.

Table 2. Comparison of several baseline methods with the proposed method.

Method	**Yoochoose 1/64**		**Diginetica**	
	HR@20	MRR@20	HR@20	MRR@20
GRU4REC [5]	60.64	22.89	29.45	8.33
NARM [6]	68.32	28.63	49.70	16.17
STAMP [8]	68.74	29.67	45.64	14.32
RepeatNet [10]	70.71	31.03	47.79	17.66
CSRM [13]	71.45	30.36	50.55	16.38
SR-GNN [15]	70.87	30.94	50.73	17.59
GC-SAN [16]	70.66	30.04	51.70	17.61
TAGNN [17]	71.02	31.12	51.31	18.03
LESSR [2]	70.64	30.97	51.71	18.15
DGTN [20]	71.18	31.35	53.05	18.07
Transformer	**72.15**	30.79	51.31	17.14
TAGNN++	70.74	**32.87**	**56.25**	**19.86**

Hyperparameters: We retain most hyperparameter settings from previous baselines to display the advantage of learning better representations using Trans-

formers. We keep 10% of our datasets for validation and use a batch size of 50, for 15 epochs. We use the Adam [3] optimizer with an initial learning rate of 10^{-4}, with momentum parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and decay it by a factor of 0.1 every 3 epochs. Additionally, we set the $L2$ penalty to 10^{-6}. For Multi-Head Attention in the Transformer, we set the number of heads as $2, 8$ and dropout of 0.1, with the embedding dimension as $100, 120$ for Yoochoose and Diginetica respectively.

Baselines: We compare our method with baseline GNN and Attention-based methods for session-based recommendation, as shown in Table 2. The RNN-based methods [6] [13] [5] and further [10] that takes repeat consumption into account were then outperformed by graph-based methods [15], those involving the notion of attention [16] [17] and further by [20], which combines both the target session and its neighbor sessions into a single session graph.

Table 3. Ablation Study

Architecture	**Yoochoose 1/64**		**Diginetica**	
	HR@20	MRR@20	HR@20	MRR@20
TAGNN++	70.74	**32.87**	**56.25**	**19.86**
- AGC	70.04	32.53	55.46	19.51
- GNN	**72.15**	30.79	51.31	17.14
- PE	70.37	32.08	54.82	19.18
- Transformer	71.02	31.12	51.31	18.03

Inferences: We can infer from the values of HR@20 and MRR@20 that we outperform the previous state-of-the-art [20] on the given benchmarks. Hence, it is evident that Deep learning-based methods are more capable of learning the user-item interaction function by capturing complex data distributions and transitions in the dataset. Incorporating Transformers to overcome the representational limitations in a GNN improves the performance of our recommendation system. The improvement observed over [20] verifies that learning better representations as proposed is helpful to model complex patterns, as shown in Fig. 3. To verify the efficacy of our architecture, we also perform an ablation study as shown in Table 3. It can be observed that the performance is only marginally affected by the removal of secondary techniques such as AGC (Adaptive Gradient Clipping) and PE (Positional Embedding). It is also evident that the removal of the Transformer decreases the overall performance, verifying that GNNs fail at learning accurate representations in some cases.

Fig. 3 indicates better performance on Yoochoose than Diginetica, because of a greater average session length in the former - as more historical data would enhance the Transformer's attention mechanism to leverage the global context. Both datasets consist of short, medium, and long sessions [14]. When the user wishes to build a session-based recommender consisting of long sessions with better rank/order as the main objective (tweets, webpages, music), we show that

our model outperforms existing methods (Yoochoose - MRR). For short/medium length sessions, where both order and the number of desired items in the top N of the ranked list are important (shopping items, movies/videos), our model outperforms the existing methods by a considerable margin (Diginetica - HR and MRR).

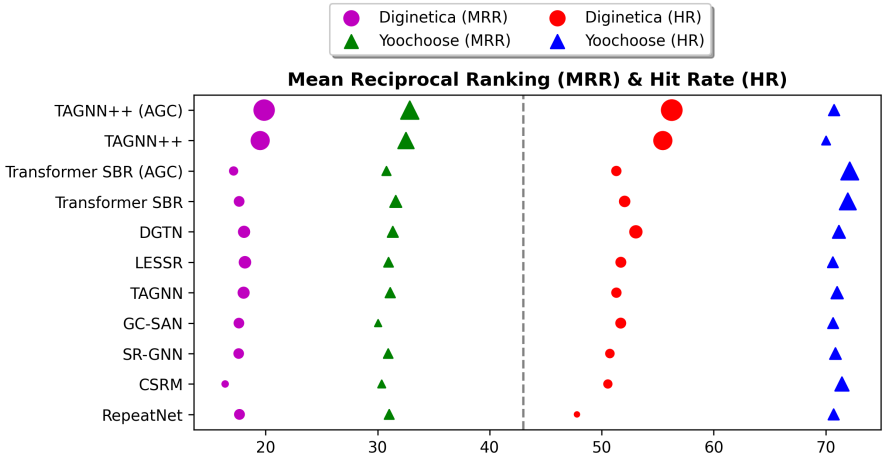


Fig. 3. Scatter plot comparison of models on the evaluation metrics

5 Conclusion

This paper proposes an alternative method for representation learning in session-based recommendation models to overcome the limitations posed by message-passing GNNs. We perform intelligent feature extraction using Transformers with target attentive GNNs. We leverage Multi-head Attention to capture both local and global context. Our method outperforms the previous state-of-the-art on the aforementioned benchmarks. We show that our methods are suitable for rank-based retrieval in long sessions (Yoochoose - MRR). Our results also conclusively show our methods' superiority over existing methods for short/medium length sessions (Diginetica - HR and MRR). Motivating research in this direction would enable making informative and practical choices in streaming, business operations, and E-Commerce.

References

1. Brock, A., De, S., Smith, S.L., Simonyan, K.: High-performance large-scale image recognition without normalization. arXiv preprint arXiv:2102.06171 (2021)

2. Chen, T., Wong, R.C.W.: Handling information loss of graph neural networks for session-based recommendation. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 1172–1180 (2020)
3. Da, K.: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
4. Garg, V., Jegelka, S., Jaakkola, T.: Generalization and representational limits of graph neural networks. In: International Conference on Machine Learning. pp. 3419–3430. PMLR (2020)
5. Hidasi, B., Karatzoglou, A., Baltrunas, L., Tikk, D.: Session-based recommendations with recurrent neural networks. arXiv preprint arXiv:1511.06939 (2015)
6. Li, J., Ren, P., Chen, Z., Ren, Z., Lian, T., Ma, J.: Neural attentive session-based recommendation. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. pp. 1419–1428 (2017)
7. Li, Y., Tarlow, D., Brockschmidt, M., Zemel, R.: Gated graph sequence neural networks. arXiv preprint arXiv:1511.05493 (2015)
8. Liu, Q., Zeng, Y., Mokhosi, R., Zhang, H.: Stamp: Short-term attention/memory priority model for session-based recommendation. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. p. 1831–1839. KDD '18, Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3219819.3219950, https://doi.org/10.1145/3219819.3219950
9. Qiu, R., Li, J., Huang, Z., YIn, H.: Rethinking the item order in session-based recommendation with graph neural networks. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management. p. 579–588. CIKM '19, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3357384.3358010, https://doi.org/10.1145/3357384.3358010
10. Ren, P., Chen, Z., Li, J., Ren, Z., Ma, J., de Rijke, M.: Repeatnet: A repeat aware neural recommendation machine for session-based recommendation. Proceedings of the AAAI Conference on Artificial Intelligence **33**(01), 4806–4813 (Jul 2019). https://doi.org/10.1609/aaai.v33i01.33014806, https://ojs.aaai.org/index.php/AAAI/article/view/4408
11. Seetharaman, P., Wichern, G., Pardo, B., Roux, J.L.: Autoclip: Adaptive gradient clipping for source separation networks. In: 2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP). pp. 1–6 (2020). https://doi.org/10.1109/MLSP49062.2020.9231926
12. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L.u., Polosukhin, I.: Attention is all you need. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc. (2017), https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
13. Wang, M., Ren, P., Mei, L., Chen, Z., Ma, J., de Rijke, M.: A collaborative session-based recommendation approach with parallel memory modules. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 345–354. SIGIR'19, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3331184.3331210, https://doi.org/10.1145/3331184.3331210
14. Wang, S., Cao, L., Wang, Y.: A survey on session-based recommender systems. ArXiv **abs/1902.04864** (2019)

15. Wu, S., Tang, Y., Zhu, Y., Wang, L., Xie, X., Tan, T.: Session-based recommendation with graph neural networks. In: Hentenryck, P.V., Zhou, Z.H. (eds.) Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence. vol. 33, pp. 346–353. AAAI Press (Jul 2019). https://doi.org/10.1609/aaai.v33i01.3301346, https://aaai.org/ojs/index.php/AAAI/article/view/3804
16. Xu, C., Zhao, P., Liu, Y., Sheng, V.S., Xu, J., Zhuang, F., Fang, J., Zhou, X.: Graph contextualized self-attention network for session-based recommendation. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19. pp. 3940–3946. International Joint Conferences on Artificial Intelligence Organization (7 2019). https://doi.org/10.24963/ijcai.2019/547, https://doi.org/10.24963/ijcai.2019/547
17. Yu, F., Zhu, Y., Liu, Q., Wu, S., Wang, L., Tan, T.: Tagnn: Target attentive graph neural networks for session-based recommendation. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 1921–1924 (2020)
18. Yun, C., Bhojanapalli, S., Rawat, A.S., Reddi, S.J., Kumar, S.: Are transformers universal approximators of sequence-to-sequence functions? arXiv preprint arXiv:1912.10077 (2019)
19. Zhang, J., He, T., Sra, S., Jadbabaie, A.: Why gradient clipping accelerates training: A theoretical justification for adaptivity. arXiv: Optimization and Control (2020)
20. Zheng, Y., Liu, S., Li, Z., Wu, S.: Dgtn: Dual-channel graph transition network for session-based recommendation. In: 2020 International Conference on Data Mining Workshops (ICDMW). pp. 236–242 (2020). https://doi.org/10.1109/ICDMW51313.2020.00041

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/352727481

Improving the Productivity of the Machining Process of a Manufacturing Company: A Six Sigma Case Study

Preprint · June 2021

CITATIONS
0

READS
47

3 authors, including:

G. Kumar
Delhi Technological University
41 PUBLICATIONS **179** CITATIONS
SEE PROFILE

M. K. Loganathan
Indian Institute of Technology Delhi
19 PUBLICATIONS **66** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Circular Economy Challenges View project

Risk assessment View project

All content following this page was uploaded by M. K. Loganathan on 25 June 2021.

The user has requested enhancement of the downloaded file.

Chapter 48

Improving the Productivity of the Machining Process of a Manufacturing Company: A Six Sigma Case Study

Vikram Narula
Department of mechanical engineering
Delhi Technological University (formerly Delhi College of Engineering), Delhi, India

Girish Kumar
Department of mechanical engineering
Delhi Technological University (formerly Delhi College of Engineering), Delhi, India

MK Loganathan
Department of mechanical engineering, Kaziranga University, Assam, India

Abstract—Micro, Small and Medium Enterprises (MSMEs) generally face problems maintaining high productivity levels in their processes due to various reasons including lack of skilled workforce and unawareness of quality tools. Low productivity in the manufacturing process results in high costs, low profits and sometimes, losses. Hence, it is crucial that productivity be increased in order to minimize the losses. In this study, the productivity of the machining process of Pole (50 XT), which is used in public address (PA) driver unit, was improved. To achieve this, DMAIC (Six Sigma) methodology was used. The process capability was increased with the help of control charts and Ishikawa diagrams. In addition to this, the process mean was shifted to better comply with specifications after surface treatment. When compared to the original process, an improvement of approximately 10 seconds in production time per piece and 15.7 percent in process productivity were achieved.

Keywords—Quality, Six Sigma, DMAIC, Productivity, Manufacturing, MSME

I. INTRODUCTION

The manufacturing sector is crucial to the economic growth of a country. Manufacturing sector in India has witnessed significant growth in the past few years. The Government of India has launched many initiatives such as 'Make in India' for the growth of this sector, recognizing the opportunities that underlie this sector for the growth of economy and the potential to generate large-scale jobs. Under this initiative, the government aims to increase the share of manufacturing sector to the country's GDP to 25% by 2025 [1]. About 31% of all MSMEs in India are of manufacturing nature and have an employment share of 32% [2]. During the years 2011-2016, MSMEs contributed about 33% towards the country's overall manufacturing Gross Value of Output [3]. In this emerging scenario, MSMEs need to improve their efficiency and productivity to be able to compete both in the local and global markets.

Productivity is a measure of the efficiency of production. It is measured as the ratio of an aggregate output to a single input. In this study, productivity has been calculated based on effective production time per good piece manufactured.

The causes of low productivity include ineffective use of resources, poor information flow, non-productive activities, low labor productivity, excessive rework, and excessive inventory [4].

This study shows the implementation of DMAIC (Define-Measure-Analyze-Improve-Control) methodology which is most often associated with Six Sigma activities. Almost all implementations of Six Sigma use this methodology for project management and completion. Quality tools such as histogram, check sheet, pareto chart, cause-and-effect diagram (Ishikawa diagram), defect concentration diagram, scatter diagram and control chart are beneficial to productivity improvement. These are often called 'The Magnificent Seven' [5]. The application of some of 'the magnificent seven' is demonstrated in this study.

There are limited studies available on improvement of productivity in the MSME sector in India. Undertaking the task of quality/productivity improvement requires resources such as time and money amongst other efforts on the organization's part. This study contributes in by providing a case study utilizing Six Sigma methodology to achieve improvement in productivity. A roadmap for improving the quality and productivity of the machining stage of the production process of a manufacturing MSME is presented in this study.

II. OVERVIEW OF SIX SIGMA

Six Sigma is a structured methodology that focuses on reducing variations, assessing defects, and enhancing the quality of products, processes, and services. Originally developed by Motorola in the 1980s, a Six Sigma process is one which targets a goal of 3.4 defects per million [6]. In 1995, Jack Welch made Six Sigma principal to his business strategy at General Electric. Six Sigma is a method that provides organizations with data-driven decision-making tools to improve the capability of their business processes. There are three aspects that Six Sigma addresses: process, defects, and variation. Process is the set of repetitive steps that are performed to achieve some result. Six Sigma implementation involves studying the process and making it as efficient as possible. Defects are the measurable features of the process or its output that do not conform to specifications. Six Sigma helps in eliminating the defects and maintaining consistency thereafter. Variation refers to any significant difference between standard measurement and actual output of the process [7]. Six Sigma helps reduce the variations by utilizing statistical tools to target a goal of 3.4 defects per million as mentioned before. This study primarily focuses on the 'variation' aspect of Six Sigma.

III. CASE STUDY

This case study is conducted on XYZ company which is an Indian MSME and a manufacturer of speaker and loudspeaker parts. The company is engaged in B2B sales.

They have been raising concerns regarding productivity of their machining process. They have observed an excessive number of pieces being reworked due to inconsistency in critical dimension. This study shows the productivity improvement project undertaken by the company. This case study follows the DMAIC methodology which comprises of 5 phases: Define, Measure, Analyze, Improve and Control. The define phase helps understand the problem by clearly pronouncing the problem as well as the goal, scope etc. [8]. The measure phase is the data collection stage, its purpose is to measure and establish the process performance baselines [9]. In the analyze phase, causes and root causes of the problem are identified. The improve phase is aimed at identifying, testing, and implementing a solution to the problem. Control phase involves tracking of improvements made to sustain gains and ensuring corrective action is taken when necessary [10].

A. Define

In the define phase, a SIPOC (Suppliers-Inputs-Process-Outputs-Customers) diagram is created to map the complete process for the manufacturing of the component 'Pole (50 XT)'. SIPOC diagram is a Six Sigma tool which is used to document business processes. It visually documents the business process from start to finish [11]. The SIPOC diagram for the process is shown in fig 1.

It was identified that the finish turning process, which generates the critical dimension, resulted in the maximum number of pieces that required reworking. Hence it was selected for productivity improvement. A drawing of the component showing the critical dimension is shown in fig 2.

The machining process is as follows: First, rough turning of the faces and diameters, keeping allowances of 0.5mm on the face and 0.3mm on the diameter is done. Then, finish turning of all dimensions is carried out. Thereafter, grooving, drilling and bore finishing with internal radius machining is done. The tools, parameters and cutting fluid used are as shown in Table 1.

Table I Tools, Parameters and Cutting Fluid

Process	Tool used	Cutting parameters	Cutting fluid
Turning	Holder: PCLNL 2525M12 Insert: CNMG 120408	Constant 1500 rpm Feed = 0.2~0.3mm/rev DOC = 0.7~1.2mm (roughing) 0.3~0.5mm (finishing)	Non-synthetic, water soluble cutting fluid used in flood condition directly on cutting edge
Grooving	Special purpose holder Insert: SOMT 060204	Constant 800 rpm Feed = 0.06mm/rev	
Drilling	U-drill: dia = 26 mm Insert: SPMG 07T308	Constant 1000 rpm Feed = 0.04mm/rev	
Boring	Boring bar: dia = 20 mm Insert: CNMG 120408	Constant 1500 rpm Feed = 0.2~0.3mm/rev	

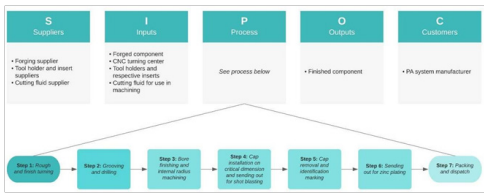


Fig. 1. SIPOC Diagram

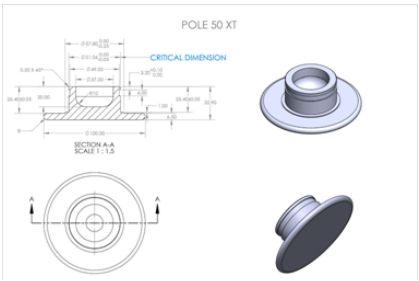


Fig. 2. Component Drawing

B. Measure

In the measure phase, the process was examined without any alteration. The CTQ (critical to quality) considered was the critical dimension as mentioned in the define phase. Variation beyond specification limits in this dimension causes variable air gap in the diaphragm coil resulting in inefficient output of the speaker. After machining, the critical dimension was measured with the help of a micrometer. For conducting the study, 28 batches of components containing 5 samples each were randomly selected over the course of 7 days. They were then used to plot control charts on a time scale according to the time of the day the samples were taken [12]. Fig. 3 shows a histogram of the total sample and figs. 4 and 5 show $\bar{X}$ and R charts of the sample data, respectively. It was observed from the control charts that the process was not in control. The process capability indices and production time per good piece were calculated as follows:

1) Process capability indices:

$$C_p = \frac{USL - LSL}{6\sigma} = 0.7143$$

$$C_{pk} = min\begin{cases}\dfrac{USL - \bar{\bar{X}}}{3\sigma} \\ \dfrac{\bar{\bar{X}} - LSL}{3\sigma}\end{cases} = 0.7143$$

2) Production time per good piece: The data for 140 samples showed that 32 pieces required rework. Each piece required 54 seconds of complete cycle time (including loading and unloading time). Each rework also required the same cycle time. Hence effective cycle time for manufacturing a good piece is:

$$\text{Time/good piece} = \frac{140\times(54)+32(54)}{140} = 66.34 \text{ seconds}$$

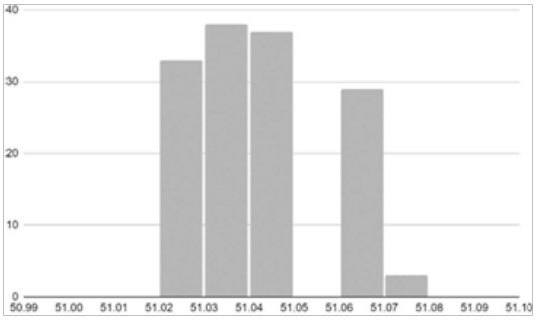


Fig. 3. Histogram

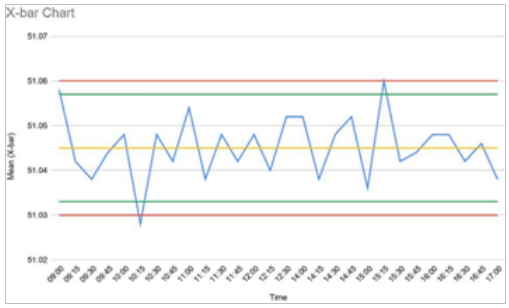


Fig. 4. X^{-} control chart (Legend: -USL and LSL; -UCL and LCL; -X^{--}; -Plot of X^{-} w.r.t. time.)

C. Analyze

In the analyze phase, a cause-and-effect diagram was made to list the possible causes of variations observed in the process. Fig. 6 shows the cause-and-effect diagram of the problem. After careful consideration, the root causes were identified as follows:

At the start of the shift, the chart shows that mean of the samples is beyond the upper control limit. The cause for this variation was identified as 'cold start' of the machine. When the machine is shut down for a long duration, all machine components tend to shrink to their minimum dimensions. After running for some duration, all the machine parts reach their operating temperature. This was the root cause for variation at the start of shift. The cause for a sudden dip in the chart was attributed to 'overcorrection' resulting from wrong inspection of the size. The root cause for sudden spike in the chart was attributed to defective forging, which on analysis showed chill points along the machined surface. Its effect is also reflected in the R-chart.

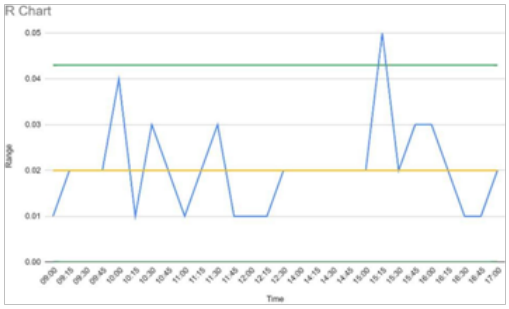


Fig. 5. Range chart (Legend: -UCL and LCL; -$\bar{R}$; -Plot of R w.r.t. time.)

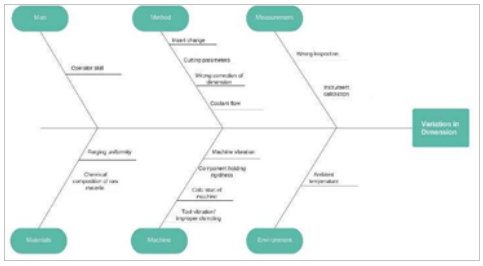


Fig. 6. Cause-and-effect diagram

D. Improve

In the improve phase, the following improvements were made: To address the problem of cold start, it was practiced to run the machine idle for at least 5-10 minutes after starting it for the first time in the shift. During this period, the machine lubrication cycle was also run manually while moving the slides to ensure that the machine attains a working temperature while getting forced lubrication of all the moving parts. Inspection using a micrometer sometimes creates a problem of wrong inspection. Instead, snap gauges were introduced which helped in easy inspection while also reducing the skill, time of inspection and the idle time. The problem of defective forging was discussed with the forging supplier. The supplier was requested to check the chemical composition of each rod before cutting them to be used for component forging and to ensure that there is no instantaneous cooling of forged components. Also, the process mean (mean dimension) was shifted towards 51.04 mm as opposed to 51.045 mm previously with the objective of keeping dimensional allowances within specification limits after zinc plating (of thickness 3-5 microns) of the component.

E. Control

In the final stage, that is the control phase, the process was re-sampled for 7 days as before and control charts were again plotted on a time scale according to the time of the day the samples were taken. Fig. 7 shows a histogram of the total sample and figs. 8 and 9 show $\bar{\bar{X}}$ and R charts of the sample data, respectively. The process capability indices and production time per good piece were calculated as follows:

1) *Process capability indices:*

$$C_p = \frac{USL - LSL}{6\sigma} = 1.3021$$

$$C_{pk} = min\begin{cases} \dfrac{USL - \bar{\bar{X}}}{3\sigma} \\ \dfrac{\bar{\bar{X}} - LSL}{3\sigma} \end{cases} = 0.9549$$

2) *Production time/good piece:* The data for 140 samples showed that 5 pieces required rework. Each piece required 54 seconds of complete cycle time (including loading and unloading time). Each rework also required the same cycle time. Hence effective cycle time for manufacturing a good piece is:

$$Time/good\ piece = \frac{140\times(54)+5(54)}{140} = 55.93\ seconds$$

3) *Process capability improvement:* Improvements of 82.29% (C_p) and 33.68% (C_{pk}) were accomplished.
4) *Productivity improvement:* The effective time required for producing a good piece went down from

66.34 seconds to 55.93 seconds. Hence productivity improvement accomplished was:

$$\frac{66.34 - 55.93}{66.34} \times 100 = 15.7\%$$

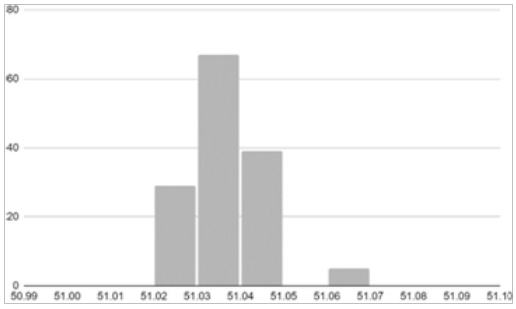


Fig. 7. Histogram

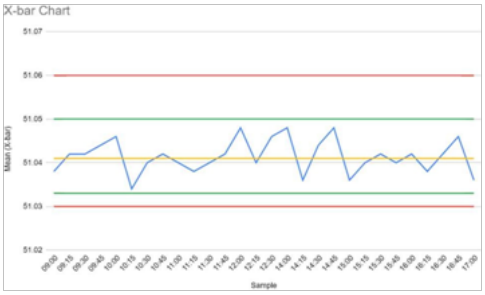


Fig. 8. $X^{=}$ control chart (Legend: -USL and LSL; -UCL and LCL; -$X^{==}$;-Plot of $X^{=}$ w.r.t. time.)

Fig. 9. Range chart (Legend: -UCL and LCL; -$R^{=}$; -Plot of R w.r.t. time.)

IV. CONCLUSION

This study presents an implementation of the DMAIC (Six Sigma) methodology to improve the productivity of the machining process of a manufacturing MSME. Baseline measures in the form of process capability indices and production time per good piece were established, control charts were prepared and shortcomings in the process were identified. Root causes were identified with the help of Ishikawa diagram. Various measures were put into place to eliminate the root causes. Improvements of 82.29% (Cp), 33.68% (Cpk) and 15.7% in productivity were accomplished.

REFERENCES

[1] Anon., 2020. Manufacturing Sector in India. [Online], Available at: https://www.ibef.org/industry/manufacturing-sector-india.aspx#:~:text=Under%20the%20Make%20in%20India,intensive%20MSME%20sectors%20by%202%25.

[2] GoI Report, 2019-2020. MSME Annual Report. [Online]. Available at: https://msme.gov.in/sites/default/files/FINAL_MSME_ENGLISH_A R_2019-20.pdf

[3] GoI Report, 2017-2018. MSME Annual Report. [Online]. Available at: https://msme.gov.in/sites/default/files/MSME-AR-2017-18-Eng.pdf

[4] Vilasini, P. P. G. N., Gamage, J. R., Kahangamage, U.P., Thibbotuwawa, N. (2012). "Low Productivity and Related Causative Factors: A Study Based on Sri Lankan Manufacturing Organisations"

[5] Montgomery, D. C. (2009) Introduction to Statistical Quality Control. 6th ed: Wiley.

[6] Gijo, E.V., Scaria, J. and Antony, J. (2011). "Application of six sigma methodology to reduce defects of a grinding process", Qual. Reliab. Engng. Int., vol. 27, pp. 1221–1234.

[7] Greg, B. (2006) Six Sigma for Small Business. Madison, Wisconsin: CWL Publishing Enterprises

[8] Webber, L and Wallace, M. (2006) Quality Control for Dummies. For Dummies. pp. 42–43. ISBN 978-0-470-06909-7.

[9] Retrieved from https://www.uce.org.uk/wp-content/uploads/2020/09/TKSBB-v1.pdf

[10] Gupta, K. and Kumar, G. (2014) "Six sigma application in warehouse for damaged bags: A case study", Proceedings of 3rd International Conference on Reliability, Infocom Technologies and Optimization, Noida, pp. 1–6.

[11] Sharma, P., Gupta, A., Malik, S. and Jha, P.C. (2019). "Quality improvement in manufacturing process through six sigma: A case study of Indian MSME firm", Yugoslav Journal of Operations Research, vol. 29(4), pp. 519–537.

[12] Lind, D. A., Marchal, W., & Wathen, S. (2011). Statistical Techniques in Business and Economics. 15th ed. Mcgraw-hill/Irwin Series Operations and Decision Sciences: McGraw-Hill/Irwin.

View publication stats

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/352733233

IoMT and DNN-enabled Drone-Assisted Covid-19 Screening and Detection Framework for Rural Areas

Article *in* IEEE Internet of Things Journal · June 2021

DOI: 10.1109/IOTM.0011.2100053

CITATIONS
0

READS
129

7 authors, including:

Vinay Chamola
Birla Institute of Technology and Science Pilani
106 PUBLICATIONS **1,879** CITATIONS

SEE PROFILE

Sainath Bitragunta
Birla Institute of Technology and Science Pilani
37 PUBLICATIONS **62** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

VANET Research View project

IoT Security View project

All content following this page was uploaded by Vinay Chamola on 25 June 2021.

The user has requested enhancement of the downloaded file.

IoMT and DNN-enabled Drone-Assisted Covid-19 Screening and Detection Framework for Rural Areas

Naren, Vinay Chamola, *Senior Member, IEEE*, Sainath Baitragunta, *Senior Member, IEEE*, Ananthakrishna Chintanpalli, Puneet Mishra, Sujan Yenuganti, Mohsen Guizani, *Fellow, IEEE*

***Abstract*—Providing rapid testing and proper treatment has become highly challenging due to the rapid and highly unpredictable spread of the coronavirus disease (COVID-19). In most developing countries, rural areas lack adequate medical facilities and medical staff for effective diagnosis and treatment. Recently, there have been several technological advancements across various engineering disciplines such as the Internet of things (IoT), unmanned aerial vehicles (UAVs) or drones, deep neural networks (DNNs), and intelligent robots. This work proposes a prototype that integrates these technologies to develop a payload deployable in a drone to help in providing rapid testing and healthcare. The proposed UAV prototype combines secure patient authentication, an automated disinfection system, and medical sensors as part of the UAV payload. It uses a DNN-model for real-time COVID-19 detection. It uses intelligent flight path planning, operational management, battery recharge planning, disinfectant refilling, and strategic location planning to quickly disseminate testing kits and essential medical services to remote locations without direct human involvement.**

***Index Terms*—COVID-19, IoMT, DNN, UAV, Contactless Diagnosis**

I. Introduction

The novel coronavirus (COVID-19) is rapidly and unpredictably spreading worldwide. Due to its highly infectious nature, healthcare and management are becoming more challenging despite significant medicine, medical facilities, and treatment improvements. We witnessed the phenomenal growth of technologies such as the Internet of Medical Things (IoMT), unmanned aerial vehicles (UAVs), or drones. Furthermore, the astonishing developments in artificial intelligence (AI), machine learning, deep neural networks (DNNs) promise to handle challenging problems, including health care. Automated technologies such as intelligent robots and drones having clinic-like payloads are undoubtedly valuable for controlling the rapidly spreading pandemic due to human contacts intelligent robots. However, the challenge still lies in integrating these technologies to develop an automated drone-centric system to provide innovative healthcare and health management. Such systems can quickly reach any required location to serve people's medical needs. There is also an urgent need for rapid automated testing, accurate diagnosis, and treatment without much human intervention.

The current study proposes a contact-less COVID-19 screening and detection named Covidrone that consists of biomedical sensors to capture COVID-19 related medical data from the users. Covidrone would have a payload consisting of a sensing module, disinfectant module, testing kits, a DNN model, and a wireless communication subsystem. These drones would land on carefully selected locations and wait for users who requested the drone's services. After authenticating the user and sanitizing the sensing equipment, smart medical sensors on the drone capture the required medical data. For COVID-19 detection and analysis, these medical data are then passed through a DNN-based model. The user will be directed for further medication or preventive measures based on his/her test report. Remote medical staff would be able to access these test reports via cloud services for further analysis. Automation of the whole process prevents human contact, thereby contributing to the combat against COVID-19. This approach can provide essential healthcare and testing facilities to elderly and physically disabled individuals. It is expected that Covidrone can significantly help in flattening COVID-19 curves, especially in rural areas where specialized medical facilities are limited.

We propose a prototype that consists of various features: (i) an effective authentication for the user to have secure testing, (ii) the integration of an automatic disinfectant system, (iii) smart medical sensors in the drone's payload, (iv) efficient and well-trained DNN model for real-time COVID-19 detection, (v) intelligent flight path planning, operational management and battery recharge planning, (vi) disinfectant refilling and strategic location planning. These features will be incorporated in Covidrone, making it highly effective in tackling COVID-19 and can help quickly disseminate testing kits and essential medical services to remote locations without direct human involvement.

II. Related Works

Integrating various next-generation information and communication technologies will help develop smart healthcare to manage the COVID-19 pandemic. These technologies include drones, the internet of things (e.g., medical things or sensors), fifth-generation cellular networks, and DNN model [1]. Traditionally, drones are used in defense-related applications, such as remotely controlled aerial missile deployers. However, in

Naren, Vinay Chamola (Corresponding author), Sainath Baitragunta, Ananthakrishna Chintanpalli, Puneet Mishra and Sujan Yenuganti are with the Department of Electrical and Electronics Engineering, BITS-Pilani, Pilani Campus, India 333031 (e-mail: f2015547@pilani.bits-pilani.ac.in, vinay.chamola@pilani.bits-pilani.ac.in, sainath.bitragunta@pilani.bits-pilani.ac.in, anantha.krishna@pilani.bits-pilani.ac.in, puneet.mishra@pilani.bits-pilani.ac.in, yenuganti.sujan@pilani.bits-pilani.ac.in).

Mohsen Guizani is with the Department of Computer Science, Qatar University, Doha 2713, Qatar (e-mail: mguizani@ieee.org).

Digital Object Identifier: XXXXXXXXXXXX

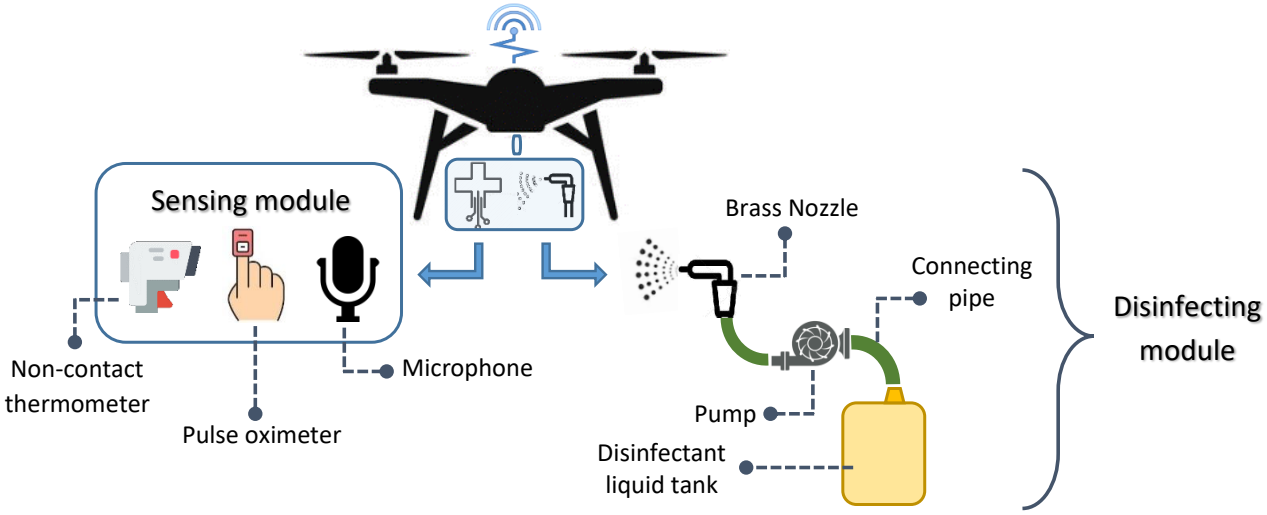


Fig. 1: Covidrone schematic showing sensing and disinfecting module details.

the past decades, drones have been used in a wide range of applications such as climate change monitoring, filming, and photography, and blood delivery. The authors of [2] propose using multiple UAV swarms to tackle different problems of COVID-19, such as social distancing, symptoms tracking, and sanitization. They propose using 6G, blockchain, and Software Defined Networks (SDN) for security and performance. The authors in [3] had utilized drones and robots to combat the COVID-19 by tracking the hotspots and delivering medical aids to the needy ones. Several COVID-19 operations such as sanitization, thermal imaging, crowd monitoring, and control are also supported in their framework. Such Systems, when deployed in considerably large regions such as a village, might require immense network resources and coordination, and thus the feasibility of its operations remains a question. In another work by [4], the authors proposed a drone-based model to provide medical service to reduce the COVID-19 spread. However, only some algorithms were proposed based on few case studies for combating COVID-19 via approaches such as social distancing. In [5], authors discuss an AI and Big Data Analytics-based prediction of COVID-19 cases and economic growth, aimed towards economy boosting.

In the existing literature on COVID-19, different research communities have used various technologies for novel COVID-19 diagnosis and drug discovery. In [6], the authors proposed an AI-enabled diagnosis based on cough droplets. More specifically, the authors developed the AI4COVID-19 app, which can diagnose COVID-19 with negligible misdiagnosis probability. The work [7] discusses a machine learning-based technique and datasets that can be used for diagnosing COVID-19 and also identifying the stage of infection. In general, the DNN model falls under a big pool of AI techniques. DNN has been used previously for detecting pneumonia and other thorax diseases with chest X-rays [8]. DNN has many other vast applications in healthcare such as Parkinson's, Breast Cancer, Electrocardiography, recognition of medical/surgical instruments, robotic surgery, among others. The authors of [8] also proposed an efficient DNN-based model for COVID-19 detection using the x-ray images. Their model achieves 93.9% accuracy. 96.8% COVID-19 Sensitivity, and positivity prediction of 100%. The authors have shown that the computational requirement of their model is such that it is highly suitable for use in medical equipment or SoCs such as a smartphone. Therefore, we used the DNN-based approach in the current study. We have added this discussion in the second paragraph of Section II.

While most existing works are simulation-based analyses, we propose Covidrone - (IoMT and DNN-enabled drone framework) that could be used in real life for contact-less COVID-19 diagnosis and testing.

The major contributions of this work are listed below.

- The integration of sensing, authentication, and disinfectant modules to an IoMT-based drone system for contact-less diagnosis of Covid-19 patients in rural areas.
- The development of a deep neural network to detect COVID-19 using the raw data collected through the drone system, and dispensing the COVID-19 test kit in case the patient is having a high likelihood of having COVID-19.
- To design and implement drone path planning and operational management while considering parameters like strategic location planning, and battery recharging.
- To integrate the various subsystems to develop a working prototype and testing the system through a pilot test run.

III. System Development

In this section, we discuss in detail four objectives that are to be achieved as part of the Covidrone framework.

A. *Objective 1 -*

To integrate sensing, authentication, and disinfectant modules to an IoMT-based drone system for contact-less

diagnosis of Covid-19 patients in rural areas.

The first step is to procure an indigenous drone and develop the drone system to have the following modules:

1) An autonomous sensing module that collects the required data as per the symptoms of the individual.
2) Authentication module that ensures a proper identification of the individual before data collection.
3) A disinfectant module ensures that the complete drone system gets sanitized after every testing.

Sensing Module: The sensing module should sense various critical parameters that help the proposed Covidrone system to detect whether a person has COVID-19 or not. This sensing will be carried out on the drone in an autonomous manner. The interface to the sensing module should be carefully designed, considering ease of use for the rural population who may not be familiar with the technology. As the symptoms of COVID-19 infection ranges from mild to severe, the Covidrone system targets the most common symptoms such as fever, shortage of breath, and dry cough. These symptoms can be measured using the body temperature, oxygen saturation level (SpO2), and audio input, respectively. These sensing parameters have been depicted in Fig. 3.

The medical sensors employed to measure the common symptoms will be a significant part of the payload of Covidrone. The outputs of the sensing module will be inputs to the on-board processing unit (e.g., Raspberry pi) for further processing and communication of the raw medical data. The salient features of the biomedical sensors of the sensing module are described below:

Body temperature: An infrared temperature sensor can measure the body temperature of an individual without physical contact. It avoids close human contact and thus reduces the chances of viral transmissions.

SpO2: Pulse oximeter sensor measures the percentage of blood hemoglobin carrying oxygen. Normal SpO2 is usually at least 95%, and some patients with chronic lung disease or sleep apnea can have normal levels of around 90%. An individual under test can be suspected with COVID-19 if the SpO2 level is less than 85%. The pulse oximeter is portable and lightweight, can be mounted on a drone as part of the payload subsystem.

Audio recording: It is used to record the patient's cough patterns through a mic system connected to the on-board processing unit. The recorded information would serve as useful raw medical data that goes as an input to the DNN model for processing. Additionally, other subjective health evaluations (e.g., sore throat, headache, fatigue, muscle pain) will also be recorded.

Authentication module: The on-board processing unit (Raspberry Pi) will authenticate the individual with the drone. This process is carried out without the need for any other human intervention (other than the user/patient of concern). Following methodologies are proposed for individual authentication:

1) A camera will be mounted on the drone, which is connected to the on-board processor.
2) A proximity sensor on the drone will detect that a person is coming near it. Once the user's presence is detected, the on-board processor on the drone will issue a voice command requesting the user to keep their identification card near the camera.
3) The user will put their identification card in front of the camera. The on-board processing will use OCR (optical character recognition) to extract the user's details.
4) Suppose a person does not have their identification card. In that case, the on-board processor unit will use voice commands to seek the user's personal information by asking a set of questions. Finally, the audio input will be converted into text to populate the necessary fields.
5) The user's photograph will also be taken for validation purposes.

Disinfectant module: It is of utmost importance to disinfect the on-board sensors on the drone after each patient is tested to prevent virus spread among different users of the Covidrone.

Currently, there are mainly two methodologies in practice to disinfect exposed surfaces or medical/surgical equipment. The first method uses a suitable chemical, Sodium Hypochlorite (NaOCl), which is sprayed over contaminated surfaces. Its 5% w/v strength can help to disinfect the surfaces and other medical equipment from various viruses, and bacterial contamination [9]. Additionally, 70% alcohol can also be used to wipe down surfaces where the use of bleach is not suitable, e.g., metal, in which Chloroxylenol (4.5-5.5%)/ Benzalkonium Chloride may be used [10]. Alternatively, Far UltraViolet C (UVC) radiation with sufficient intensity for a suitable time duration can be used to disinfect medical equipment. However, UVC radiation at 254 nm can be a potential health hazard to the skin and eyes. In [11], authors had shown that 222-nm far-UVC radiation can inactivate COVID-19. However, many other sources, including WHO in its advisory, still argue about the safety of UVC radiation on human skin cells and eyes if directly contacted [12]. Hence, a chemical-based disinfectant module is a safe and effective choice. This module will be autonomous and will automatically disinfect the used sensing module after each testing. A pictorial representation of the working of this disinfecting module is shown in Fig. 1. After each testing, the disinfection process will be needed, and hence a suitable sensing mechanism will be developed. Once a single sense cycle is completed, the brass nozzle will spray the disinfectant in the form of small droplets for a pre-defined time period. This complete control will be done by a single master control unit, keeping track of the synchronization of all the tasks to be performed by Covidrone.

*B. **Objective 2:***

To develop a deep neural network to detect COVID-19 using the raw data collected through the drone system, and dispensing the COVID-19 test kit in case the patient is having a high likelihood of having COVID-19.

Following are the various stages of the proposed algorithm for detecting COVID-19 (see Fig. 2).

1) The input data to the proposed algorithm will be the weighted combination of the acquired raw data from the biomedical sensors. [1].

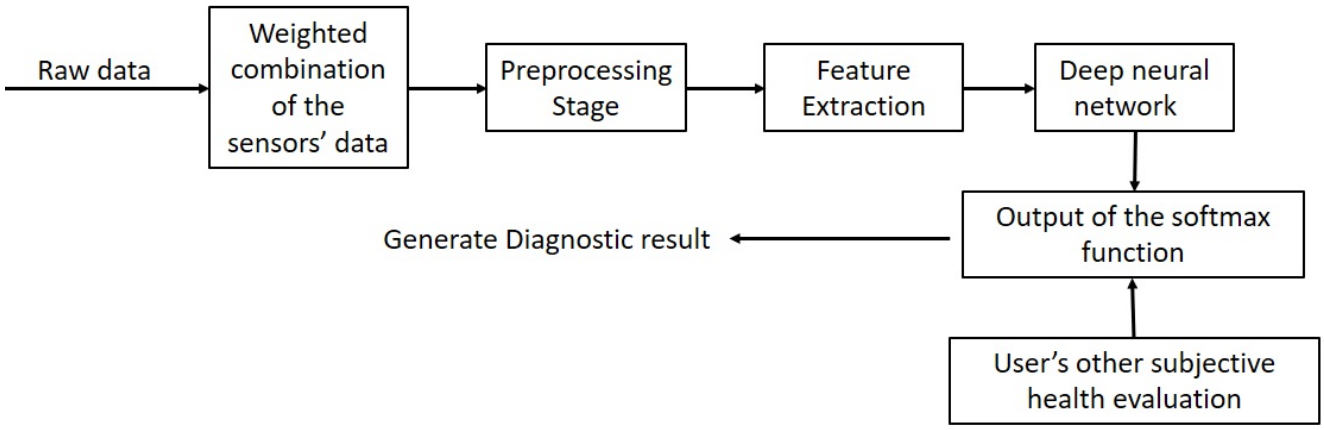


Fig. 2: Block diagram illustrating the proposed algorithm for detecting COVID-19.

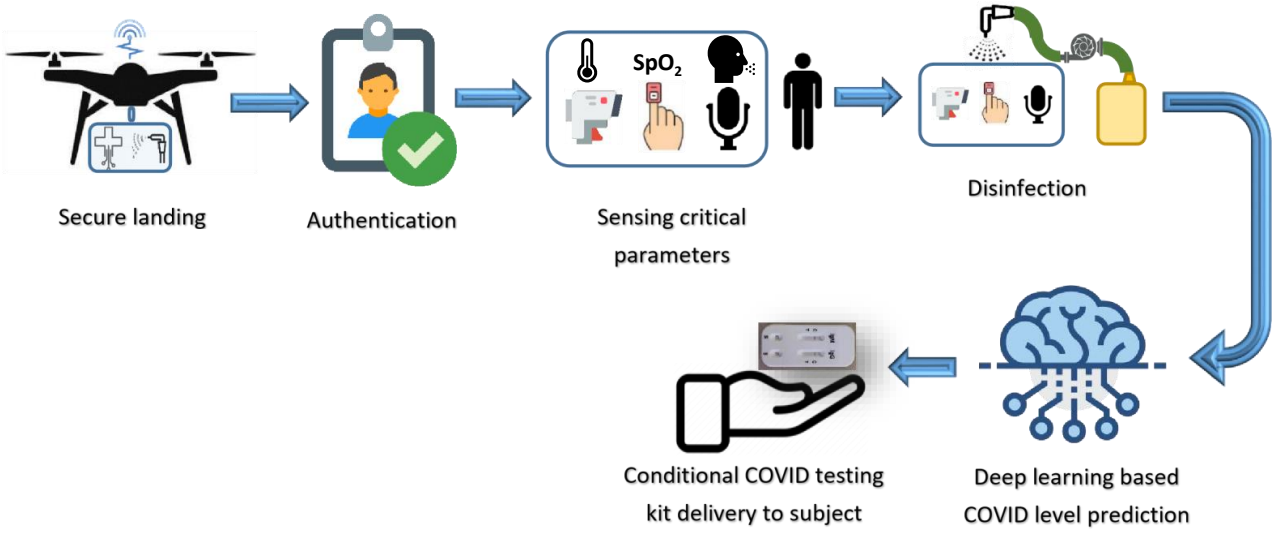


Fig. 3: Process schematic for dispensing COVID testing kit to the user.

2) The highest weight value (less than 1) will be assigned to the sensor data coming from the most common symptom of COVID-19, followed by the lower weight values to the lesser common symptoms. The sum of the proportional weights associated with these symptoms will add up to 1. Furthermore, the exact weight value for each symptom will be decided such that the network meets the performance criteria (see step 7).
3) The input data will then be passed through a preprocessing stage to filter out the unwanted noises acquired during data collection.
4) The signal processing transform (e.g., spectrogram) will be used to extract the important features of the input data.
5) These feature vectors will then be fed to the deep neural network (e.g., convolutional neural network). The output layer of this network will be passed through a soft-max function to obtain the probabilistic value for each class. There will be three classes: negative (no COVID-19), positive (COVID-19), and test inconclusive. The network will be trained and validated using the dataset collected from Objective I. If required, a clustering algorithm can also be adopted for diagnosis.
6) The output of the soft-max function, along with the other subjective information gathered (e.g., sore throat, headache, fatigue, muscle pain) will be used for decision making.
7) The overall accuracy rate, COVID-19 sensitivity, and positive prediction will be computed to evaluate the performance of the network. Additionally, these quantitative metrics will assist us in determining the number of hidden layers and the exact weight values (of step 2) such that these values will be $\geq 90\%$.
8) The network will be tested against the data collected from objective 1 (that were not included in the training and validation stages) and from the pilot test run.
9) The diagnostic result will be generated and stored on the cloud so that the doctor/hospital staff/patient can view or download the report for further action (if required).

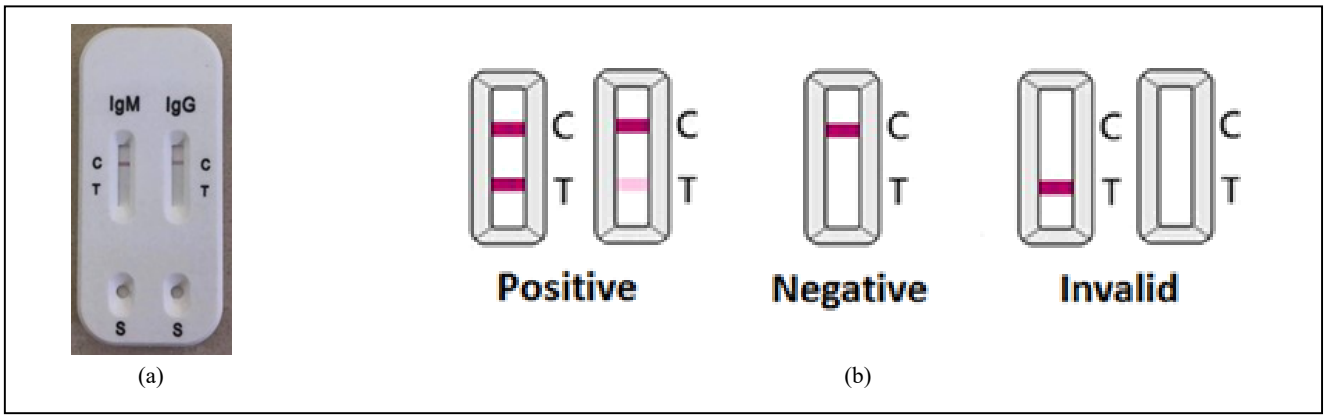


Fig. 4: (a) Typical COVID-19 IgM/IgG Rapid Test Kit and (b) Indications of Test Kit.

COVID-19 testing kit dispensing setup: Our proposed solution, Covidrone, will be equipped to dispense antigen test kits (IgG/IgM) to the user based on his/her likelihood to be infected with COVID-19, as predicted by the DNN model. As the maximum number of kits that the drone can carry could be limited, the kit will only be dispensed to the person who is having a high likelihood of infection. Fig. 3 shows the process flow diagram for distributing the kit by Covidrone.

National Institute of Virology, Pune, has successfully developed the first indigenous anti-SARS-CoV-2 human IgG ELISA test kit for antibody detection of COVID-19. According to the Indian Council of Medical Research (ICMR), ELISA kits are more reliable and cheaper than rapid antibody testing kits. The ELISA (enzyme-linked immunosorbent assay) antibody test is a kind of blood test, which is quite similar to rapid antibody tests that detect antibodies in the blood to determine if an individual is infected with COVID-19 or not. Other rapid testing kits, e.g., Pathocatch COVID-19 by Mylab Discovery Solutions and Standard Q COVID-19 Ag rapid antigen detection test by SD Biosensor, have also been developed to detect the antibodies present in the body. Most of these kits are inexpensive and can be made available under $7 USD per kit. A picture of a typical rapid antibody detection kit for COVID-19 is shown in Fig. 4.

IV. Objective 3.

To design and implement drone path planning and operational management while considering parameters like strategic location planning and battery recharging.

This objective describes the path planning of the drone to reach the destination village site for testing. The overall flow of the drone hosting, path planning, recharging, and landing is shown in Fig. 5. The following are the practical aspects that need to be considered for the implementation of drone battery charging:

1) A single drone will be responsible for medical operations in approximately 8 to 10 villages, with a central facility hosting the setup required for the drone. For example, this central facility could be a major hospital in the given area.
2) The drone visits will be scheduled for testing in a particular village/ locality as per the requirement or can take turns visiting various localities. The drone visits will be planned as unmanned visits using mission planner software to choose the drone's trajectory to reach the required site. The Covidrone will be equipped with an on-board GPS module to enable dynamic path planning.
3) The Covidrone architecture can communicate on existing cellular networks (2G/3G/4G/5G). Its communication requirements are very minimal - to send and receive details of the patient, details of next area to visit, or to communicate emergency notifications - which does not require large bandwidths or high data rates. A single drone will be responsible for medical operations in approximately 8 to 10 villages, with a central facility hosting the setup required for the drone. In the absence of network connectivity, the drone will function in a fully autonomous manner (for example in the region between two villages where network connectivity may not be available). Within the villages, the drone is expected to be in range of network connectivity (either 2G/3G/4G/5G) and it will exchange the required information with the central facility using the available network. If the drone gathers information in regions where there is no internet connectivity, then it will log this information in its on-board storage for future use.
4) As the drone relies on batteries for its operations, careful battery recharging is crucial to ensure smooth operations and longer battery life cycles. Suppose the battery recharging is not done properly, and the battery has to undergo deep discharge cycles. In that case, it may degrade the battery life and might reduce the drone's flight time. To avoid this problem, every village can have an emergency setup for drone charging to reach the central medical facility. This emergency setup could be the village medical center or the residence of a responsible person.
5) Drone scheduling and charging can be implemented using novel energy trading frameworks as described in [13, 14]. In [13], the authors discussed an energy trading framework containing UAVs and charging stations. It is based on blockchain and allows UAVs to charge at charging stations in exchange for digital tokens. In this work,

charge scheduling (deciding which UAV gets access to the charging station at a particular point in time) is implemented using a game-theoretic model. Similarly, in [14], the authors described a peer-to-peer and distributed network comprising UAVs and charging stations. Additionally, an optimal scheduling algorithm was proposed that resulted in cost-optimal charging for the UAVs and higher profits for the charging station. Careful and strategic placement of charging stations combined with efficient charge scheduling schemes can result in less frequent recharge requirements for the drones and thus can increase the number of clients who can be served in a particular time duration.

6) There will be a caretaker for the drone in the central facility who will be responsible for scheduling the drone flights. As the drone will be on unmanned missions, the caretaker can set/ choose the GPS locations for the testing sites.
7) There can be multiple spots in a village where the drone can be scheduled to land for testing, taking into account the affected areas. Initially, each village will be divided into clusters. There will be a drone landing spot in every cluster, enabling the drone to land in any of the clusters based on the population targeted for testing. The clustering of the village will be achieved based on population density across various regions of the village. Regions of the village with higher population density will have smaller clusters so that more drones can be deployed to these regions to cater to the larger number of people. Population information collected from local authorities will be fed into the system's back-end to which would use an efficient clustering algorithm to decide the number of clusters and their boundaries. This will be modeled as an optimization problem in which the number of people catered to in a given amount of time should be maximized for a given number of drones. Parameters such as drone payload capacity, the time required to serve each patient, drone recharging, the closeness of emergency setups, and time of return to the central facility should be taken into account while making these clusters. This would also ensure that older/physically disabled persons in that cluster would not have to travel far distances for testing.
8) The drone can also be used for other healthcare emergencies (e.g., blood-delivery). In such healthcare applications, the caretaker for the drone at the central health facility shall have to select the path of the drone on the mission planner to the desired delivery location.

V. Objective 4

To integrate the various subsystems to develop a working prototype and test the system through a pilot test run.

This objective aims at integrating the various subsystems. Specifically, the biomedical sensing setup, the disinfecting setup, and the on-board signal processing boards will be mounted on the drone. We will choose a central medical set up and carry out an unmanned mission by sending the drone to a test site and conduct a test trial on patients to verify the working status of the system.

End-to-end performance testing can find and eliminate time-consuming bottlenecks and reduce security and operational risks in the entire system. The following are the major factors that need to be considered while testing:

Security and Privacy: During the entire operation, it is critical to ensure security and privacy for the Covidrones and the patients. The communication of the Covidrones with the central medical facility should be secure. Specifically, this communication must satisfy the following features: Mutual Authentication, Data Integrity, Confidentiality, and Identity Privacy [15]. Additionally, since the Covidrone would operate mainly in remote areas, there is a risk of being captured physically or taken control of by people/groups with malicious intent. It may be possible that they may try to gain control of the drone's communication with the central facility, such as extracting secret credentials stored on the drone. They may also even try to steal/replace the on-board testing kits or some components of the on-board sensing module, such as the non-contact thermometer or the pulse oximeter, for monetary gains. By using tamper-proof hardware root of trust (e.g., Physical Unclonable functions), these attempts can be detected and stopped quickly before significant damage occurs to the Covidrone system or the clients.

Sensing functionality and delay: On many occasions, the Covidrone could be responsible for catering to several clients at the same landing location. Hence, it is essential that the authentication, sensing, disinfection, COVID-19 prediction, and conditional kit delivery must be conducted quickly. Before the drone departs to a particular location, the patient's basic details (e.g., name, photograph, identification number) can be gathered from Government databases, encrypted, and stored locally on the Covidrone for quick authentication. A possible time-consuming bottleneck at this stage could be OCR and can be eliminated using a suitable camera (with high resolution, fast shutter, and image stabilization features) to capture high-quality images. Hence, a fast and accurate OCR model can significantly decrease the time consumed at this stage. An outer layer of this model can identify the type of ID Card (e.g., Driver's License, Insurance Card), following which an OCR model can quickly recognize the text for the specific ID card.

At the sensing stage, it is expected that the major delay would arise while the microphone records the cough patterns of the patient. Suitable modifications to the neural network can reduce both the prediction time and the length of the audio clip required for prediction.

Disinfection Stage: The disinfection of the on-board sensors can happen soon after the patient's inputs are recorded. However, depending on the strength of the chemical disinfectant used, there would be a minimum wait time before a new patient's pulse oximeter data can be obtained. There is no such wait time required for the non-contact thermometer and the microphone.

Test Kit Delivery: The conditional testing kit delivery could take a few seconds (5 - 10 seconds) using suitable actuation.

Final Report Generation: The prediction of the DNN model will be uploaded to the cloud server that has virtually unlimited resources. Hence, the report will be generated

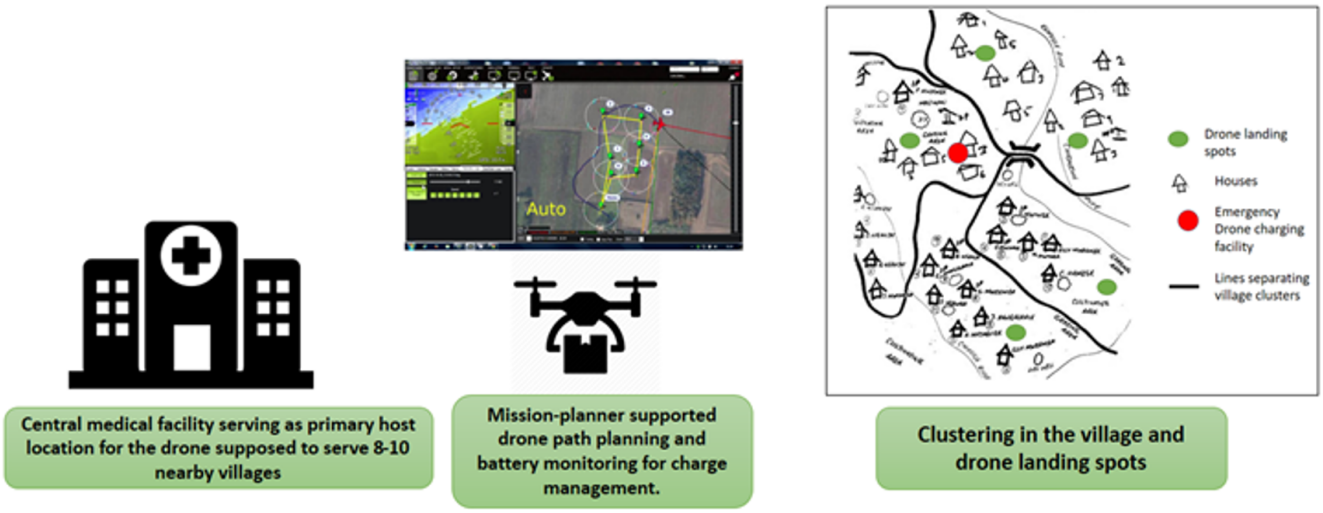


Fig. 5: Overall flow of the drone hosting, path planning, recharging and landing.

immediately and available to the patient/client and the doctor. However, the entire time for generating the report depends on the available network connectivity to the drone at the testing location. If network connectivity is unavailable at the testing location, the result will be stored on the Covidrone itself and uploaded later.

VI. Open Challenges and Future Work

The following improvements and modifications can be explored in future works on drone-assisted covid-19 screening and detection frameworks.

- **Patient Health History**: The present work proposes the use of a camera-based setup on the drone to scan the patient's identification card and also take a photograph of the user for authentication purposes. Future works can explore the usage of more advanced camera and processing systems on board the Covidrone which can scan the previous health and diagnostic reports of the patient. This can be immediately relayed to authorized healthcare personnel for personalized medical advice.
- **Reconfigurable Platforms**: The use of reconfigurable platforms such as Field Programmable Gate Arrays (FPGAs) can add virtually unlimited functionality to the drone. Partial Reconfiguration of the FPGA can be performed at any time to add/remove/modify any of the functionalities of the drone with the only constraint that the necessary medical sensors (such as heart rate sensor, ECG sensor, etc., are already mounted on the drone). With optimal resource partitioning and hardware-level parallelization, an FPGA-based drone system can provide the best balance of both performance and flexibility.
- **Generic Payloads**: While the proposed Covidrone system is specifically meant to be used for tackling COVID-19, future works can implement generic payloads containing a wide variety of medical sensors and processing systems (which could be implemented as per situational requirements using partial reconfiguration as mentioned in the previous point). These systems would help healthcare bodies to respond quickly in the case of any future outbreaks of infectious diseases which occur without warning and spread rapidly.
- **Integration with healthcare databases** Future works can also work on integrating drone-based COVID-19 screening and detecting systems with Local/National/International healthcare databases to enable real-time updates and more accurate statistics of total positive cases, active cases, number of deaths, number of asymptomatic cases along with their demographic distributions. This will help healthcare authorities make better decisions to contain the spread and strategize better healthcare emergency responses.
- **Drone Navigation in challenging environments**: In several cases, the drone might be required to navigate within remote regions that have a thick forest cover or in densely packed rural areas. Unexpected weather phenomena such as rain, fog, and haze could also add to these challenges. To solve this, future works can focus on including computer vision-based video-feed enhancement systems, which could result in better and safer autonomous/manual flying of the drones.
- **Health and Safety Risks** Using drones for healthcare applications where the drones have to interact with patients in very close proximity presents several health and safety risks. The drone's battery has to be properly sealed and secured to prevent harmful chemical leakage. The drone's rotors should be properly guarded to prevent untoward accidents which can harm nearby people. Damage to the rotors may also render the drone unable to return back to its central facility. Hence, future works can include an exclusive focus on these issues and ways to solve them.

VII. Conclusions

COVID-19 has spread rapidly to almost all countries across the globe. It has resulted in millions of human casualties and severely affected economies worldwide. Due to its highly contagious nature, there is an urgent need for technological

innovations which can tackle this pandemic with minimal or no human involvement. The proposed Covidrone system utilized state-of-the-art technologies, including IoMT, drone, and DNN model, for COVID-19 detection. It can authenticate a user, disinfect sensing equipment, and detect COVID-19. Its intelligent and automated path planning, battery recharging, and disinfectant refilling functions will enable secure, reliable, and affordable COVID-19 diagnosis. Thus, it would help disseminate testing kits and essential medical services quickly to remote locations without direct human involvement (fostering contact-less diagnosis and rapid testing). The proposed approach has great importance for combating COVID-19, especially in rural areas where specialized medical facilities are limited.

REFERENCES

[1] V. Chamola, V. Hassija, V. Gupta, and M. Guizani, "A comprehensive review of the covid-19 pandemic and the role of iot, drones, ai, blockchain, and 5g in managing its impact," *Ieee access*, vol. 8, pp. 90 225–90 265, 2020.

[2] R. Gupta, A. Kumari, S. Tanwar, and N. Kumar, "Blockchain-envisioned softwarized multi-swarming uavs to tackle covid-i9 situations," *IEEE Network*, 2020.

[3] A. Kumar, K. Sharma, H. Singh, S. G. Naugriya, S. S. Gill, and R. Buyya, "A drone-based networked system and methods for combating coronavirus disease (covid-19) pandemic," *Future Generation Computer Systems*, vol. 115, pp. 1–19, 2021.

[4] M. Angurala, M. Bala, S. S. Bamber, R. Kaur, and P. Singh, "An internet of things assisted drone based approach to reduce rapid spread of covid-19," *Journal of Safety Science and Resilience*, vol. 1, no. 1, pp. 31–35, 2020.

[5] D. Vekaria, A. Kumari, S. Tanwar, and N. Kumar, "ξboost: An ai-based data analytics scheme for covid-19 prediction and economy boosting," *IEEE Internet of Things Journal*, 2020.

[6] A. Imran, I. Posokhova, H. N. Qureshi, U. Masood, M. S. Riaz, K. Ali, C. N. John, M. I. Hussain, and M. Nabeel, "Ai4covid-19: Ai enabled preliminary diagnosis for covid-19 from cough samples via an app," *Informatics in Medicine Unlocked*, vol. 20, p. 100378, 2020.

[7] H. Kumar, A. Anuradha, and S. Tanwar, "Machine learning-based scheme to identify covid-19 in human bodies," *Emerging Technologies for Battling Covid-19*, vol. 324, p. 35, 2021.

[8] E. Luz, P. L. Silva, R. Silva, L. Silva, G. Moreira, and D. Menotti, "Towards an effective and efficient deep learning model for covid-19 patterns detection in x-ray images," *arXiv preprint arXiv:2004.05717*, 2020.

[9] "COVID-19: Guidelines on disinfection of common public places including offices," Guideline by Ministry of Health and Family Welfare, Government of India, Tech. Rep. [Online]. Available: https://www.mohfw.gov.in/pdf/Guidelinesondisinfectionofcommonpubliccplacesincludingoffices.pdf

[10] W. A. Rutala, D. J. Weber, and Healthcare Infection Control Practices Advisory Committee (HICPAC), "Guideline for Disinfection and Sterilization in Healthcare Facilities, 2008 (updated May, 2019)," Tech. Rep. [Online]. Available: https://www.cdc.gov/infectioncontrol/pdf/guidelines/disinfection-guidelines-H.pdf

[11] M. Buonanno, D. Welch, I. Shuryak, and D. J. Brenner, "Far-uvc light (222 nm) efficiently and safely inactivates airborne human coronaviruses," *Scientific Reports*, vol. 10, no. 1, pp. 1–8, 2020.

[12] "Could Far-UV Light Provide Widespread Disinfection of Airborne Viruses? - IEEE Spectrum." [Online]. Available: https://spectrum.ieee.org/tech-talk/biomedical/devices/ultraviolet-revolution-could-faruv-light-provide-widespread-safe-disinfection-of-airborne-viruses

[13] V. Hassija, V. Chamola, D. N. G. Krishna, and M. Guizani, "A distributed framework for energy trading between uavs and charging stations for critical applications," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 5, pp. 5391–5402, 2020.

[14] V. Hassija, V. Saxena, and V. Chamola, "Scheduling drone charging for multi-drone network based on consensus time-stamp and game theory," *Computer Communications*, vol. 149, pp. 51–61, 2020.

[15] T. Alladi, N. Naren, G. Bansal, V. Chamola, and M. Guizani, "Secauthuav: A novel authentication scheme for uav-base station scenario," *IEEE Transactions on Vehicular Technology*, 2020.

Naren completed his B.E in Electrical and Electronics Engineering, and M.Sc(Hons) in Physics from Birla Institute of Technology and Science (Pilani). He is currently pursuing his M.E In Embedded Systems from BITS-Pilani, Pilani Campus. His research interests include Internet of Things, Internet ofDrones, Healthcare IoT, Vehicle 2 Grid, and their security aspects

Vinay Chamola is an Assistant Professor in the EEE Department & APPCAIR, BITS-Pilani, India. He received his B.E. (2010) and M.E. (2013) degrees from BITS, Pilani and PhD (2016) from National University of Singapore (NUS), Singapore. His research interests include IoT, 5G network management, Blockchain and Security. He is an Area Editor of the Ad Hoc Networks, Elsevier, and also Associate Editor of IEEE Internet of Things Magazine, IEEE Networking Letters, IET Quantum Communications and IET Networks.

Sainath Bitragunta received his Ph.D. degree in Electrical Communication Engineering from the Indian Institute of Science (IISc.), Bangalore. He published peer-reviewed journal articles in reputed journals of IEEE, IET communications, and Elsevier (PhyCom). He has more than ten years of teaching, research, and industry experience. He is currently working as an Assistant Professor in the Department of Electrical and Electronics Engineering at BITS Pilani, Pilani campus. His research interests lie in communication systems analysis, design, and modeling.

Ananthakrishna Chintanpalli received his Ph.D. in Biomedical Engineering from Purdue University in 2011. Subsequently, he moved to the Medical University of South Carolina and completed his Postdoctoral training in 2014. His Master's degree was in Electrical and Computer Engineering (with signal processing as a specialization) from the University of Rhode Island. He is currently an Assistant Professor with the Department of Electrical and Electronics Engineering, BITS-Pilani. His research interests include signal processing, biomedical signal processing, speech perception, auditory computational modelling and neural networks.

Puneet Mishra received his M.E. (2011) and Ph.D. (2017) in the field of Instrumentation and Control Engineering from Delhi College of Engineering, New Delhi and Netaji Subhas Institute of Technology, University of Delhi, New Delhi, respectively. He is currently with Department of Electrical and Electronics Engineering, BITS Pilani as an Assistant Professor. His research interests include intelligent instrumentation, intelligent adaptive control, and bio-inspired optimization algorithms

Sujan Yenuganti received his Master and Ph.D. degrees in Instrumentation Control Engineering, National Institute of Technology Trichy in 2012 and 2017. He worked as a Post-doctoral research fellow in Smart Materials Laboratory, University of North Texas, Texas, USA from January 2017 to 2018. He is now currently working as an Assistant Professor in Electrical Electronics Engineering Department, BITS Pilani, Pilani from January 2018. His research interests include the design and development of sensors using smart materials, MEMS sensors and Actuators and also into Piezoelectric Energy harvesting from ambient wind and vibrations.

Mohsen Guizani received his B.S. (with distinction) and M.S. degrees in electrical engineering and M.S. and Ph.D. degrees in computer engineering from Syracuse University, New York, in 1984, 1986, 1987, and 1990, respectively. He is currently a professor in the Department of Computer Science and Engineering, Qatar University. He is the author/coauthor of nine books and more than 450 publications in refereed journals and conferences. His research interests include wireless communications and mobile computing, smart grid, cloud computing, and security.

View publication stats

KCNet: Kernel-based Canonicalization Network for entities in Recruitment Domain

Nidhi Goyal[1], Niharika Sachdeva[2], Anmol Goel[3], Jushaan Singh Kalra[4], and Ponnurangam Kumaraguru[5⋆]

[1] Indraprastha Institute of Information Technology, New Delhi, India
nidhig@iiitd.ac.in
[2] InfoEdge India Limited, Noida, India
niharika.sachdeva@infoedge.com
[3] Guru Gobind Singh Indraprastha University, Delhi, India
agoel00@gmail.com
[4] Delhi Technological University, Delhi, India
jushaan18@gmail.com
[5] International Institute of Information Technology, Hyderabad, India
pk.guru@iiit.ac.in

Abstract. Online recruitment platforms have abundant user-generated content in the form of job postings, candidate, and company profiles. This content when ingested into Knowledge bases causes redundant, ambiguous, and noisy entities. These multiple (non-standardized) representation of the entities deteriorates the performance of downstream tasks such as job recommender systems, search systems, and question answering. Therefore, making it imperative to canonicalize the entities to improve the performance of such tasks. Recent research discusses either statistical similarity measures or deep learning methods like word-embedding or siamese network-based representations for canonicalization. In this paper, we propose a Kernel-based Canonicalization Network (KCNet) that outperforms all the known statistical and deep learning methods. We also show that the use of side information such as industry type, url of websites, etc. further enhances the performance of the proposed method. Our experiments on 351,600 entities (companies, institutes, skills, and designations) from a popular online recruitment platform demonstrate that the proposed method improves the overall F1-score by 23% compared to the previous baselines, which results in coherent clusters of unique entities.

Keywords: Entity Canonicalization · Recruitment Domain · Entity Normalization

1 Introduction

Recruitment platforms such as LinkedIn, Indeed.com ingest an enormous amount of user-generated content in form of job postings, CVs, and company profiles.

⋆ Major part of this work was done while Ponnurangam Kumaraguru was a faculty at IIIT-Delhi.

This content includes diverse set of recruitment domain entities (company names, institute names, skills, designations) that become part of Knowledge base. As the content is user-generated, multiple variations (e.g., *'economictimes.com'; 'eco. times'; 'the economic times'; 'economic times'; 'ET'*) of each entity name also come up into the KBs. Employing these noisy, redundant, and ambiguous variations directly into downstream applications such as semantic search, question answering, and recommender systems results in poor system performance. Therefore, canonicalization of the entities i.e., mapping various references of a unique entity into a representative cluster, is imperative for recruitment platforms.

Canonicalizing named entities involves various challenges including spelling mistakes and variations (*java developer* & *java deveoper*), overlapping but different entities (*Emerald Bikes pvt limited* & *Emerald Jewellery Retail Limited*), hierarchical variations (*Oracle Financial Services Software* & *Oracle Corporation*), domain-specific concepts (*SOAP* & *REST*), short forms (*umbc* & *University of Maryland, Baltimore*), and semantically similar variations (*Accel Frontline* & *Inspirisys*).

Previous approaches focus on statistical methods [5] for entity canonicalization. However, these methods use handcrafted features that are unable to scale well for advanced (semantic, domain-specific) variations of entities. Fatma et al. [4] employ a deep learning method that overcomes challenges of statistical methods by eliminating the need for explicit feature engineering and using character-based word-embeddings for unknown and emerging entities. Recent literature [9] shows that deep learning methods are often very good at minimizing the training errors but fail to generalize. Literature [9] suggests the introduction of learnable kernels in deep neural networks often improves generalizability.

Therefore, we study a kernel-based neural network designed for entity canonicalization in the recruitment domain. Our proposed method outperforms all the known statistical and deep learning methods on canonicalization tasks. We further enhance the performance of the kernel-based network using side-information which is underexplored in the literature. This literature suggests the use of external side information (morphological, IDF token overlap, PPDB [17]) which is rudimentary [21] and has limited utility in domain-specific settings. We leverage more prosperous meta and semantic side information from external sources (Wikipedia, Google KG) [22,10] to improve the entity canonicalization.

In this paper, we propose a novel multi-tier framework using a learnable kernel network [7,9] which implicitly maps the data into high-dimensional feature space. Our framework captures the non-linear mapping between contextual, meta, and semantic representations through learning objective to output the pairwise similarity between recruitment domain entities. Furthermore, we generate the canonicalized clusters for each entity. We demonstrate and validate the efficacy of our approach on proprietary as well as open source datasets including DBpedia and ESCO [1,3] for generalizability of our solution. We summarize the main contributions of *this* paper as follows:

- We propose a Kernel-based Canonicalization Network (KCNet), which induces a non-linear mapping between the contextual vector representations

while capturing fine-granular and high-dimensional relationships among vectors. To the best of our knowledge, this is the first approach towards exploring kernel features for canonicalizing Knowledge Base entities in the recruitment domain.
- KCNet efficiently models more prosperous semantic and meta side information from external knowledge sources to canonicalize domain-specific entities.
- We perform extensive experiments on real-world proprietary and publicly available datasets in the recruitment domain to show the effectiveness of our proposed approach as compared to baselines.

The organization of the rest of the paper is as follows: Section 2 contains related works; Section 3 elaborates our proposed framework KCNet; Section 4 reports the datasets. Section 5 describes the experimental setup, Section 6 has results and discussion followed by conclusion in Section 7.

2 Related Works

This section briefly describes some of the related works on KB Canonicalization, domain-specific methods, kernel methods, and clustering.

KB Canonicalization. Existing work [5] use manually defined feature spaces to perform the canonicalization task. This approach encodes limited similarity between different semantic representations. Hence, it results in degradation of performance for real-time applications. Vashishth et al. [21] jointly handle noun and relation phrases using knowledge graph embedding models [16] by optimizing its objective function along with using information from external sources called *'side information'*. However, these state-of-the-art knowledge graph embedding methods [2] achieve below par performance for real-world recruitment domain datasets due to noisy, sparseness [6], and context-sensitive information present in triples. Additionally, the side information methods used in literature [21] is rudimentary and lack domain-specific information. Considering these limitations, we leverage external knowledge sources such as Wikipedia Infobox and Google search API which provides additional knowledge for noisy entities.

Domain-Specific methods. Despite the importance of named-entity canonicalization in the recruitment domain, only a few recent studies have explored the problem with respect to unique domain challenges [12,13]. Yan et al. [24] propose a company name normalization system that employs LinkedIn social graphs and a binary classification approach. In this work, the authors use complete profile information as the context. However, this information will be hard to get for new and emerging entities. Lin et al. [11] uses side information and learns domain knowledge from the source text based on the type of entities. Popular state-of-the-art entity linking tools [14] are probabilistic and requires sufficient contextual information to connect to candidate entity and perform well when standard surface forms are available. For example, recruitment domain-specific documents may contain *'Python'* or *'Python Programming'* while, the former is linked to a different type of entity with a high confidence score using these tools. Similarly, Fatma et al. [4] utilizes word and character-based representations based

similarity model along with the attention mechanism to cluster similar entities. However, these works fail to generalize and have limited understanding for more complex and emerging entities.

Kernel-based architectures. Kernel methods have proven effective in exploring larger feature space implicitly in deep learning architectures [23]. Customized kernel [8] based deep learning architectures enhance the performance of the model and map data to an optimized high-level feature space where data may have desirable features toward the application. Recent works utilize deep embedding kernel architectures for identity detection, transfer learning, classification and other tasks [9]. We use kernel infused neural networks to capture the latent semantic relationships and non-linearity between different pair of entities in KBs. These kernel methods are robust for collaboration with neural networks and less expensive than training deep learning architectures.

Clustering methods. Research works have used various clustering techniques for the canonicalization task. Among these methods, Hierarchical Agglomerative Clustering (HAC) is the most extensively used in the literature [5,21].

Our research is uniquely placed at the intersection of the vast literature on kernel-based neural network learning, and clustering approaches for the canonicalization of domain-specific KBs.

3 Kernel-based Canonicalization Network (KCNet)

In this section, we introduce the proposed KCNet approach. We elaborate on the problem definition and each component of our network architecture in detail.

3.1 Problem Definition

Consider $\mathcal{E}$ be the set of entities extracted from job postings, CVs, and company profiles. For each entity x_i, we consider its side information $s_i \in \mathcal{S}$ $\forall x_i \in \mathcal{E}$ acquired from heterogeneous sources (elaborated in detail in section 4.2). Given a pair of entities x_1 and x_2 and their corresponding side information s_1 and s_2 where $x_1, x_2 \in \mathcal{E}$ and $s_1, s_2 \in \mathcal{S}$, the main objective is to find a function $f(x_1, s_1, x_2, s_2) \rightarrow sim(x_1, x_2)$. A pairwise similarity matrix ($\mathcal{M}_{sim}$) is formed by applying f over the set of all entity pairs and then a clustering algorithm is used to form unique canonical clusters of similar entities.

3.2 Network Architecture

We propose a multi-tier novel architecture consisting of three modules: Entity embedding generation, Side Information embedding generation, and Kernel network. We apply clustering technique after obtaining output on our proposed architecture. Fig. 1 shows the overall architecture of our proposed approach (KCNet).

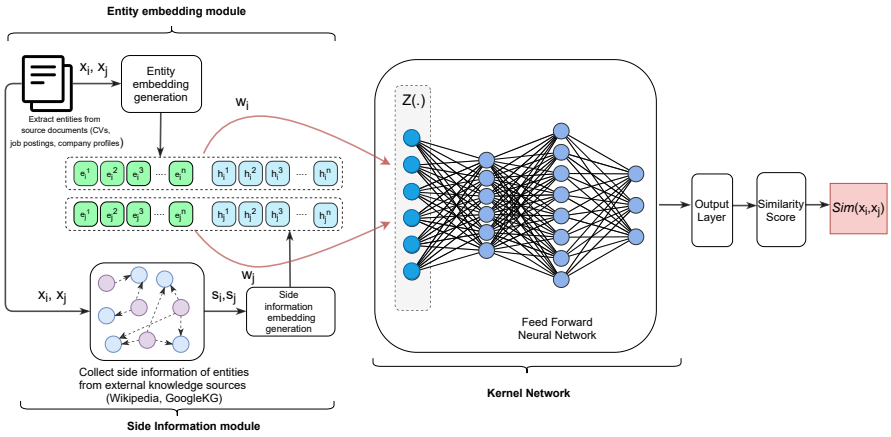


Fig. 1. Kernel-based Canonicalization Network for entities in recruitment domain. We first extract entities and combine it with side information. The combination (concat) is passed through the Kernel network. The output is a pairwise similarity matrix.

- ***Entity embedding generation.*** We obtain an m-dimensional (m=768) vector for each entity pair (x_i, x_j) producing (e_i, e_j) in the space $C \in \mathcal{R}^m$. We use a pre-trained distilled version (fewer parameters, less space and time complexity) of S-BERT [19] to generate initial contextual† [6] embeddings for all entities.
- ***Side Information embedding generation.*** We represent (h_i, h_j) as n-dimensional vector (n=768) for the side information acquired for each entity pair (x_i, x_j) in the space $H \in \mathcal{R}^n$. The formation of side information vector is described in Section 4.2. These representations h_i and h_j are concatenated with the corresponding entity representations e_i and e_j to obtain infused vector representations w_i and w_j for the input pair (x_i, x_j). Here, $w_i = e_i \odot h_i$ and $w_j = e_j \odot h_j$, Note that, $\odot$ is the concatenation function of two vectors producing (w_i,w_j) in the space $W \in \mathcal{R}^{m+n}$. The $(m+n)$ -dimensional vector representation is fed into the kernel network to learn the similarity function, $sim(x_i,x_j)$.
- ***Kernel Network.*** We introduce a kernel network to learn the similarity function f to model complex relationships between the data representations of input pairs in an optimized space. The input to this network is denoted as Z, formed by the combined representation w in the equation (1).

$$Z = (w_i \circ w_j) \odot |w_i - w_j| \tag{1}$$

Here, $\circ$ is a Hadamard (element-wise) product which exploits interactions between two vectors at each dimension. We also determine the $\mathcal{L}_1$ distance

[6] † specifies that an entity name such as '*University of Maryland, Baltimore*' contains the location specific context i.e. '*Baltimore*'. The representation of the entire entity is termed as contextual embedding.

for each dimension w_i and w_j and concatenate the interactions of both the components as shown in equation (2).

$$Z = \left\{ w_i^1 * w_j^1, \ldots, w_i^{m+n} * w_j^{m+n}, |w_i^1 - w_j^1|, \ldots, |w_i^{m+n} - w_j^{m+n}| \right\} \tag{2}$$

where w_i^k represents the k^{th} dimension of w_i. The dimensionality of Z is $2*(m+n)$. Kernel function takes in account both element-wise product (design of polynomial kernel) and differences (design of RBF kernel) over each dimension of original entity. This configuration of inputs allows the network to learn a non-linear relationship between the input pairs and symmetric representations at a fine granular level over each input dimension. Therefore, the learned kernel can map a more robust similarity function over the input space in comparison to traditional methods such as RBF and polynomial [8]. Similar observations for the customized kernel have been made in [9]. The newly obtained vector Z captures the latent semantic relationships between the two input entities. This vector is fed into a multi-layer feed-forward neural network with *sigmoid* output, facilitating the learning of a highly non-linear mapping f to predict similarity over entity pairs. The size of hidden layers (number of neurons) in the kernel network is chosen using a hyperparameter k. We define $k = \alpha * d$ where d is the dimensionality of Z and typically, $\alpha = \{1, 2, 3\}$, say, $f(x_i, x_j) = f(x_j, x_i) > 0$. The kernel network outputs the probability that input pair (x_i, x_j) belong to the same cluster. Therefore,

$$f(x_i, x_j) = P(y^i = y^j | x_i, x_j) \tag{3}$$

– ***Clustering using pairwise similarity scores.*** We compute the pairwise similarity matrix $\mathcal{M}_{sim}$ for all the entity pairs (x_i, x_j) using probability obtained from previous step and apply Hierarchical Agglomerative Clustering (HAC) to form a unique cluster of entities. HAC is popular technique used in literature [5,21] for canonicalization tasks. Each entity is finally mapped to a unique cluster. We choose the number of clusters k using the silhouette index [20]. We repeat the same process for all the datasets (skills, designations, institutes, companies).

4 Datasets

In this section, we describe our datasets and side information collection process in detail.

4.1 Dataset Description

Proprietary Datasets. We use real-world datasets from one of the largest recruitment platforms in India. The dataset i.e., Recruitment Domain Entities (RDE) consists of 25,602 company clusters (RDE (C)), 23,690 institute clusters (RDE (I)), 607 skill clusters (RDE (S)), and 3,894 designation clusters (RDE (D)). The ground truth clusters are manually annotated by domain experts

with a kappa agreement of 0.83. Next, we generate the variation pairs- positive and negative samples. Each name variation of entity $e_x \in \mathcal{E}$ is defined as $\{e_x^1, e_x^2, e_x^3, \ldots e_x^n\}$, which belong to same annotated cluster. We remove Non-ASCII characters to form a set of all unique name variations of e_x. For each entity pair, (e_x^i, e_y^j), training data is prepared using the mapping function g, such that, $g(e_x^i, e_y^j) = 1, \forall(x, y)$ where $i \neq j$ and $x = y$ belongs to same annotated cluster (positive pairs). Similarly, $g(e_x^i, e_y^j) = 0$ where $x \neq y$ belongs to different clusters (negative pairs) using a random sampling approach [15].

Open Datasets. We test the effectiveness of our framework (KCNet) using open-source datasets i.e., DBpedia(C) and ESCO. DBpedia(C) [4] dataset is prepared by querying DBpedia for company names to extract *dbo:Company* which contains 2,944 entity clusters and 22,829 variation pairs. ESCO [3] i.e., ESCO(S) and ESCO(D) are open-source recruitment domain datasets for ESCO-skills and ESCO-designations. ESCO(S) has 35,554 variation pairs and 2,644 clusters of ESCO-skills. ESCO(D) has 62,969 variation pairs and 2,903 clusters. Authors [4] prepared and released these datasets for research community.

4.2 Side information Collection

We leverage two sources for side information extraction, *Wikipedia infoboxes* and *Google KG*.

Wikipedia infobox. We query Wikipedia using its advanced search options [7] and extracted knowledge from Wikipedia infoboxes for different datasets such as {*'title_wikis', 'websites', 'types'*} for RDE(S); {*'Names', 'websites', 'title_wikis'*} for RDE(D); {*'Names', 'websites', 'affiliation'*} for RDE(I); {*'Names', 'websites', 'title_wikis','types'*} for ESCO(S); {*'Names', 'websites', 'title_wikis'*} for ESCO(D); {*'types', 'industries', 'websites', 'native names', 'title_wikis'*} for DBpedia(C).

Google KG. For some entities with short forms, noisy variations, etc. we are unable to fetch knowledge using Wikipedia search; therefore, we leverage Google KG Search API [8] to extract rich semantic textual descriptions of entities to supplement the model with semantic knowledge. Other attributes such as {*location*, *type*, *established*} are extracted that form a part of meta knowledge. Finally, we combine the side information extracted from Google KG and Wikipedia infoboxes. For example, an entity *'vb script'* and its combined side information is defined as *'descriptions'*:'VBScript is an Active Scripting language ... advanced programming constructs'; *'title_wikis'*:'VBScript'; *'websites'* docs.microsoft.com/en-us/previous-versions/t0aew7h6}. We create side information embeddings s_i, a concatenated sequence of side information vector representations $\{s_i^1, s_i^2, \ldots s_i^p\}$, where p is the number of attributes obtained from external sources. We generate the side information embeddings using a pre-trained distilled version of S-BERT [19] model. We follow the same process across all the entity types.

[7] https://www.mediawiki.org/wiki/MediaWiki
[8] https://serpapi.com/

5 Experimental Setup

In this section, we describe our baselines, model configurations, and evaluation metrics.

Baselines. We compare our approach against the following methods:

Galarraga-IDF. Authors [5] uses AMIE algorithm and handcrafted features to find the similarity between entity e_x and entity e_y. We utilize the weighted word overlap approach as a baseline method.

Entity embeddings (Distilled S-BERT(*)) +cosine. We generate our entity embeddings (see Section 3.2) to obtain the vector representation for the entity pair (x_i, x_j). Instead of using the next module, i.e. Kernel Network, we apply cosine similarity measure to get a pairwise similarity matrix.

Entity and Side information embeddings (Distilled S-BERT(**)) +cosine. We obtain entity embedding of (x_i, x_j) and side information embedding of (s_i, s_j) to get (w_i, w_j) (see Section 3.2). The pairwise similarity matrix is generated using cosine similarity.

Char-BiLSTM+A. Fatma et al. [4] describe the architecture which utilize a siamese network that takes characters as input and passes it through the pair of BiLSTM layers enhanced by the attention layer.

Word-BiLSTM+A. This baseline modifies the previous method (Char-BiLSTM+A) [4] by replacing character-based representations with word-based representations followed by attention layer.

Char-BiLSTM+A+Word+A. Authors [4] combine Char-BiLSTM+A and Word-BiLSTM+A architectures combining word and character representations followed by attention mechanism.

Model Configurations. We learn pairwise similarity models using the proposed architecture for different datasets (companies, institutes, designations, skills). The training and testing dataset split is taken as *(80, 20)*. The optimal value of hyperparameters (*size of hidden layer,* α) for companies, designations, and skills is (*1536, 2*) whereas for institutes, (*768, 2*). Batch-size is *512* and the number of fully connected layers are *3*. Rectified linear units ($ReLU$) is used as activation function and dropout rate is 0.3. *Binary cross-entropy loss* and *Adam* as an optimizer is used to train the kernel network and learn the pairwise similarity matrix ($\mathcal{M}_{sim}$).

Evaluation metrics. For pairwise similarity results (Table 1), we use *Precision* (P), *Recall* (R) and *F1-score* (F) [18]. We evaluate clusters (see Table 2) using *Micro Precision*, *Macro Precision*, *Micro Recall*, and *Macro Recall* used in the literature [21].

6 Results and Discussion

Pairwise Similarity. Table 1 summarizes the test results of the pairwise similarity of KCNet along with other baseline approaches. We observe that *Galarraga IDF* (a weighted word overlap similarity measure) and the entity embeddings generated using *Distilled S-BERT(*)+cosine* results in low pairwise similarity

Table 1. Test Results of pairwise similarity using our proposed model in comparison with different baselines. Here S, D, I, C refers to Skills, Designations, Institutes, and Companies datasets respectively. Results of † are taken from [4]. P and F refers to Precision and F1-scores. Distilled S-BERT (*, **) refers to (entity, entity ⊙ side information) embedding using distilled S-BERT model.

Model	Performance													
	Proprietary								Open					
	S		D		I		C		ESCO(S)		ESCO(D)		DBpedia(C)	
	P	F	P	F	P	F	P	F	P	F	P	F	P	F
Galarraga-IDF†	33.2	12.5	63.0	60.3	64.3	66.5	75.8	71.2	50.8	32.8	61.7	38.9	22.6	23.6
Distilled S-BERT(*)+cosine	47.8	47.5	49.7	48.8	49.7	49.1	49.2	49.1	49.3	44.4	49.3	39.0	49.6	45.3
Distilled S-BERT(**)+ cosine	47.5	48.8	49.8	49.9	34.6	41.5	56.2	48.4	49.5	50.0	49.4	49.7	50.0	49.8
CharBiLSTM+A†	81.8	86.9	72.6	77.2	84.5	84.8	99.3	98.9	85.9	86.9	76.3	75.1	72.1	59.7
WordBiLSTM+A†	80.1	86.5	90.5	94.8	80.6	83.3	95.3	95.6	85.6	89.6	83.1	83.7	77.6	70.7
CharBiLSTM+A+Word+A†	82.7	88.5	94.4	96.3	86.7	86.7	99.5	99.2	87.3	90.7	84.2	85.4	78.0	71.3
KCNet (without sideinfo)	**96.7**	**90.6**	**99.6**	90.9	**92.4**	**89.3**	99.4	98.8	**99.0**	**95.1**	**98.8**	**86.9**	**99.0**	**92.5**
KCNet (with sideinfo)	**99.5**	**99.4**	**99.7**	**99.6**	**99.5**	**99.5**	**99.5**	**99.3**	**99.2**	**98.3**	**98.8**	**89.4**	**99.1**	**97.0**

for non-overlapping variations and different surface forms of entities. For eg. (*'mdx'*,*'MultiDimensional eXpressions'*) has similarity of 0.73 using Distilled S-BERT(*)+cosine. *KCNet* gives the similarity of 0.84 as it learns the structure and non-linear mapping in latent space, even in the absence of side information. With side information, *KCNet* learns the latent semantic relationships between these two entities and returns a high similarity score of 0.99. Another example is overlapping variations (*uplholstery fillings, upholstery paddings*); *Distilled S-BERT(*)+cosine* returns a pairwise similarity score of 0.86 for these same entities, whereas *KCNet* learns a better representation and gives a pairwise similarity score of 0.99. *KCNet* generalizes well across all entity types, it gives higher P and F even for all open datasets where it outperforms with 21% F1-score as compared to best baseline.

Clustering results. Test results after applying the clustering approach is reported in Table 2. Overall, *KCNet* significantly outperforms the best baseline [4] by an improved micro F1-score by 23% and macro F1-score by 25%. A one-way repeated measures ANOVA test was conducted to determine significance for all evaluation metrics ($p < 0.00003$).

Side information for KCNet. We evaluate the performances of different versions of *KCNet* (with and without side info). From Table 1, we observe that P and F performance benefits from increased performance in the presence of side information. *Char-BiLSTM+A+Word+A* captures limited patterns and unable to model similar semantic variations (*mycology, fungi studies*) for which *KCNet* gives a pairwise similarity score of 0.98. This shows that *KCNet* is able to model these variations well when supplemented with side-information. Even though side information might be unavailable for a some entities, the proposed framework results in overall better entity canonicalization.

Error analysis: Although KCNet gives promising results across all datasets, it wrongly clusters some entities; for example, some skills such as *bees wax* and *natural wax* signify same concept but occur in the different cluster. One possible reason could be that the representation of words *bees* and *natural* are far

Table 2. Test Results over HAC using pairwise similarity. Here, β: baseline (*Char-BiLSTM+A+Word+A*) and γ: proposed model (*KCNet*) with sideinfo.

Dataset	Model	Metrics					
		Micro			Macro		
		P	R	F	P	R	F
S	β	0.71	0.64	0.67	0.94	0.31	0.47
	γ	**0.99**	**0.97**	**0.98**	**0.96**	**0.97**	**0.96**
D	β	0.95	0.53	0.67	0.83	0.15	0.24
	γ	0.86	**0.78**	**0.82**	**0.85**	**0.54**	**0.66**
I	β	0.84	0.75	0.79	0.96	0.48	0.64
	γ	0.83	**0.85**	**0.84**	0.74	**0.71**	**0.72**
C	β	0.98	0.99	0.98	0.97	0.96	0.96
	γ	0.98	0.97	**0.98**	**0.98**	**0.97**	**0.98**
ESCO(S)	β	0.84	0.82	0.83	0.65	0.49	0.55
	γ	**0.93**	**0.92**	**0.92**	**0.89**	**0.75**	**0.81**
ESCO(D)	β	0.49	0.79	0.61	0.21	0.32	0.25
	γ	**0.91**	0.61	**0.73**	**0.81**	0.22	**0.34**
DBpedia(C)	β	0.88	0.52	0.65	0.92	0.25	0.39
	γ	**0.93**	**0.75**	**0.83**	0.86	**0.60**	**0.70**

apart in the contextual vector representation space, so the model assigns a lower similarity score and hence, incorrectly classifies it. Similarly, *'packager'* is incorrectly placed in cluster of [*'dozer driver'*, *'dozer/crawler driver'*, *'packager'*]. The possible reason could be the complete absence of side information for three entities confuses KCNet with closer contextual vector representations. Despite this, KCNet addresses the challenge of handling abbreviations, short forms, and non-overlapping entities by learning vector representations of these entities in the kernel space.

7 Conclusion

Our research focused upon canonicalizing real-world entities from the recruitment domain such as companies, designations, institutes, and skills by designing a novel multi-tier framework Kernel-based Canonicalization Network (KCNet). KCNet induces a non-linear mapping between the contextual vector representations while capturing fine-granular and high-dimensional relationships among vectors. KCNet efficiently models more prosperous semantic and meta side information from external knowledge towards exploring kernel features for canonicalizing entities in the recruitment domain. Furthermore, we applied Hierarchical Agglomerative Clustering (HAC) using the pairwise similarity matrix $\mathcal{M}_{sim}$ to create unique clusters of entities. Experiments revealed that the Kernel-based neural network approach achieves significantly higher performance on both proprietary and open datasets. We demonstrate that our proposed methods are also generalizable to domain-specific entities in similar scenarios.

Acknowledgements.

We would like to acknowledge the support from SERB, InfoEdge India Limited, and FICCI. We are grateful to PreCog Research Group and Dr. Siddartha Asthana for critically reviewing the manuscript and stimulating discussions.

References

1. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: Dbpedia: A nucleus for a web of open data. In: The semantic web, pp. 722–735. Springer (2007)
2. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: Advances in neural information processing systems (2013)
3. European Commission: ESCO handbook. EU publications (2019)
4. Fatma, N., Choudhary, V., Sachdeva, N., Rajput, N.: Canonicalizing knowledge bases for recruitment domain. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. pp. 500–513. Springer (2020)
5. Galárraga, L., Heitz, G., Murphy, K., Suchanek, F.M.: Canonicalizing open knowledge bases. In: Proceedings of the 23rd acm international conference on conference on information and knowledge management. pp. 1679–1688 (2014)
6. Gupta, S., Kenkre, S., Talukdar, P.: Care: Open knowledge graph embeddings. 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). pp. 378–388 (2019)
7. Hofmann, T., Schölkopf, B., Smola, A.J.: Kernel methods in machine learning. The annals of statistics pp. 1171–1220 (2008)
8. Kuo, B.C., Ho, H.H., Li, C.H., Hung, C.C., Taur, J.S.: A kernel-based feature selection method for svm with rbf kernel for hyperspectral image classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **7**(1), 317–326 (2013)
9. Le, L., Xie, Y.: Deep embedding kernel. Neurocomputing **339**, 292–302 (2019)
10. Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., Van Kleef, P., Auer, S., et al.: Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. Semantic Web **6**(2), 167–195 (2015)
11. Lin, X., Chen, L.: Canonicalization of open knowledge bases with side information from the source text. In: 2019 IEEE 35th International Conference on Data Engineering (ICDE). pp. 950–961. IEEE (2019)
12. Liu, Q., Javed, F., Dave, V.S., Joshi, A.: Supporting employer name normalization at both entity and cluster level. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1883–1892 (2017)
13. Liu, Q., Javed, F., Mcnair, M.: Companydepot: Employer name normalization in the online recruitment industry. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 521–530 (2016)
14. Mendes, P.N., Jakob, M., García-Silva, A., Bizer, C.: Dbpedia spotlight: shedding light on the web of documents. In: Proceedings of the 7th international conference on semantic systems. pp. 1–8 (2011)

15. Neculoiu, P., Versteegh, M., Rotaru, M.: Learning text similarity with siamese recurrent networks. In: Proceedings of the 1st Workshop on Representation Learning for NLP. pp. 148–157 (2016)
16. Nickel, M., Rosasco, L., Poggio, T.A., et al.: Holographic embeddings of knowledge graphs. In: AAAI. vol. 2, pp. 3–2 (2016)
17. Pavlick, E., Rastogi, P., Ganitkevitch, J., Van Durme, B., Callison-Burch, C.: Ppdb 2.0: Better paraphrase ranking, fine-grained entailment relations, word embeddings, and style classification. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). pp. 425–430 (2015)
18. Raghavan, V., Bollmann, P., Jung, G.S.: A critical investigation of recall and precision as measures of retrieval system performance. ACM Transactions on Information Systems (TOIS) **7**(3), 205–229 (1989)
19. Reimers, N., Gurevych, I.: Sentence-BERT: Sentence embeddings using Siamese BERT-networks. 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). pp. 3982–3992. Association for Computational Linguistics, Hong Kong, China (Nov 2019)
20. Starczewski, A., Krzyżak, A.: Performance evaluation of the silhouette index. In: International Conference on Artificial Intelligence and Soft Computing. pp. 49–58. Springer (2015)
21. Vashishth, S., Jain, P., Talukdar, P.: CESI: Canonicalizing open knowledge bases using embeddings and side information. In: Proceedings of the 2018 World Wide Web Conference. pp. 1317–1327. WWW '18, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland (2018)
22. Vrandečić, D., Krötzsch, M.: Wikidata: a free collaborative knowledgebase. Communications of the ACM **57**(10), 78–85 (2014)
23. Weston, J., Ratle, F., Mobahi, H., Collobert, R.: Deep learning via semi-supervised embedding. In: Neural networks: Tricks of the trade, pp. 639–655. Springer (2012)
24. Yan, B., Bajaj, L., Bhasin, A.: Entity resolution using social graphs for business applications. In: 2011 International Conference on Advances in Social Networks Analysis and Mining. pp. 220–227. IEEE (2011)

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-021-11092-8

Learning image by-parts using early and late fusion of auto-encoder features

Seba Susan [1] · Jatin Malhotra [1]

Received: 12 September 2020 / Revised: 7 January 2021 / Accepted: 21 May 2021
Published online: 03 July 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

A novel sub-part learning scheme is introduced in our work for the purpose of recognizing handwritten numeral images. The idea is borrowed from the concept of visual perception and part-wise integration of visual information by the cortical regions of the brain. In this context, each numeral image is divided into four half-parts: top-half, bottom-half, left-half and right-half; the other half of the image being kept masked. An efficient data representation is derived in an unsupervised manner, from each image part, using convolutional auto-encoders (CAE), for our learning scheme that involves both early and late fusion of features. The chief advantage of the features derived from convolutional auto-encoders is the preservation of 2D spatial locality while the features are being filtered layer-by-layer through the convolutional architecture. The features derived from each individual CAE are fused by concatenation in our early fusion scheme, and learnt using an appropriate classifier. The late fusion strategy involves learning the probability density pertaining to the predicted values emanating from the four base classifiers using a meta-learner classifier. The early-cum-late fusion is proposed in the later stage of our work to combine the goodness of both schemes and enhance the performance. The support vector machine is used in all the classification stages. Experiments on the benchmark MNIST dataset of handwritten English numerals prove that our method competes favorably to the state of the art, as inferred from the high classification scores achieved. Our method thus provides a computationally simple and effective methodology for sub-part learning and part-wise integration of information from different parts of the image. The method also contributes to saving in computational expense since, at a time, only a small part of the image is processed, speeding up the inferencing process.

Keywords Handwritten numeral recognition · Sub-part learning · Convolutional auto-encoder · Early fusion · Late fusion · Early-cum-late fusion

✉ Seba Susan
seba_406@yahoo.in

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

1 Introduction

The automated recognition of handwritten numeral images assumes significance in real-world applications involving digitization of financial and accounting documents [52], deciphering digits from signboards in natural scenery images or videos [28], reading license plates from blurred images [33] and so on. Encapsulating text and numeral characters from document images, is by itself a complex task due to the varying handwriting styles, noisy artifacts, blur, presence of foreign objects, occlusion etc. that characterize handwritten digitized documents [37]. The next stage involves recognizing the bounded text or numeral character as belonging to one of the alphabets of a particular language form. In our work, we specifically undertake the problem of classification of handwritten numeral images that are categorized into ten numeral classes representing the digits 0 to 9, in most languages.

Understanding image content has been made much easier with the advent of deep neural networks that take various forms of (deeper): - fully connected dense networks, convolutional neural networks (CNN), restricted Boltzmann machines and auto-encoders [19]. Much before convolutional neural networks (CNN) came into vogue, handcrafted features were derived from numeral images using distinctive feature extraction schemes such as box-partitioning [44] and curvature measurements [31], and learnt using a suitable classifier such as the Multi-Layer Perceptron (MLP) neural network or Support Vector Machine (SVM). Typical image features like those derived from Gabor filters, Local Binary Pattern (LBP), Histogram of gradients (HOG) etc. have also been popularly applied [6, 9, 36]. However, latent representations derived from CNN have till date outperformed several of these handcrafted features in image processing tasks. This is attributed to their intrinsic ability to hierarchically process and retain filtered information in 2D [16]. Most of the neural networks require supervised training with the exception of the auto-encoders [2] that learn a latent representation of the input image in a greedy unsupervised manner. Auto-encoders have been applied for feature engineering from Arabic numeral images in a recent work [21]. Some other works that deserve mention are [48] for Chinese character recognition and [29] for Farsi numeral recognition, both of which used stacked auto-encoders for unsupervised feature extraction. Convolutional auto-encoder (CAE), that hierarchically retains 2D information on spatial locality, was proposed by Masci et al. [25] to improve the efficiency of the deep features extracted. The authors applied stacked CAE for classification of handwritten English numerals in [25]; the work has been extended to multi-lingual applications as well [7].

In this paper, we integrate latent representations emanating from convolutional auto-encoders that are trained separately on four half-parts of the numeral images. A novel early-cum-late fusion approach is proposed for the integrating the information. The contributions of our work are as follows:

1. The handwritten numeral image is divided into four half-parts-top, bottom, left and right, retaining the pixel intensities in one half while keeping the second half masked. Our work thus advocates learning the image by-parts rather than as a whole. The part-wise information extracted is integrated both in the early and later stages of the recognition pipeline, in a manner inspired by the human visual system.
2. Structure learning involves learning the hidden state activations of neural networks called the latent representation which is, in our case, the auto-encoder features that are derived in an unsupervised manner from the raw input image.

3. A novel early-cum-late fusion of features derived from convolutional auto-encoders (CAE) is proposed for handwritten numeral recognition. Early fusion involves concatenation of the latent representations from the individual CAEs while late fusion involves concatenation of the posterior class probabilities of the base classifiers.
4. Our method advocates the combination of part-wise learning with structure learning as a successful venture for explainable AI that helps to integrate and interpret visual information in a better manner.

The rest of this paper is organized as follows: a review of the related works and the feature extraction mechanism from convolutional auto-encoders is given in Section 2, the proposed part-learning approach and the early, late, and early-cum-late fusion schemes are discussed in Section 3, the experimentation and results are analyzed in Section 4 and the conclusions are drawn in Section 5.

2 Related work

The problem of learning directly from raw pixels is the high-dimensionality of the input. To reduce the dimensionality, we can induce some form of feature computation and compression. This idea gave rise to feature engineering that dominated the research area for several decades until the resurgence of deep learning (refer [1]) that learns directly from raw images and computes features intrinsically to the network. The deep convolutional neural networks process the input raw images and extract low-level features in the lower layers and high-level features in the higher layers of the network [15]. Deep CNNs trained on handwritten images in [50] were found to perform well by adjustments in dropout rate and regularization. Particularly in [50], the mis-identification rate reduced when only half the hidden units were involved in the computations, using dropout. Deep generative models are the current state of the art in several computer vision tasks. In a recent work [3], a variety of Generative Adversarial Network (GAN) models were compared for recognition of handwritten numerals; Auxiliary Classifier Generative Adversarial Network (ACGAN) variant was found to give the best classification accuracy on the benchmark Modified National Institute of Standards and Technology (MNIST) dataset. A very deep highway network is proposed by Srivastava et al. in [35] that involves training of 151 K parameters for handwritten numeral recognition. A comparison of deep learning models for handwritten numeral recognition is found in [27], where the best performance was observed for the capsule networks that record orientation and position attributes unlike CNNs, closely followed by the CNN and the pre-trained CNN ResNet, with logistic regression performing the worst among all methods. All these works, generally, reiterate the utility of convolutional networks for efficient processing of 2D numeral images.

Neural networks are, however, sensitive to weight initialization since the optimization procedure is gradient-based [45]. Recently, alternatives to directly training the neural networks have been explored. Instead of iteratively training a neural network, some researchers have proposed non-iterative techniques using neural-like structures as in [14, 46, 47]. Two separate CNNs were used in [41] for weight pre-initialization, that were pre-trained on two separate halves of the handwritten numeral image using only 5% of the dataset. The outputs of the two CNNs were integrated by probabilistic score fusion. Extracting hidden state activations from the trained model, also known as latent representation, is yet another technique for mitigating

the weight initialization issue. Latent representations derived from Multi-layer perceptron neural networks [42] and convolutional networks [43] for images of handwritten characters were classified using a third-party classifier such as the support vector machine.

Feature engineering thus permeated through deep neural networks as well, an instance of this is the auto-encoder that learns lower dimensional features from the raw input in an unsupervised manner. The auto-encoder (encoder-decoder) is a three layer fully-connected network proposed by Bengio et al. in 2007 [2] where the input layer is of the size of the (linearized and flattened) input image and the output of the auto-encoder is the reconstructed (linearized) input image. The first two layers comprising of the input layer and the hidden layer constitute the encoder part while the last layer is the decoder. The features extracted are the hidden state (latent) activations emanating from the trained auto-encoder (output of the encoder part) for each sample input.

Consider the input image (reshaped to 1D) be denoted by $\mathbf{x} \in \mathbb{R}^d$ and let $\mathbf{h} \in \mathbb{R}^c$ denote the associated latent representation. The latent code is derived from the input layer through a weighted linear combination of inputs followed by a non-linearity function σ (such as the hyperbolic tangent function that is used in our experiments), as shown below.

$$\mathbf{h} = \sigma(\mathbf{w}\mathbf{x} + b) \tag{1}$$

Let $\mathbf{y} \in \mathbb{R}^d$ be the reconstructed input at the output of the auto-encoder given by

$$\mathbf{y} = \sigma(\mathbf{w}'\mathbf{h} + b') \tag{2}$$

Here, the weight parameters ($\mathbf{w}$, $\mathbf{w}$') are learnt during the training phase while the bias (b, b') is used for regularization. The conventional 3-layer auto-encoder [2] does not retain information of the 2D image layout. This defect is removed in the Convolutional Auto-Encoder (CAE) introduced by Masci et al. [25] that accepts and processes a 2D image in a manner similar to convolutional neural networks. The CAE contains a set of convolutional layers as opposed to AEs that contain only fully connected layers. The input is filtered in 2D (by convolution operation on the input) through an identical though reverse hierarchy of convolutional layers in the encoder and decoder modules of the CAE. The max-pooling layers in the encoder section downsample the input to generate the latent code at the output of the encoder stage. In the decoder section, the steps are retraced by appropriate 2D up-sampling to generate the reconstructed 2D image. The decoding stage is thus akin to a linear combination of image patches. For the k^{th} feature map, the latent code or hidden state activation for CAE is given by,

$$\mathbf{h}^k = \sigma\left(\mathbf{w}^k\mathbf{x} + b^k\right) \tag{3}$$

Let $\mathbf{y} \in \mathbb{R}^d$ be the reconstructed input at the output of the CAE given by the non-linear transformation of the linear weighted combination of all latent codes pertaining to the feature maps, the weights being flipped ($\tilde{\mathbf{w}}$) in the case of the decoder, as shown below.

$$\mathbf{y} = \sigma\left(\sum_k\left(\tilde{\mathbf{w}}^k\mathbf{h}^k + l\right)\right) \tag{4}$$

The latent code is extracted as per (3) and linearized to a 1D vector $\mathbf{h}$, in this manner, for all images. These are then applied as input features to any suitable classifier, for learning from the training set and subsequent classification of the test set. Such a scheme involving local sharing of weights is found acceptable for learning from 2D images, due to the preservation of spatial locality while the input filters through successive convolutional layers, as proved from the

higher classification scores in experiments on CAE in [2, 12, 49]. As in all other forms of convolutional networks, the weights of the CAE are learnt using stochastic gradient descent. The auto-encoder processes the whole image at a time and the features derived are fed to a suitable classifier for recognizing the handwritten character, specifically- softmax classifier in [21], neural networks in [48] and softmax regression in [29].

3 Proposed half-part learning with early and late fusion

3.1 Motivation from the human visual and cognitive system

A novel sub-part learning scheme is introduced in our work for the purpose of recognizing handwritten numeral images. Part-based information integration is advantageous in terms of both accuracy and speed, as compared to the convention of processing the whole image at one go. The idea is borrowed from the concept of visual perception and part-wise integration of visual information by the cortical regions of the brain [24]. The human visual system (HVS) employs a selective attention mechanism that pays greater attention to certain sub-parts that are more informative on the category of the scene or the primary object in the scene. A suggestive work that is modelled on human cognition is [34] that breaks up the numeral image into patches, creates a dictionary for feature extraction, and applies the features to a two-stage hierarchical neural network model. Another work that is indicative is [54] that aggregates auto-encoder features parallelly into a 3D shape descriptor. We affirm the theory that learning the numeral image by-parts is more useful than learning the image as a whole since the variation is rather learnt locally for a part rather than for the whole image, and it also involves less computational expense. The parts of a numeral image that can be decoded semantically would involve, inarguably, at least 50% of the numeral image, since any less would not relate to the shape and the curvature of the particular numeral being recognized. In this context, each numeral image is divided into four half-parts: top-half, bottom-half, left-half and right-half; the other half of the image being kept masked. An efficient data representation is derived in an unsupervised manner, from each image part, using convolutional auto-encoders (CAE), for our learning scheme that involves both early and late fusion of features.

3.2 Early-cum-late fusion for part-learning

Part-learning has been a part of several pattern recognition experiments. The matching results obtained by comparing parts of the image is more useful than comparing the image as a whole, as observed in the facial expression recognition experiment in [39] wherein the facial image was divided into grids. In most practices, the parts that are learnt pertain to specific parts of objects like car wheels [8], cat heads and human bodies and faces [22] and specific facial features like the eye pair [38]. In our case, for numeral images that are static over time, we assume that semantic interpretations are related to the top-half, bottom-half, left-half and right-half parts of the numeral image. We propose two distinct strategies for matching: the early fusion of features by concatenation of all four feature sets, and the late fusion of probabilistic scores obtained after processing each part separately. Early-cum-late fusion is proposed for further enhancing the performance. The early-cum-late fusion of features indicates the same mechanism as HVS that hierarchically integrates the total scene information, in the brain cortical regions. The early, late, and early-cum-late fusion strategies are described next.

Early and late fusion schemes are important feature engineering strategies employed in diverse pattern recognition problems that involve a double decision layer [23, 32]. The probabilistic confidence scores ensuing from two-class random decision tree forest classifiers were compared in [40] to decide the multi-class label. Early and late fusion strategies differ in the manner in which the multimodal results are integrated [32]. While early fusion combines unimodal representations extracted individually, late fusion integrates predictive scores extracted from individual probabilistic models that are trained on the unimodal features.

a) Early fusion In our experiments, the features extracted from the CAE trained on each half-part are 128-dimensional each. The chief advantage of the features derived from convolutional auto-encoders is the preservation of 2D spatial locality while the features are being filtered layer-by-layer through the convolutional architecture. In our early fusion strategy, we fuse the features derived from the four half-parts, by concatenation. The 512-dimensional concatenated vector formed by the early fusion of the four CAE representations is given by

$$\mathbf{h}_{concat} = \left[\mathbf{h}_{top}, \mathbf{h}_{bottom}, \mathbf{h}_{left}, \mathbf{h}_{right}\right]^T \tag{5}$$

b) Late fusion In the late fusion strategy, we fuse or concatenate the posterior class probability scores associated with the ten numeral classes, that emanate from four base classifiers trained on each of the half-parts. The CAE representations derived from each half-part are learnt by the respective base classifiers. Let $\varphi(x)$ denote the predicted value of a sample vector $\mathbf{x}$. Let the posterior class probability associated with class c be denoted by $p_c(\mathbf{x})$ that forms a complete distribution $\mathbf{p}(\mathbf{x})_{part} = \forall\, c\{p_c(\mathbf{x})\}$ such that

$$\sum_c p_c(\mathbf{x}) = 1 \tag{6}$$

Thereby, fusing the four probability vectors pertaining to the decision scores of CNNs trained on the four half-parts, results in a 40-dimensional feature vector represented by

$$\mathbf{p}(\mathbf{x})_{late} = \left[\mathbf{p}(\mathbf{x})_{top}, \mathbf{p}(\mathbf{x})_{bottom}, \mathbf{p}(\mathbf{x})_{left}, \mathbf{p}(\mathbf{x})_{right}\right]^T \tag{7}$$

c) Early-cum-late fusion Let $\mathbf{p}(\mathbf{x})_{concat}$ represent the vector of ten class posterior probabilities emanating from the base classifier that is trained on the 512-dimensional fused auto-encoder features $\mathbf{h}_{concat}$ in (5). In our early-cum-late fusion strategy, $\mathbf{p}(\mathbf{x})_{concat}$ is fused by concatenation with the 40-dimensional probability vector $\mathbf{p}(\mathbf{x})_{concat}$ in (7), as shown below.

$$\mathbf{p}(\mathbf{x})_{early-cum-late} = \left[\mathbf{p}(\mathbf{x})_{concat}, \mathbf{p}(\mathbf{x})_{late}\right] \tag{8}$$

The feature vector derived in (8) is learnt using a meta-learner classifier in the second layer of classification.

The overall process flow depicting a double decision layer of early-cum-late fusion is shown in the block diagram in Fig. 1. The configuration of CAE for our experiments is summarized in Table 1. The CAE is trained on the (half-part of) raw input images of the training set without any requirement of target class labels (unsupervised learning). The output of the encoder stage of the CAE after training is complete (highlighted in gray in Table 1) is of dimension 4x4x8 that when linearized forms a 128-dimensional feature vector for a single

image, that is trained on the base classifier as observed in Fig. 1. The predicted probabilistic scores of the four base classifiers are fused by concatenation in the late fusion strategy. The 128-dimensional feature vector emanating from each half-part is concatenated and learnt using a separate classifier (early fusion). In early-cum-late fusion, the predicted probabilistic scores of late fusion (40-dimension) and early fusion (10- dimension) are fused by concatenation and classified using the meta-classifier.

4 Results and discussions

4.1 Experimental setup

The most researched handwritten numeral dataset is inarguably MNIST, the dataset of English handwritten numeral images, comprising of 60,000 training images and 10,000 testing images segregated at the source [18]. Each image is set to a standard size of 28 × 28. Our experiments were performed using the Keras deep learning library of Python software (3.7 version) for CAE implementation. The different convolutional layers in our CAE architecture are shown in Table 1. The gray-highlighted rows in Table 1 indicate the encoder part that is stripped off after training and used to extract the latent codes for (half-parts of) the input images. Each (training / test image) is processed; half-part images are extracted; the corresponding CAEs are trained followed by latent code extraction and fusion. The Adam optimizer and cross-entropy loss function was used for learning in 50 epochs with a batch size of 128. The code took roughly about 20 min to execute on an Intel Quad Core Pentium PC with graphics card. The software

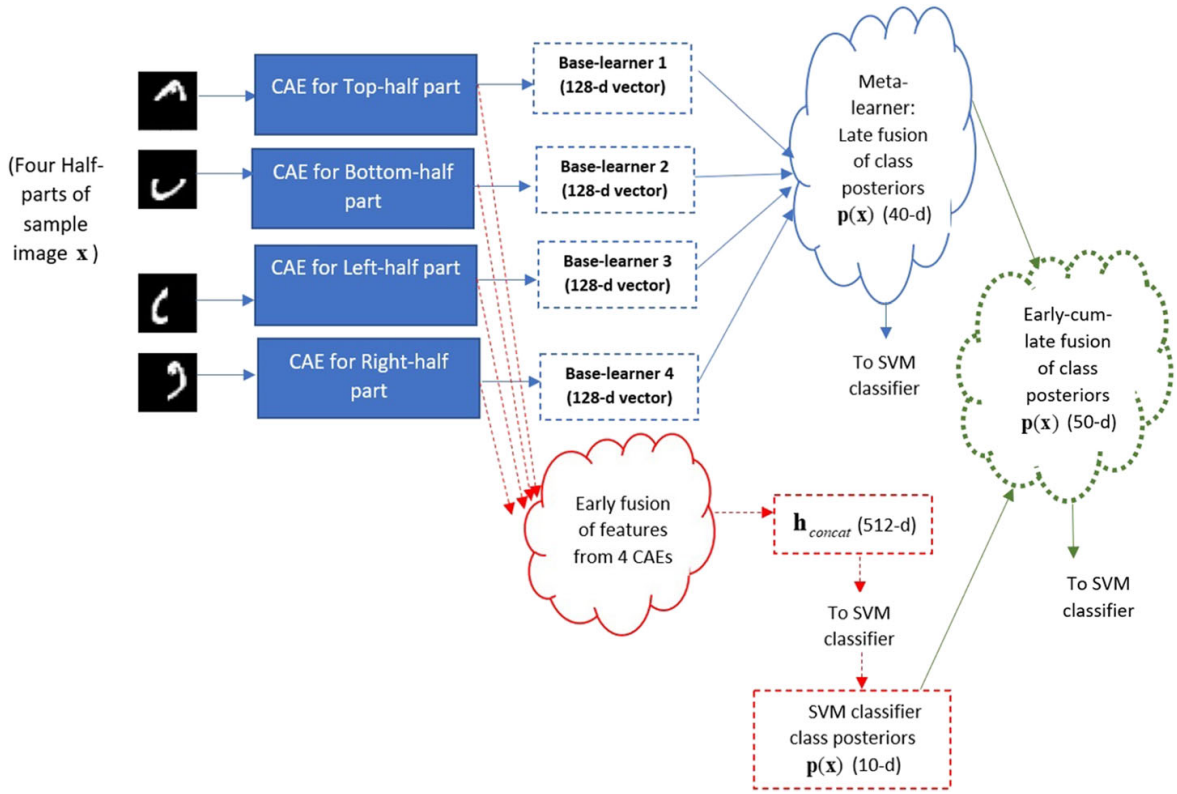


Fig. 1 Process flow of our part-learning scheme with early and late fusion

Table 1 CAE architecture (The rows highlighted in gray denote the encoder and the remaining rows indicate the decoder)

Layer	Kernel size	Output channels	Output shape	Stride	Activation function
Convolutional (2D)	**3x3**	16	28x28	1	ReLU
Max pooling (2D)	**2x2**	16	14x14	1	--
Convolutional (2D)	**3x3**	8	14x14	1	ReLU
Max pooling (2D)	**2x2**	8	7x7	1	--
Convolutional (2D)	**3x3**	8	7x7	1	ReLU
Max pooling (2D)	**2x2**	8	4x4	1	--
Convolutional (2D)	**3x3**	8	4x4	1	ReLU
Up-sampling (2D)	**2x2**	8	7x7	1	--
Convolutional (2D)	**3x3**	8	7x7	1	ReLU
Up-sampling (2D)	**2x2**	8	14x14	1	--
Convolutional (2D)	**3x3**	16	14x14	1	ReLU
Up-sampling (2D)	**2x2**	16	28x28	1	--

implementation of our code is provided online at [13]. The images were divided in turn into four half-parts: - top-half, bottom-half, left-half and right-half, with the other half of the image masked with black pixels, as shown in Fig. 2 for few MNIST samples. It is to be noted that the size of the part-images is same as that of the original images i.e. 28 × 28, only the relevant half is retained as it is, and the other half is kept masked with black pixels. The support vector machine (SVM) [4] is used in all the classification stages. The linear kernel was selected for ease of computations for all the base-learners that learn the CAE features, while the meta-learner SVM (for all late fusion schemes involving learning of probabilistic scores) has a non-linear rbf kernel; these settings being the result of several empirical trials. A grid search algorithm on SVM parameters was conducted for determining the optimal values of SVM parameters: (C for linear kernel; C and γ for rbf kernel). The timing characteristics of our implementation are summarized below.

a) Time taken for training each CAE = 18*50 = 900 s (18 s for each epoch)
b) Training time for SVM (linear kernel; C = 1) = 700 s
c) Training time for SVM (RBF kernel; C = 100, gamma = 0.1) = 700 s
d) Training time for Fully connected dense layers (20 epochs, batch size = 128, cross-entropy loss function) = 100 s

4.2 Discussion on results

In Table 2, we compare the performance of the proposed method, for different training scenarios, with the baseline comprising of CAE features extracted from the original (whole) image and trained on the SVM classifier with a linear kernel (accuracy = 97.59%). The first strategy is early fusion of CAE features extracted from the four image parts (512-dimension), the second is the late fusion of the ten probabilistic scores emanating from the four base-learners using the meta-learner SVM (40-dimension). The third strategy is the early-cum-late fusion in which the probabilistic scores emanating from the four part CAEs and their base-

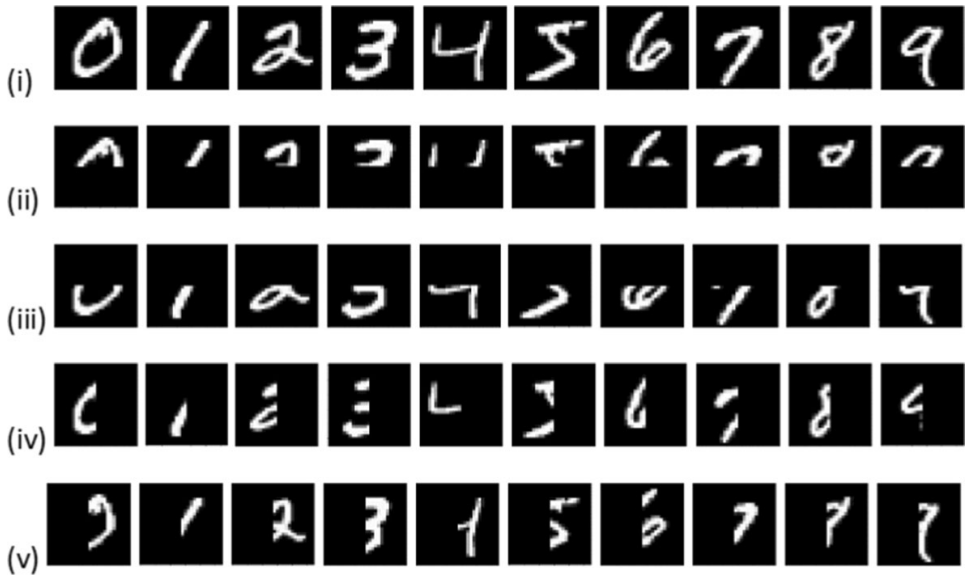


Fig. 2 MNIST English numeral dataset (*Left to Right*: 10 numerals ranging from 0 to 9) (i) Original numeral image (ii) Top-half part (iii) Bottom-half part (iv) Left-half part (v) Right-half part

Table 2 Test accuracy (%) for the MNIST dataset using proposed scheme involving half-part learning with convolutional auto-encoder representations

Method	Base classifier	Meta-learner	Image Part	Number of features to last stage SVM	Percentage Accuracy
CAE+Original	SVM (linear kernel)	–	Original (Whole) image	128	97.59%
CAE+Early fusion	SVM (linear kernel)	–	4 Parts	512	98.35%
CAE+Late fusion	SVM (linear kernel)	SVM (rbf kernel)	4 Parts	40	98.15%
CAE+Early-cum-late fusion	SVM (linear kernel)	SVM (rbf kernel)	4 Parts	50	98.6%
CAE+Early fusion	SVM (rbf kernel)	–	4 Parts	40	98.94%
CAE+Early-cum-late fusion	SVM (linear kernel)	SVM (rbf kernel)	4 Parts	50	99%
CAE+Early-cum-late fusion	SVM (linear kernel)	Fully connected dense layers (50–512–512-10)	4 Parts	50	99%
CAE+Early-cum-late fusion	Fully connected dense layers (50–1024–512-10)	Fully connected dense layers (50–1024–512-10)	4 Parts	50	98.57%

learners are fused with the ten probabilistic scores emanating from the SVM that learnt the early fusion features in the first step (50-dimensions). The results in Table 2 indicate that the third step involving early-cum-late fusion yields the highest classification accuracies with an improvement in error rate by over 1% over the conventional CAE trained on the whole image. The highest accuracies of 99% are achieved for the early-cum-late fusion scheme when the 50-dimensional concatenated feature vector is trained on the SVM with rbf kernel (meta-classifier). This constitutes one of the highest accuracies reported for the SVM classifier for MNIST. We also tested logistic regression instead of SVM for all experiments. However, the scores dipped by 1% as compared to the SVM results. The substitution with linear classifiers such as the nearest-neighbor showed a sharp dip in the results by 8%.

One particular reading in Table 2 corresponds to the use of fully-connected dense network (FCN) as the meta-learner instead of SVM. The results are same at 99% which implies that change in the meta-learning algorithm does not have an impact on the performance. A complete substitution of SVM by FCN also does not improve results, as observed from the last row in Table 2. Overall, the readings in Table 2 indicate that learning the image by-parts is indeed more beneficial than learning from the whole image at one go. The method also contributes to saving in computational expense since, at a time, only a small part of the image is processed, speeding up the inferencing process. Introducing distributed computing and parallel processors for learning from each image sub-part, in an independent and synchronized manner, is a possible future line of research that would showcase the benefits of our scheme to the maximum. Part-based information integration is thus advantageous in terms of both accuracy and speed.

Table 3 compares the classification results of some of the state-of-the-art methods for MNIST. As observed, our results compete favorably with the state of the art due as evident from the high classification scores achieved. Our highest accuracy (99%) is comparable to the scores of computationally complex ELM algorithm for shallow neural networks (99.1%) and LeNet (99.14%) deep model; our method performs better than all the other methods listed in Table 3 except for the very deep Highway network with 10 layers and 151 K parameters (Srivastava et al., 2015) that yields an accuracy of 99.5%. It is observed in general that the deeper the neural network the higher the performance. This inspires us to create computationally feasible deep versions of our model in future applications. Higher performance could be achieved by increasing the complexity of our model (at the cost of an increase in computational complexity) by using deeper CAEs. In its present form, our model offers an optimal tradeoff between computational complexity and high performance.

An ablation analysis was conducted for our best scheme of early-cum-late fusion (50-dimension), to understand the relative importance of the sub-components in our feature set. Our findings conclude that removal of a single 10-dimensional subset pertaining to any of the half-parts would result in a drop in performance by 0.4%. The removal of a pair of (10-dimension) features: (top-half, bottom-half) or the removal of (left-half, right-half) half-parts would bring down the scores by 0.3%. The test accuracies of individual half-parts when used as a single-set (10- dimension) feature for classification are: top-half only: 91.96%; bottom-half only: 89.54%; left-half only: 92.84%; right-half only: 93.16%. This reveals that the right-half part of the numeral image is more conclusive about the numeral category. Our ablation studies conclude that the individual parts by themselves are insufficient for recognition experiments, and removal of a single half-part from the computation process leads to a drop in performance.

Table 3 Comparison with other methods for the MNIST dataset: Test accuracy (%)

Method	Percentage accuracy
Interpretable Deep CNN [20] (Liu et al., 2018)	92.8%
Deep representation learning with target coding [53] (Yang et al., 2015)	85.5%
Very deep Highway network [35] (Srivastava et al., 2015)	99.5%
Adversarially robust neural network [30] (Schott et al., 2018)	99%
Neural Network + Extreme learning machine algorithm [26] (McDonnell et al., 2015)	99.1%
CNN-LeNet [51] (Xie et al., 2016)	99.14%
Regression model [11] (Hosmer et al., 2013)	92.5%
CNN [17] (Kuo, 2016)	98.4%
ResNet pretrained deep neural network [10] (He et al., 2016)	97.7%
Unsupervised STDP Spiking neural network [5] (Diehl and Cook., 2015)	97.7%
Auxiliary Classifier GAN [3] (Cheng et al., 2020)	96.25%
Patch-based MLP hidden state features + SVM [42] (Susan and Malhotra, 2020)	97.11%
Proposed method (with SVM-linear kernel as the base classifier)	99%

5 Conclusion

A novel sub-part learning scheme is presented in our work for classifying handwritten numeral images. The numeral image is divided into four half-parts: - top-half, bottom-half, left-half and right-half, with the other half of the image kept masked. Each half-part is trained on a convolutional auto-encoder that learns features from the training set in an unsupervised manner. Early fusion involves fusing the four 128-dimensional feature vectors to form a 512-dimensional early fusion vector that is trained on a classifier. The features extracted from each half-part are learnt individually by base classifiers. The probabilistic class posteriors are derived from the trained base classifiers for each sample image, and concatenated to form the late-fusion vector. The results of late fusion are compared with the early fusion of the features. A combination of both schemes called early-cum-late fusion is also investigated by fusing the 10 class posterior probabilities from the four base classifiers with the 10 class posterior probabilities associated with the early fusion features. SVM is used in all stages of classification. It is observed from the results on MNIST English numeral dataset that the early-cum-late fusion strategy performs the best. The Gaussian rbf kernel works best for the meta-learner SVM, while the base classifiers are best trained using the linear kernel of SVM. Ablation studies concluded that the right-half part is more conclusive of the image category than the other half-parts, and removal of any one half-part led to a drop in performance. Our method thus provides a computationally simple and effective methodology for learning numeral images by-parts and integrating the information learnt from individual parts. Future work

would involve creating deeper versions of our convolutional auto-encoder model, and finding ways to reduce the computational complexity that would arise as the depth of the network increases. The method also contributes to saving in computational expense since at a time, only a small part of the image is processed, speeding up the inferencing process. Introducing parallel processors for learning from each image sub-part, in an independent and synchronized manner, is a future line of research that would showcase the benefits of our scheme to the maximum. Part-based information integration is thus advantageous in terms of both accuracy and speed, rather than processing the whole image at a time. The concept can be extended to other problems of computer vision by focusing on and learning from parts of the image and exploring fusion strategies for integrating the scene information. The limitation of our approach is that though the scheme works well in the context of handwritten numeral images in which the content of each half-part is predictable, it may not work as well for other computer vision tasks such as object recognition where the objects are randomly placed in the scene. Localization of objects using bounding boxes followed by normalizing the detected object to a standard shape and size yields an image window to which the methodology in this paper could be applied, the investigation of which forms the future scope of our work.

Declarations

Conflict of interest The two authors are with the Department of Information Technology, Delhi Technological University, New Delhi, India.

References

1. Bengio Y (2012) Deep learning of representations for unsupervised and transfer learning. In Proceedings of ICML workshop on unsupervised and transfer learning, pp. 17–36
2. Bengio Y, Lamblin P, Popovici D, Larochelle H (2007) Greedy layer-wise training of deep networks. In Advances in neural information processing systems, pp. 153–160
3. Cheng K, Tahir R, Eric LK, Li M (2020) An analysis of generative adversarial networks and variants for image synthesis on MNIST dataset. Multimed Tools Appl:1–28
4. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297
5. Diehl PU, Cook M (2015) Unsupervised learning of digit recognition using spike-timing-dependent plasticity. Front Comput Neurosci 9:99
6. Ebrahimzadeh R, Jampour M (2014) Efficient handwritten digit recognition based on histogram of oriented gradients and SVM. International Journal of Computer Applications 104(9):10–13
7. Gao X, Zhou C, Chao F, Yang L, Lin C-M, Xu T, Shang C, Shen Q (2019) A data-driven robotic Chinese calligraphy system using convolutional auto-encoder and differential evolution. Knowl-Based Syst 182: 104802
8. Geng Q, Lu F, Huang X, Wang S, Cheng X, Zhou Z, Yang R (2018) Part-level car parsing and reconstruction from single street view. arXiv preprint arXiv:1811.10837
9. Hassan T, Khan HA (2015) Handwritten bangla numeral recognition using local binary pattern. In 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), pp. 1–4. IEEE, 2015.
10. 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
11. Hosmer Jr DW, Lemeshow S, Sturdivant RX (2013) Applied logistic regression 398. John Wiley & Sons
12. Hou B, Yan R (2019) Convolutional auto-encoder model for finger-vein verification. IEEE Trans Instrum Meas
13. https://github.com/JMalhotra7/Learning-image-by-parts-using-early-and-late-fusion-of-auto-encoder-features [Last accessed on 27th Dec 2020]
14. Izonin I, Tkachenko R, Kryvinska N, Tkachenko P (2019) Multiple linear regression based on coefficients identification using non-iterative SGTM Neural-Like Structure. In International Work-Conference on Artificial Neural Networks, pp. 467–479. Springer, Cham

15. Khan A, Sohail A, Zahoora U, Qureshi AS (2020) A survey of the recent architectures of deep convolutional neural networks. Artificial Intelligence Review 53(8):5455–5516
16. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pp. 1097–1105
17. Kuo C-CJ (2016) Understanding convolutional neural networks with a mathematical model. J Vis Commun Image Represent 41:406–413
18. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324
19. Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE (2017) A survey of deep neural network architectures and their applications. Neurocomputing 234:11–26
20. Liu X, Wang X, Matwin S (2018) Interpretable deep convolutional neural networks via meta-learning. In 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–9. IEEE
21. Loey M, El-Sawy A, EL-Bakry H (2017) Deep learning autoencoder approach for handwritten arabic digits recognition. arXiv preprint arXiv:1706.06720.
22. Lorenz D, Bereska L, Milbich T, Ommer B (2019) Unsupervised part-based disentangling of object shape and appearance. arXiv preprint arXiv:1903.06946
23. Malinowski M, Doersch C (2018) The visual QA devil in the details: The impact of early fusion and batch norm on clevr. arXiv preprint arXiv:1809.04482
24. Malowany D, Guterman H (2020) Biologically inspired visual system architecture for object recognition in autonomous systems. arXiv preprint arXiv:2002.03472
25. Masci J, Meier U, Cireşan D, Schmidhuber J (2011) Stacked convolutional auto-encoders for hierarchical feature extraction. In International Conference on Artificial Neural Networks, pp. 52–59. Springer, Berlin, Heidelberg
26. McDonnell MD, Tissera MD, Vladusich T, van Schaik A, Tapson J (2015) Fast, simple and accurate handwritten digit classification by training shallow neural network classifiers with the 'extreme learning machine'algorithm. PLoS One 10(8):e0134254
27. Palvanov A, Cho YI (2018) Comparisons of Deep Learning Algorithms for MNIST in Real-Time Environment. Int J Fuzzy Log Intell 18(2):126–134
28. Park J, Lee G, Kim E, Lim J, Kim S, Yang H, Lee M, Hwang S (2010) Automatic detection and recognition of Korean text in outdoor signboard images. Pattern Recogn Lett 31(12):1728–1739
29. Safdari R, Moin M-S (2016) A hierarchical feature learning for isolated Farsi handwritten digit recognition using sparse autoencoder. In 2016 Artificial Intelligence and Robotics (IRANOPEN), pp. 67–71. IEEE, 2016.
30. Schott L, Rauber J, Bethge M, Brendel W (2018) Towards the first adversarially robust neural network model on MNIST. arXiv preprint arXiv:1805.09190
31. Shi M, Fujisawa Y, Wakabayashi T, Kimura F (2002) Handwritten numeral recognition using gradient and curvature of gray scale image. Pattern Recogn 35(10):2051–2059
32. Snoek, Cees GM, Worring M, Smeulders AWM (2005) Early versus late fusion in semantic video analysis. In Proceedings of the 13th annual ACM international conference on Multimedia, pp. 399–402. ACM
33. Špaňhel, Jakub, Jakub Sochor, Roman Juránek, Adam Herout, Lukáš Maršík, and Pavel Zemčík (2017) Holistic recognition of low quality license plates by cnn using track annotated data. In 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–6. IEEE, 2017.
34. Spratling MW (2017) A hierarchical predictive coding model of object recognition in natural images. Cognitive computation 9(2):151–167
35. Srivastava, Rupesh K, Greff K, Schmidhuber J (2015) Training very deep networks. In Advances in neural information processing systems, pp. 2377–2385
36. Sung J, Bang S-Y, Choi S (2006) A Bayesian network classifier and hierarchical Gabor features for handwritten numeral recognition. Pattern Recogn Lett 27(1):66–75
37. Susan S, Devi KMR (2019) Text area segmentation from document images by novel adaptive thresholding and template matching using texture cues. Pattern Analysis and Applications:1–13
38. Susan S, Kadyan P (2013) A supervised fuzzy eye pair detection algorithm. In 2013 5th International Conference and Computational Intelligence and Communication Networks, pp. 306–310. IEEE
39. Susan S, Kakkar G (2015) Decoding facial expressions using a new normalized similarity index. In 2015 Annual IEEE India Conference (INDICON), pp. 1–6. IEEE
40. Susan S, Keshari J (2019) Finding significant keywords for document databases by two-phase maximum entropy partitioning. Pattern Recogn Lett 125:195–205
41. Susan S, Malhotra J (2019) CNN Pre-initialization by minimalistic part-learning for handwritten numeral recognition. International Conference on Mining Intelligence and Knowledge Exploration:320–329. Springer, Cham

42. Susan S, Malhotra J (2020) Learning interpretable hidden state structures for handwritten numeral recognition. In 2020 4th International Conference on Computational Intelligence and Networks (CINE), pp. 1–6. IEEE
43. Susan S, Malhotra J (2020) Recognising devanagari script by deep structure learning of image quadrants. DESIDOC J Libr Inf Technol 40(5):268–271
44. Susan S, Singh V (2011) On the discriminative power of different feature subsets for handwritten numeral recognition using the box-partitioning method. In 2011 Annual IEEE India Conference, pp. 1–5. IEEE
45. Susan S, Ranjan R, Taluja U, Rai S, Agarwal P (2019) Neural net optimization by weight-entropy monitoring. In Computational intelligence: theories, applications and future directions-volume II, pp. 201–213. Springer, Singapore
46. Tkachenko R, Izonin I (2018) Model and principles for the implementation of neural-like structures based on geometric data transformations. In International Conference on Computer Science, Engineering and Education Applications, pp. 578–587. Springer, Cham
47. Tkachenko R, Tkachenko P, Izonin I, Tsymbal Y (2018) Learning-based image scaling using neural-like structure of geometric transformation paradigm. In Advances in Soft Computing and Machine Learning in Image Processing, pp. 537–565. Springer, Cham
48. Wang M, Chen Y, Wang X (2014) Recognition of handwritten characters in chinese legal amounts by stacked autoencoders. In 2014 22nd International Conference on Pattern Recognition, pp. 3002–3007. IEEE
49. Wang Y, Xie Z, Xu K, Dou Y, Lei Y (2016) An efficient and effective convolutional auto-encoder extreme learning machine network for 3d feature learning. Neurocomputing 174:988–998
50. Wang Y, Li F, Sun H, Li W, Cheng Z, Wu X, Wang H, Wang P (2020) Improvement of MNIST Image Recognition Based on CNN. In IOP Conference Series: Earth and Environmental Science 428(1):012097. IOP Publishing
51. Xie L, Wang J, Wei Z, Wang M, Tian Q (2016) Disturblabel: regularizing cnn on the loss layer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4753–4762
52. Xu X (2013) Receipt digitizing method for retail customers. U.S. Patent Application 13/507,291, filed March 7, 2013.
53. Yang S, Luo P, Loy CC, Shum KW, Tang X (2015) Deep representation learning with target coding. In Twenty-Ninth AAAI Conference on Artificial Intelligence
54. Yang Z-X, Tang L, Zhang K, Wong PK (2018) Multi-view cnn feature aggregation with elm auto-encoder for 3d shape recognition. Cogn Comput 10(6):908–921

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

Lower Fin Modulation Analysis for a Novel 5nm Top Bottom Gated Junctionless FinFET for improved performance

Lakshya Gangwani
Department of Applied Physics
Delhi Technological University
New Delhi, India
lakshyagangwani1@gmail.com

Ruchika Chakravarti
Department of Applied Physics
Delhi Technological University
New Delhi, India
ruchika.chakravarti1@gmail.com

Rishu Chaujar
Department of Applied Physics
Delhi Technological University
New Delhi, India
chaujar.rishu@dtu.ac.in

2021 Devices for Integrated Circuit (DevIC) | 978-1-7281-9955-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/DevIC50843.2021.9455834

***Abstract*— In this paper, lower fin modulation is performed for a Novel 5nm Top Bottom Gated Junctionless Fin shaped Field Effect Transistor (FinFET). The analysis is done for fin lengths varying from 3nm to 12nm with and without source, drain contacts present at the bottom fin and compared with the Conventional Junctionless FinFET structure. The Novel Top Bottom Gated Junctionless FinFET with 3nm Lower Fin length and contacts on bottom fin shows 197 times improvement in switching ratio as compared to the Conventional Junctionless FinFET at the same gate length. The results reveal that the device shows enhancement in terms of switching ratio and allows better gate controllability when stacked against the Conventional Junctionless FinFET for various fin lengths**.

Keywords— Top Bottom Gated, FinFET, Fin Modulation, Switching Ratio, Junctionless

I. INTRODUCTION

Technology and along with it our way of life is ever evolving. Innovation is happening in all fields of research with greater importance being given to portability, efficiency and speedy operation. Keeping in line with this trend the semiconductor and solid state technology field has also experienced development with greater impetus for designing compact and low power consumption devices. Major players in the industry are already adapting and catering to these needs [1, 2]. These innovations entail scaling down of devices with more integration and lower power consumption. But during the process of scaling down we face certain obstacles. These obstacles take the shape of Short Channel Effects (SCEs) which reflect inefficient device performance and delays. These include leakage current, Hot Carrier Effects, Velocity Saturation and Drain Induced Barrier Lowering (DIBL) [3]. New improvements in technologies like SOI and FinFET are being made to address these effects. In this work we have performed lower fin dimension modulation for a Novel 5nm Top Bottom Gated (TBG) Junctionless (JL) Fin shaped Field Effect Transistor (FinFET). The analysis is done for fin lengths varying from 3nm to 12nm with and without source, drain contacts present at the bottom fin and compared with the Conventional Junctionless (JL) FinFET structure to record the improvements in switching ratio, TGF, transconductance and transfer characteristics thus reducing Short Channel Effects (SCEs).

II. DEVICE STRUCTURE AND METHODOLOGY

Fig. 1 depicts the Novel 5nm Top Bottom Gated Junctionless FinFET device having 5nm lower fin with source, drain contacts. The gates on both sides of the fins enable the device to perform better in terms of output current and switching ratio. It also helps in enhancing the control of gate over the device as compared to the Conventional Junctionless (JL) FinFET.

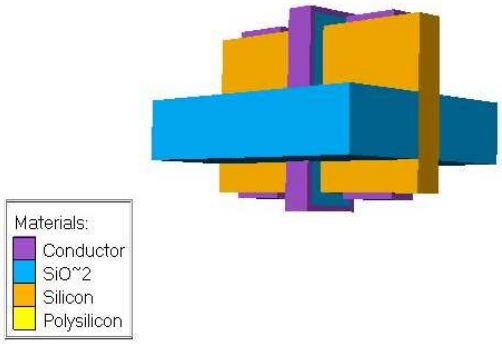


Fig. 1 Novel Top Bottom Gated Junctionless FinFET Structure with 5nm Lower Fin

In contrast to the Novel Top Bottom Gated (TBG) Junctionless (JL) FinFET, the Conventional Junctionless (JL) FinFET has gate above the substrate only due to the absence of the lower fin. As a result, it has less gate controllability over the device.

The device structures of both the Novel Top Bottom Gated (TBG) Junctionless (JL) and Conventional Junctionless (JL) FinFET have 5nm gate length, 35nm device length and 20nm source, drain length. For the case of Novel TBG JL FinFET, the device height varies from 31nm for 3nm lower fin to a maximum of 40nm for 12nm lower fin. Whereas on the other hand, the Conventional JL FinFET has fixed device height of 25nm due to absence of the lower fin. The additional lower fin in Novel TBG JL FinFET with contacts helps in enhancing the overall device performance in comparison to the Conventional JL FinFET.

Both the Novel TBG JL FinFET and Conventional JL FinFET are n-type devices with 5.0×10^{18} (cm^{-3}) uniform doping for source, drain and channel regions.

The Novel TBG JL FinFET has a gate length of 5nm and the lower fin dimension is varied from 3nm to 12nm. Also source, drain contacts are placed on the lower fin for all the device lengths and their results are analyzed and compared with the Conventional JL FinFET. The device has 2nm thickness of gate oxide for both the Conventional JL FinFET and Novel TBG JL FinFET [3].

Larger fin heights implies more gate control over the channel as the gate contact is covering significantly larger area over the channel for the same fin width. Thus it improves the output current and transconductance. But on the other hand, larger gate area leads to more surface scattering and thus leads to delays in switching. Due to this, smaller fin heights

978-1-7281-9955-9/21/$31.00 ©2021 IEEE
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:30:35 UTC from IEEE Xplore. Restrictions apply.

have better switching ratio as compared to larger ones. Also placing the contacts on the lower fin provides another path for the carriers to flow, thereby improving the output current considerably as compared to the structures without the lower fin contacts.

The three-dimensional (3D) device simulations for all the structures having different fin lengths is done on Silvaco Atlas TCAD Simulator using DeckBuild and TonyPlot tools [4]. The device simulation in Silvaco Atlas is divided into five main steps, videlicet the Specification of the Device Structure, Specification of Material Models, Selecting Numerical Models, Specification of the Solution and Analyzing the Results. Structure Specification basically involves defining the mesh grid for the device by using the mesh points and spacing, defining various regions in device, electrode and doping definition for the device. Material Model Specification involves defining the materials to be used in the device and choosing the physical models for the device. Numerical Method Selection is used to decide the mathematical models to be used for device simulation [4]. The device temperature was kept at T=300K for the analysis on Silvaco TCAD. Solution Specification and Result Analysis uses the extract and solve for running the simulations and extracting the required parameters.

Table 1. STRUCTURAL PARAMETERS SPECIFICATIONS

Parameters	**Dimensions of Simulated Devices**	
	Novel TBG JL FinFET	***Conventional JL FinFET***
Gate Length	5nm	5nm
Upper Fin Height	12nm	12nm
Lower Fin Height	3nm to 12nm	0nm
Fin Width	5nm	5nm
Effective Oxide Thickness	2nm	2nm
Substrate Width	5nm	5nm
Source,Drain Length	20nm	20nm

III. SIMULATED RESULTS AND DISCUSSIONS

Several device performance parameters namely Transconductance (g_m), Output Conductance (g_d), Transconductance Generation Factor (TGF) (g_m/I_d), Early Voltage (V_A) (I_d/g_d) are germane to the discussion while analyzing the Analog Performance of FinFET devices [5].

The ratio of ON state current I_{ON} and OFF state current I_{OFF} of a switch is a design parameter and is known as switching ratio. The ON current is recorded when the device allows current to flow through it and hence is 'ON' while the OFF current is ideally null [6]. But when we employ solid state technologies like FinFET to perform switching operations we do not obtain zero as OFF state current but rather it has a very small value and we get this value when $V_{gs} = 0$ Volts and $V_{ds} = V_{dd}$. As we scale down the device this small value can increase many folds and so becomes more important and is a reflection of Short Channel Effects (SCEs) and leakage in the device. Hence we try to minimize the OFF current which leads to high switching ratio and lower the losses.

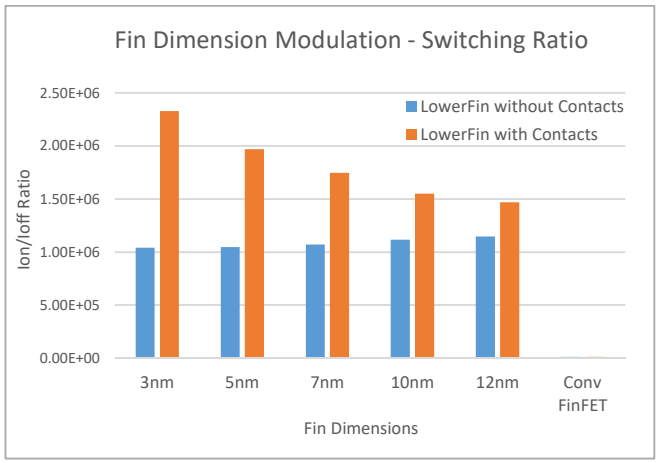


Fig. 2 Switching Ratio Variation with Lower Fin Dimensions for Novel TBG JL FinFET

Fig. 2 shows the switching ratio for different lower fin dimensions varying from 3nm to 12nm [7]. The ratios are given for both with contacts and without source, drain contacts placed on lower fin for the Novel Top Bottom Gated Junctionless FinFET. The trend can be observed that the switching ratio decreases as we increase the lower fin length for the device with contacts whereas the switching ratio increases with increase in lower fin length in the case of no contacts. Further it can be seen that the maximum switching ratio is obtained for the configuration of 3nm lower fin length with contacts. We can also observe that the switching ratio for all fin lengths of Novel Top Bottom Gated Junctionless FinFET is higher than Conventional JL FinFET. Thus indicating improvement.

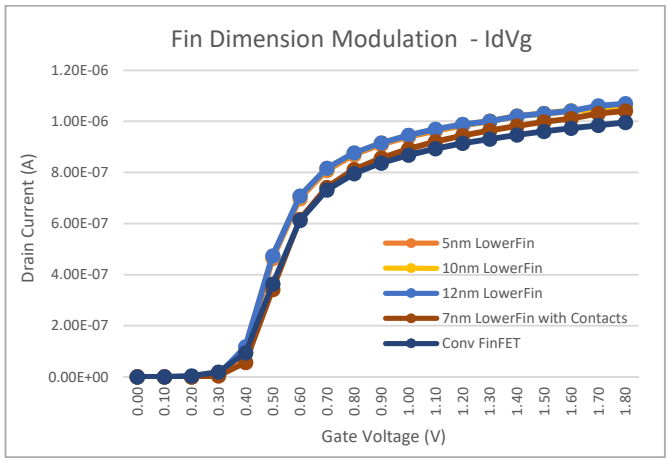


Fig. 3 Transfer Characteristics for Novel TBG JL FinFET for different lower fin lengths

Transfer Characteristics essentially shows the variation of drain current or the output current with the gate voltage (input voltage) [7]. **Fig. 3** depicts the $I_{ds} - V_{gs}$ curves for different lower fin lengths ranging from 3nm to 12nm with and without source, drain contacts placed on lower fin for the Novel Top Bottom Gated Junctionless FinFET. It is observed that 12nm lower fin without contacts shows increased output current for the same gate voltages as compared to the other fin lengths. All the fin lengths from 3nm to 12nm give better results in terms of output current as compared to the Conventional Junctionless FinFET.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:30:35 UTC from IEEE Xplore. Restrictions apply.

Thus 12nm lower fin Novel TBG JL FinFET structure without contacts gives best performance in terms of $I_{ds} - V_{gs}$ closely followed by 10nm lower fin Novel TBG JL FinFET structure. Higher drain current for 12nm lower fin depicts that gate has enhanced control over the device and V_{ds} effect on output is less, thus helping in improvement of device performance.

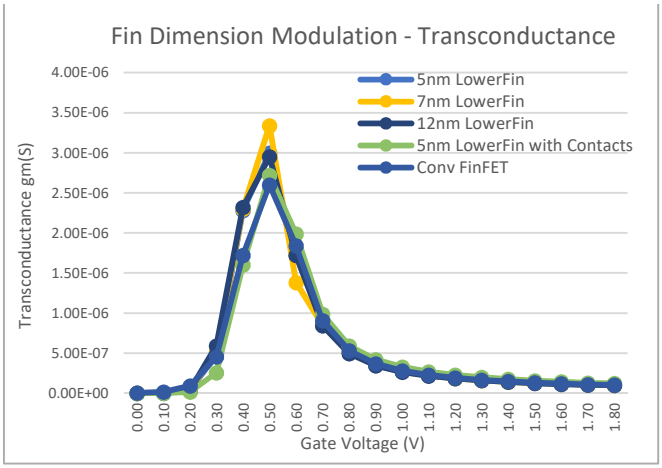


Fig. 4 Transconductance variation with fin dimension for Novel TBG JL FinFET

The linear gain of a device decides the input conductance (g_m or also known as transconductance). The transconductance is thus proportional to it. It depicts the fluctuations in output current with changes in input voltage V_{gs} under the condition where drain voltage V_{ds} is constant [8].
Transconductance is given by:

$$g_m = \frac{\partial I_d}{\partial V_g}$$

So we easily calculate g_m by the slope of $I_{ds} - V_{gs}$ plot [5]. **Fig. 4** shows the transconductance (g_m) for different lower fin lengths ranging from 5nm to 12nm with and without source, drain contacts placed on lower fin for the Novel Top Bottom Gated Junctionless FinFET with gate bias varying from 0 to 1.8 Volts. And the same is also plotted for Conventional Junctionless FinFET for comparison. We can see that we get highest transconductance value for 7nm lower fin without contacts as compared to the other fin lengths at the same gate voltage of 0.5 Volts. All the fin lengths from 5nm to 10nm give improved g_m values as compared to Conventional Junctionless FinFET.

Transconductance Generation Factor (TGF) is a measure of the efficiency of a device. It conveys how adequately the current is used to achieve a certain transconductance [8].
It is calculated as:

$$TGF = \frac{g_m}{I_d}$$

The merit of high TGF is the capacity for functioning of device at low voltages and therefore this leads to more efficient utilization and performance [5].
Fig. 5 shows the variation of TGF for different lower fin lengths of 5nm and 12nm with and without source, drain contacts placed on lower fin for the Novel Top Bottom Gated Junctionless FinFET. Again we have presented the same for Conventional Junctionless FinFET to compare performance.

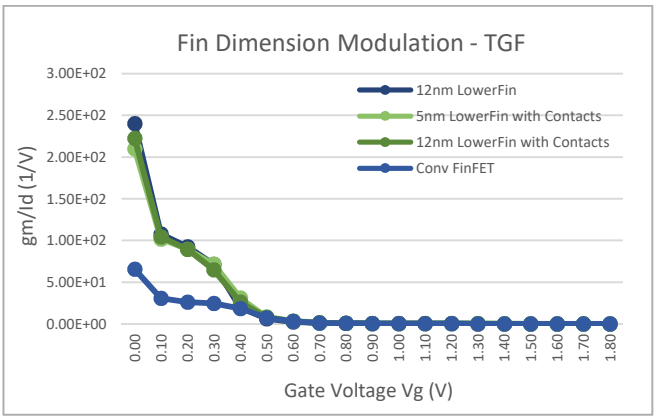


Fig. 5 Variation of Transconductance Generation Factor for different lower fin lengths for Novel TBG JL FinFET

We observe that TGF is highest at 0V gate supply and has maximum value 12nm lower fin length without contacts and shows quiet improvement over Conventional JL FinFET. Thus indicating attainment of superior parameters [7].

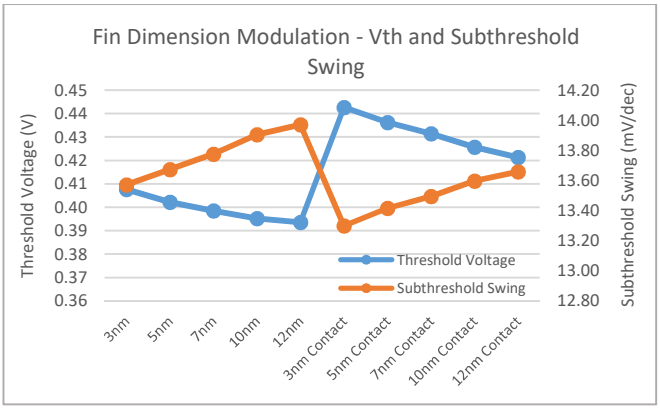


Fig. 6 Variation of Threshold Voltage and Subthreshold Swing for Novel TBG JL FinFET for varying lower fin lengths

V_{th} or the Threshold Voltage is the voltage required at the gate terminal to turn ON the device. At this voltage, the current in the device begins to appreciably [6].
Consequently Subthreshold Swing or also known as the Slope Factor is the inverse of the slope of log (I_d) vs. V_{gs} plot in the Subthreshold region. Conventional MOSFET have their Subthreshold Swing around 60mV/decade [7].
Fig. 6 shows the variation of threshold voltage and the Subthreshold Swing for different lower fin lengths with and without source, drain contacts for the Novel TBG JL FinFET structure.
It can be clearly inferred from the plot that threshold voltage decreases with increase in fin length and the lowest threshold voltage is achieved for 12nm lower fin without contact at V_{th} = 0.39V. Also threshold voltage for devices having lower fin with source, drain contacts is comparatively higher than their counterparts without the source, drain contacts. Thus lower fin without contacts leads to less threshold voltages, thereby low power consumption and device ON at lower voltages [9].
As for the Subthreshold Swing, it increases with increase in lower fin length and the lowest Subthreshold Swing is observed for the 3nm lower fin Novel TBG JL FinFET with contacts. The Subthreshold Swing for 3nm Lower Fin Novel TBG JL FinFET with contacts is S.S=13.29 mV/decade. It is

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:30:35 UTC from IEEE Xplore. Restrictions apply.

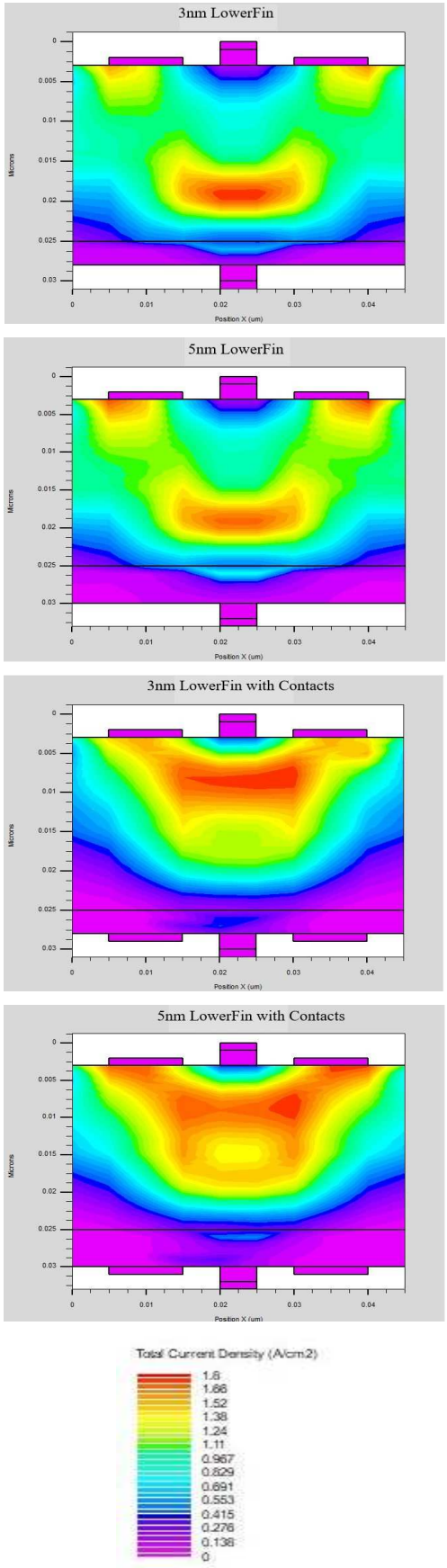


Fig. 7 Current Density Contours for Novel TBG JL FinFET for various lower fin dimensions

Also recognized that devices having lower fin with contacts have lower values for Subthreshold Swing as compared to their equivalent structures without contacts. Minimizing the Subthreshold Swing helps in reducing the leakage current up to a huge extent. The device has Subthreshold Swing less than 60mV/decade for all the fin lengths [3].

Fig. 7 shows the current density contours for the Novel Top Bottom Gated Junctionless FinFET at various lower fin lengths with and without source, drain contacts on lower fins. The current density primarily depicts the current flowing through the device per unit area. It is clearly visible from the contours that Novel TBG JL FinFET devices having lower fin contacts have higher current densities as compared to their complement structures without the contacts. This is due to the fact that contacts on lower fins provide another junction for the carriers to flow from and to, thus leading to increase in the current densities [10].

The 3nm Lower Fin Novel TBG JL FinFET with contacts shows the maximum current densities out of all the simulated devices, thus leading to improved switching ratio [11]. But higher current densities have tradeoffs as well. Higher current densities can lead to melting or heating of the insulating material and can also lead to electromigration.

IV. Conclusion

In this paper we have performed the lower fin modulation for a Novel 5nm Top Bottom Gated (TBG) Junctionless (JL) Fin shaped Field Effect Transistor (FinFET). The 3nm lower fin Novel TBG JL FinFET device showed considerably improved switching ratio (197 times over Conventional FinFET) and lowest Subthreshold Swing 13.29mV/decade is observed for the 3nm lower fin length as well. Whereas 7nm fin length without contacts had best I_dV_g characteristics and 12nm lower fin length without contacts has most improvement (2.65 times improvement over Conventional FinFET) for the Transconductance Generation Factor in comparison to other fin dimensions [12]. Thus the simulation results reveal that all the fin lengths present enhancement over the Conventional JL FinFET and the device can be utilized for high performance analog use [7].

References

[1] TSMC. Leading Edge Technology. http://www.tsmc.com/english/dedicate-Foundry/technology/16nm.htm. Accessed 3rd February 2021

[2] Semiconductor Engineering. IBM, Intel and TSMC Roll out FinFETs. http://semiengineering.com/ibm-intel-and-tsmc-roll-outfinfets. Accessed 3rd February 2021

[3] M. P. Kashyap and R. Chaujar, "Temperature investigation of a novel 3nm TF-Bulk FinFET for improved performance," 2020 IEEE 20th International Conference on Nanotechnology (IEEE-NANO), Montreal, QC, Canada, 2020, pp. 382-387, doi: 10.1109/NANO47656.2020.9183594.

[4] I. SILVACO, "ATLAS User's Manual," Santa Clara, CA, Ver, vol. 5, 2011.

[5] N. Kaur, M. Rattan & S.S Gill, "Design and optimization of novel shaped FinFET". *Arab J Sci Eng* 44, 3101–3116 (2019). https://doi.org/10.1007/s13369-018-3428-3

[6] S. Tripathi, N. Gupta, P. Tripathi & R. Chaujar. (2019). "GaN Silicon-on-Insulator (SOI) N-Channel FinFET for high-performance low power applications". 1-4. 10.1109/NMDC47361.2019.9084011.

[7] A. Kumar, N. Gupta & R. Chaujar, "Analysis of novel transparent gate recessed channel (TGRC) MOSFET for improved analog behaviour".

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:30:35 UTC from IEEE Xplore. Restrictions apply.

Microsyst Technol **22,** 2665–2671 (2016). https://doi.org/10.1007/s00542-015-2554-z

[8] M Singh, S Mishra, S S Mohanty, and G P Mishra. "Performance analysis of SOI MOSFET with rectangular recessed channel" Advances in Natural Sciences: Nanoscience and Nanotechnology, vol. 7, no. 1, 2016. doi:10.1088/2043-6262/7/1/015010

[9] A. Bughio, S. D. Guerrieri, F. Bonani, and G. Ghione, "Multi-Gate FinFET Mixer Variability Assessment through Physics-Based Simulation," IEEE Electron Device Letters, vol. 38, no. 8, pp. 1004-1007, 2017.

[10] S.Y. Kim, J. Seo, Y.J. Yoon, G.Yoo, Y.J. Kim, H. Eun et al. (2014). "Design and Analysis of Sub-10 nm Junctionless Fin-Shaped Field-Effect Transistors". Journal of Semiconductor Technology and Science. 14. 508-517. 10.5573/JSTS.2014.14.5.508.

[11] M. Tan & R. Ismail (2007). "Modeling of Nanoscale MOSFET Performance in the Velocity Saturation Region". Elektrika: Journal of Electrical Engineering.

[12] A. Kumar, B. Tiwari, S. Singh, M. M. Tripathi, and R. Chaujar, "Radiation Analysis of N-Channel TGRC-MOSFET: An X-Ray Dosimeter," IEEE Transactions on Electron Devices, vol. 65, no. 11, pp. 5014-5020, 2018.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:30:35 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Trends in Electronics and Informatics (ICOEI).
IEEE Xplore Part Number:CFP21J32-ART; ISBN:978-1-6654-1571-2

Machine Learning and IoT based solutions for detection of arrhythmia using ECG signals

Vinod Kumar
Dept. of CSE
Delhi Technological University
New Delhi, India
vinod_k@dtu.ac.in

Awishkar Ghimire
Dept. of CSE
Delhi Technological University
New Delhi, India
awishkar.ghimire@gmail.com

Hitesh Kumar Hoon
Dept. of CSE
Delhi Technological University
New Delhi, India
hitesh.hoon101@gmail.com

2021 5th International Conference on Trends in Electronics and Informatics (ICOEI) | 978-1-6654-1571-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICOEI51242.2021.9452907

***Abstract*—The World Health Organization recognizes heart diseases as being the leading cause of death globally bringing about 17.9 million deaths annually. Arrhythmia is one of the most common heart diseases and hence it is of quintessential importance that there be computational systems that are able to automatically diagnose arrhythmia. ECG is a way of detecting the heart beat using electrical changes in the skin caused by the cardiac cycle. ECG can be used to automatically diagnose various heart diseases including arrhythmia. It is very well known that artificial intelligence is known as the 4th industrial revolution and various subfields of artificial intelligence such as machine learning and deep learning have been used to solve problems across a wide range of fields. Also, IoT is a technology that is increasingly being used to solve various real world problems computationally. In this paper we have reviewed the most impactful and noteworthy research papers in recent years that deal with the subject matter of arrhythmia detection using ECG signals based on machine learning and IoT.**

***Index Terms*—ECG, Arrhythmia, Machine Learning, Deep Learning, IoT, Long Short Term Memory(LSTM), Support Vector Machine(SVM), MIT-BIH Arrhythmia database.**

I. Introduction

Arrhythmia can be defined as a problem regarding the rate of rhythm of the heart beat of a person. It means that the heart is beating at a faster rate or a slower rate than normal. It is also characterized by an irregular heartbeat. When the heart beats too fast the term given to that particular type of arrhythmia is tachycardia whereas when the heart beats faster than average the term given to it is bradycardia. One of the most important types of arrhythmia is atrial fibrillation. Atrial fibrillation is a type of quivering, rapid heartbeat that can lead to heart attacks, strokes, blood clots and heart palpitation. Many factors can cause arrhythmia such as heart attack, coronary artery disease, smoking, stress, congenital heart defects among other causes. Treatment to restore the normal heart rhythm of an arrhythmia patient includes medicines, an implantable cardioverter-defibrillator as well as surgery in the most extreme cases. Symptoms of arrhythmia can include heart palpitations, lightheadedness, passing out, shortness of breath. It can cause heart failures, strokes and death in the most extreme cases. Arrhythmia can be broadly classified into 4 main classes namely, bradyarrhythmias, supraventricular tachycardias, ventricular arrhythmias, and extra beats[14].

According to the World Health Organization, heart diseases also known as cardiovascular diseases is the leading cause of death globally, bringing about 17.9 million deaths per year. Cardiovascular diseases are a group of disorders of the heart and various blood vessels around the heart and include diseases like coronary artery disorder, cerebrovascular disease. Among all the deaths caused due to cardiovascular diseases four out of five such deaths are due to heart attacks and strokes, and about one third of the deaths happen prematurely to people below the age of 70 [3]. All cardiovascular disease can cause different types of arrhythmia. Specifically coronary artery disease can cause a certain type of a very dangerous arrhythmia called the atrial fibrillation. Thus it is of paramount importance to create systems specifically computational systems that can accurately diagnose arrhythmia in patients.

The most common and effective way of diagnosing arrhythmia is through the use of ECG. Electrocardiogram (ECG) is a graph in which the y axis is the voltage and x axis is the time. It is a voltage vs time graph. The voltage recorded in the ECG is driven by the electrical activity of the heart. The voltage is recorded using electrodes placed on the skin. These electrodes can detect the electrical changes in the skin that is caused by the cardiac cycle(heartbeat). The voltage change is caused by the depolarization followed by repolarization of the cardiac muscle. There are different types of ECG. The most common one is the 12 lead ECG which consists of 10 different electrodes placed on the skin of the patient. The electrodes are placed in different parts of the body which includes the limbs and the chest. The different electrodes detect the electrical activity of the heart in different angles giving different perspectives of the same cardiac cycle. A cardiac cycle in an ECG consists of three main components, namely the P wave, the QRS complex and the T wave. The P wave illustrates the depolarization of the atria, the QRS complex illustrates the depolarization of the ventricles and the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:29:02 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Trends in Electronics and Informatics (ICOEI).
IEEE Xplore Part Number:CFP21J32-ART; ISBN:978-1-6654-1571-2

T wave illustrates the ventricular repolarization . Usually for basic heart monitoring 1 lead ECG is used which consists of only two electrodes and it detects the electrical activity of the heart from only a single angle of view. Lead 1 ECG has been gaining popularity as it can be used easily and is cheaper. It is being extensively used for basic heart monitoring .

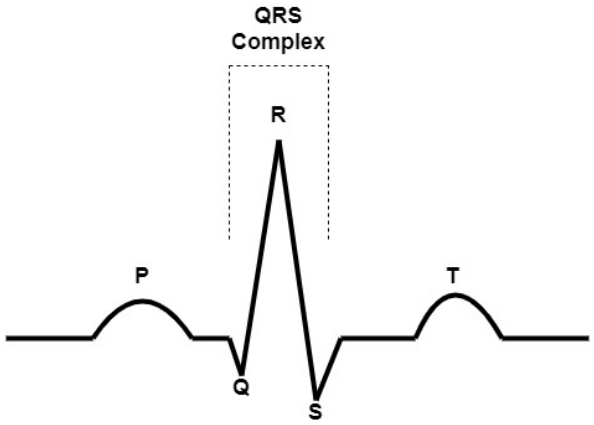


Fig. 1. Representation of the ecg signal of a regular heartbeat

Many research has been done in the field of automated diagnosis of different arrhythmias using machine learning and artificial intelligence on ECG data. Machine learning is a subfield of artificial intelligence and artificial intelligence has been successfully used to solve a wide array of tasks. Machine learning has been and is being used extensively in various businesses particularly to improve the growth of a business[6]. Machine learning and artificial intelligence is also being successfully used in the healthcare industry to combat various different diseases such as alzheimer's, parkinson's and also COVID-19 [22][21][7]. Artificial intelligence and machine learning is also being used successfully in various disciplines of engineering such as civil engineering, electrical engineering and such. It is being used to crack detection in roads and also in niche fields such as urban planning [13][12]. It is also being used for the detection of illegal hacking [8]. The current world is heavily reliant on the field of artificial intelligence and machine learning and plethora of research has been conducted on various fields using this technology. So it is of no surprise that this technology can also be used in the diagnosis of heart disease. It is extensively being used for diagnosing arrhythmia using ECG signals and this paper attempts to review some of the most notable and worthy papers in this particular subject.

Internet of things(IoT) is another technology that is sweeping the current world. It finds applications in various different areas such as healthcare, business, communication systems and other fields[5]. IoT helps in making environment smarter and improve living standards, for example, with the help of IoT people can make their home smart, generate smart energy, build smart cities[20]. The internet of things is a technology that can be broadly defined as a network of intelligent devices working together in harmony to solve a particular problem. The architecture mainly consists of sensors, computational processors, microcontrollers communicating over a network. Many IoT architectures also incorporate cloud based strategy. IoT has also been extensively used in the detection of arrhythmia using ECG signals. Recent research work are performed in IoT based monitoring systems for health care which proposes management of big data using IoT environment helping in finding big data based solutions for various problems in health care[17]. It is mainly incorporated in remote monitoring of patients and is used for real time ECG analysis. In this paper we attempt to review the most recent advancements in the use cases of IoT for arrhythmia detection.

II. Related Works

One of the most impactful papers on ECG arrhythmia detection has been by Rajpurkar et al. Rajpurkar et al [18] have developed a machine learning algorithm that has been shown to work better than board certified cardiologist in detecting a wide range of different arrhythmias using an ECG recorded from a single lead ECG machine. They have used a convolution neural network model consisting of 33 layers of convolution after which comes a fully connected layer and a softmax layer, so in total their model has 34 layers. The input consists of a sequence of numbers denoting the ECG signal and an output consists of arrhythmia classes of the different heartbeats in the input. The model makes use of batch normalization and the activation function used in the model is the Relu function. The model also makes use of a dropout algorithm to reduce overfitting. The data set Rajpurkar et al have used was collected by the team and it consists of 64,121 ECG records which were detected from 29, 163 patients., and were sampled at 200Hz. The records were annotated by a group of experts. The number of output classes in the study were 14, out of which 12 were different arrhythmia classes, one was the sinus rhythm which is the normal heartbeat and one was noise. The data was partitioned into the training set and test set and they tested their model against board certified cardiologist and it was found that their model performed better. The precision of their model turned out to be 0.8 whereas the precision of the board certified cardiologist came out to be 0.723 . This is a very relevant study because it shows that computational models are better at detecting heart arrhythmias using single lead ECG than human experts.

Yao et al[25] have proposed an attention based time incremental convolutional neural network(ATI-CNN) for the detection of arrhythmia using 12 lead ECG. They have used the ATI-CNN model to address the issues of convolutional neural networks(CNN) in detecting arrhythmia such as CNN's inability to handle varied length ECG signals. Their proposed model consists of a 13 layer CNN module , a 2 layer LSTM module and an attention module one after another and hence the model achieves both spatial and temporal fusion of information from ECG signals. The dataset that was used in this study was from the 1st China Physiological Signal challenge and it consisted of 12 lead ECG recordings ranging from 6 seconds to 60 second. There were 6877 ECG recordings in

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:29:02 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Trends in Electronics and Informatics (ICOEI).
IEEE Xplore Part Number:CFP21J32-ART; ISBN:978-1-6654-1571-2

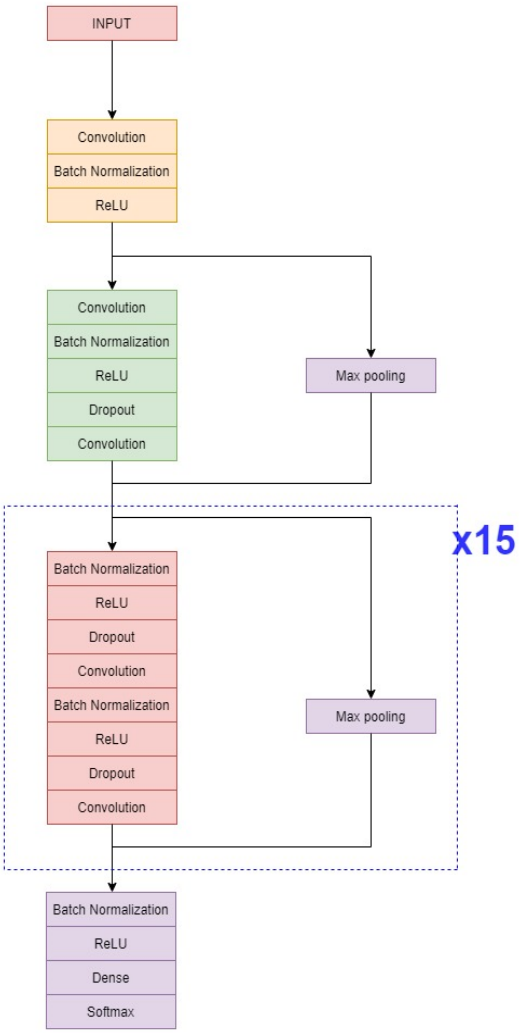


Fig. 2. The model proposed by Rajpurkar et al [18] for the detection of arrhythmia from one lead ECG.

the training set and 2954 ECG recordings in the test set. The data was sampled at 500Hz. The number of output classes that were considered in this study was 9 including, normal beat. The precision of their model turned out to be 0.826 and it's accuracy was 81.2% which was 7.7% more than a standard 16 layer CNN named VCGNet.

Yildirim et al[26] have proposed an efficient and real-time classification model for detecting 17 unique classes of arrhythmia using 1-lead ECG(MLII). they used 10-second long fragments of ECG signal for classification. They used a 16-layer deep 1D-Convolution Neural Network along with BatchNormalization and Max Pooling. Input passes through 1D convolution layer and then the outputs are normalized with the help of Batch Normalization layer, then normalized output passes through 1D max pooling layer. These operations are repeated until the 14th layer. The 14th layer of this deep

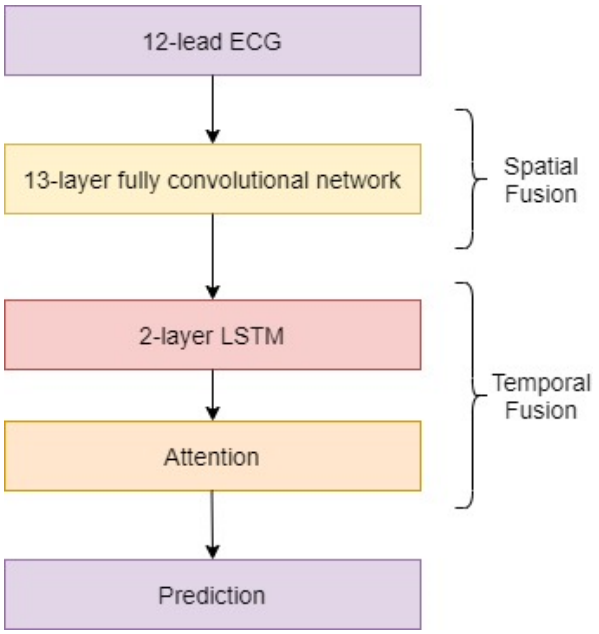


Fig. 3. The model proposed by Yao et al[25] for arrhythmia detection using 12-lead ECG

neural network is a flattened layer which transforms multi-dimensional input feature vector into 1D output data. The feature vectors are then passed to a Dense layer of size 512 units. The final layer uses Softmax as the activation function and predicts which arrhythmia class the input belongs to. The dataset used is the MIT-BIH Arrhythmia dataset. Among the 1000 ECG signal snippets, the training set includes 700 samples and the validation and testing set contains 150 samples each. After training for 50 epochs, a training accuracy of 100% and a validation accuracy of 86.67% were obtained. The proposed 16 layered deep 1D-CNN results in an accuracy of 91.33% on testing data and the classification time per single sample was 0.015 second.

Wang et al[23] have proposed a unique model for the purpose of arrhythmia detection. They have used deep multi scale fusion neural networks that can effectively classify multiple arrhythmias. The model they have used in their research consists of three key components. The first is a backbone network that is able to learn shared low level features. The second is a collection of sub-networks that can effectively learn scale specific high level signal features and different scales convolution kernel collaboratively. The third component of the model is a multi scale features fusion that is used for integrating all of the different features from the sub networks in the previous component and further uncovering correlated information from different scales using attention. Two different data sets have been used in this particular research. The first data set is the CPSC_2018 data set which consists of 6877 12 lead ECG records for training and 2954 ECG records for testing, sampled at 500 Hz and ranging from 6 to 60 seconds. This data set includes 9 different types of heartbeat

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:29:02 UTC from IEEE Xplore. Restrictions apply.

namely, normal beat and 8 different types of arrhythmia. The second data set that has been made use of in this paper is the PhysioNet/CinC_2017 data set. This data set consisted of 8, 528 records of single lead ECG data sampled at 300 Hz. The lengths of the recordings ranges between 2700-18, 300 data points. The output of these ECG signals were labeled with four classes namely, sinus rhythm, atrial fibrillation, alternative rhythm and noisy recordings. Their model achieved an F1 score and precision of 0.828 and 0.838 respectively on the CPSC_2018 data set. On the CINC_2017 data set their model achieved a F1 score of 0.841 and a precision of 0.856.

Miquel et al[1] developed a fast and fully automatic machine learning ECG arrhythmia classifier, that only requires a single ECG lead. They used ensemble learning to their advantage helping in training the classifier with less time. Their ECG arrhythmia classifier is evaluated on MIT-BIH AR data set and (MIT-BIH) AHA data set. This classifier is an ensemble of Echo State Networks, which uses raw ECG waveforms and heartbeats time interval for input. ESNs have recurrent connections which allow them to take time dependencies between two heartbeats into account.The MIT-BIH AR(arrhythmia) database consists of 48 half-hour ECG records sampled at 360 Hz. The AHA database contains 3 hours long, 154 ECG records. Each data set is divided into two sets, for training and for testing. The data is resampled at 250 Hz with a bandwidth of 0.5Hz-35Hz segmented around annotated position with a size of 240ms. The classifier used in this paper is based upon Echo State Network(ESN), which is considered as implementation of Reservoir Computing(RC). RC is a machine learning paradigm which consists of input, reservoir and output layer. In ESN the reservoir is a Recurrent Neural Network which allows context-dependent computations.The reservoir consists of sigmoid function which shows optimal results for classification. The output is obtained from the ensemble by majority voting over individual ESN output. The classifier used in this paper is an ensemble of thirty ESNs following ring structure. The proposed classifier classifies the input into two classes. For MIT-BIH AR data set the accuracy of the classifier is 98.6% and for MIT-BIH AHA data set the accuracy achieved by the classifier is 97.8%.

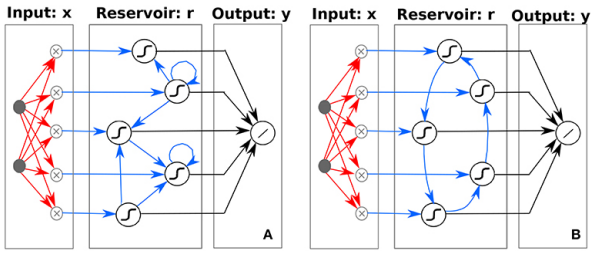


Fig. 4. The Reservoir Computing method proposed by Miquel et al[1] for arrhythmia detection using ECG signals.

Sannino et al[19] proposed a Deep Neural Network of 7 hidden layers for heartbeat classification. The data set used in this paper is MIT-BIH Arrhythmia data set. This data set is used to create a new preprocessed data set. Noises are eliminated, position of peaks are detected, signals are segmented into single beats classified as normal or abnormal. After these preprocessing steps the new data set consists of a total of 84,615 items, out of these 66,750 are normal beats(N), 2288 are abnormal beats and rest 14,828 represents unclassifiable beats(Q). These unclassifiable beats are dropped and only 2288 randomly selected normal beats out of 66,750 normal beats are considered so that the classification model does not suffer from bias. Thus the final data set consists of 4576 items. This data is used to create a Deep Neural Network which predicts two classes: normal beat and abnormal beat. After various manual configurations, the DNN with best classification performance contains seven hidden layers, with 5, 10, 30, 50, 30, 10 and 5 neurons respectively. The activation function for the hidden layer is ReLU and for the output layer it is Softmax. The proposed classifier has an accuracy of 100% on training data and an accuracy of 99.09% on test data, while the sensitivity and specificity for training data is 100% and for testing data it is 98.55% and 99.52% respectively.

Isin et al[11]have proposed a transfer learning method for detection of arrhythmia from ECG signals. A very popular convolutional neural network named AlexNet is used for feature extraction and then the extracted features is fed to a simple feed forward neural network that is trained using backpropagation. The workflow in their study undergoes various steps. The first is data acquisition in which they use a very popular dataset known as the MIT-BiH dataset. Then various signal smoothing algorithms are applied to the signal. Further Pan-Tompkins algorithm is used for QRS detection after which the signal is fed onto AlexNet for feature extraction. They test 3 different neural networks for classification and the highest accuracy that was obtained was 92.4%.

Petmezas et al[16] proposed a hybrid CNN-LSTM approach for arrhythmia detection. Atrial fibrillation is an arrhythmia that can cause failure of heart. There are four classes of ECG beats predicted by the proposed architecture. The dataset used is MIT-BIH Atrial Fibrillation database, which includes 23 long term 2-lead ECG readings sampled at 250 Hz. The data preprocessing step consists of removing unwanted noises, beat segmentation. The preprocessed data is then fed to Convolutional Neural Networks. The CNN is used to extract meaningful ECG features which are then inputted to a Long Short-Term Memory(LSTM) model, this results in memorization of temporal dynamics which help in more accurate classification of ECG beats. To deal with the imbalance data focal loss is implemented. The proposed hybrid CNN-LSTM architecture has thirteen hidden layers other than input and output layers. First, ECG signal is passed through CNN parts which have triplets of layers with each triplet including Convolutional, Batch Normalization and MaxPool layer. The CNN part is responsible for feature extraction as well as dimensionality reduction. The results from CNN part move to LSTM part. The LSTM part is responsible for memorizing time related dependencies between data. The LSTM part includes LSTM layer, flatten, fully-connected, dropout and output layers. Activation function for the convolutional and fully connected layer

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:29:02 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Trends in Electronics and Informatics (ICOEI).
IEEE Xplore Part Number:CFP21J32-ART; ISBN:978-1-6654-1571-2

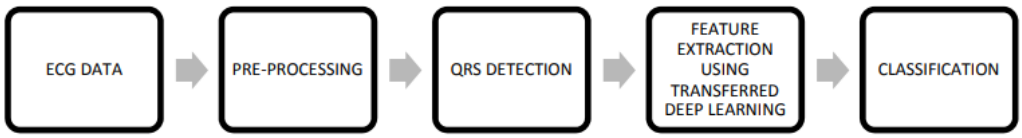


Fig. 5. The transfer learning method proposed by Isin et al[11] for arrhythmia detection using ECG signals.

is ReLU and for the output layer it is Softmax. Then a 10-fold CV is used for model evaluation. The training process ends in 100 epochs resulting in an overall sensitivity of 97.87% and an overall specificity of 99.29%.

Hong et al[10] have performed a study on arrhythmia detection using single lead ECG signals. In their study they have attempted to combine hand engineered features and features extracted from deep neural networks. The combined features are then further inputted to a classifier for classification of the detected arrhythmia. They have used various types of engineered features such as statistical features, signal features, morphological features. They have used various different types of deep neural networks like CNN, RNN, ResNet for feature extraction. The extracted and the engineered features were then trained and tested using different classifiers such as softmax, decision tree, XGBoost. Ensemble XGBoost. The best performance was given by ensemble XGBoost. The performance measure used by them in their study was the F-score and the F-score obtained by their model was 0.7605. The data set used in this study was obtained from PhysioNet/Computing in Cardiology Challenge 2017. This data set consisted of 8, 528 records odf single lead ECG data sampled at 300 Hz. The lengths of the recordings ranges between 2700-18,300 data points. The output of these ECG signals were labeled with four classes namely, sinus rhythm, atrial fibrillation, alternative rhythm and noisy recordings.

Warrick et al[24] have proposed a model which combines CNN architecture and LSTM architecture for the detection of arrhythmia using ECG signals. The model consists of 2 main components namely the representative learning component and the sequence learning component. The representative learning component consists of 1D CNN layer which is used for feature extraction, and the output of this layer is fed onto the sequence learning layer which consists of a stack of three LSTM layers. After the LSTM layer the data is fed onto a softmax layer for classification. The data set they used was from the PhysioNet Challenge 2017 which is a fairly common and standard data set. The number of output classes used in this study was 4. The primary performance measure they used to evaluate their model. They conducted a series of experiments and the highest F1 score that was obtained was 0.83.

III. Further Discussion

Moghadas et al[15] have proposed an IoT based framework for the automatic detection of arrhythmia using remote monitoring of a patient's health. Their study makes use of fog computing The patient monitoring system is based on fog computing. Fog computing is an IoT architecture in which the edge devices are responsible for a significant amount of computation and storage. In their Iot based systems three main hardware components are used. The first is the Arduino Uno which is a microcontroller, The second is the Raspberry Pi 3 B+. which is a computer that is responsible for computation of the pre trained machine learning algorithm. The third is the AD8232 board which is responsible for detecting ECG signals and heartbeat of the patient. The machine learning algorithm they used for the classification was the KNN algorithm. The data set they used to train the machine learning algorithm was the UCI data set for arrhythmia. In the particular dataset 16 different output classes were present and the model used in the study gave an accuracy of 94.44%. In their particular IoT model, the sensor sensed the ECG data and the computing was done by the processors. The data was then sent to the cloud for storage. In case of an emergency or heart attack the patient's family or physician was notified through an alert system.

Jinyuan et al[9] have proposed a method for detection of arrhythmia from IoT-based ECGs. The proposed method consists of two parts, first is data cleaning part and the other one is heartbeat classification part.The data set used is MIT-BIH database. The proposed model predicts five classes. Input data first passes through a data cleaning module in which ECG recordings are re-sampled to 360 Hz, then these signals are filtered of any unwanted noise and then 230 heartbeat samples are obtained using heartbeat segmentation. This paper proposes two solutions: Dynamic Heartbeat Classification with Adjusted Features(DHCAF) and Multi-channel Heartbeat Convolution Neural Network(MCHCNN). DHCAF method is based on feature selection. DHCAF contains 4 stages: first being Feature extraction , the second step is building a classifier training pool with various classifiers(eg.,SVM, decision tree, etc), in the third step most competent classifiers are selected for making predictions and then at the last step the aggregated results from previous steps are refined. MCHCNN takes in two inputs, one being an ECG beat and other being RR-intervals. MCHCNN performs parallel convolutions using 3 channels on the input beats to retrieve frequency and time related information of ECG heartbeats. After every convolution operation there is a batch normalization step and the activation function used is ReLU. The features learned from three parallel implemented convolution layers are integrated and then passed to the pooling layer. Then a Rhythm Integration layer is

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:29:02 UTC from IEEE Xplore. Restrictions apply.

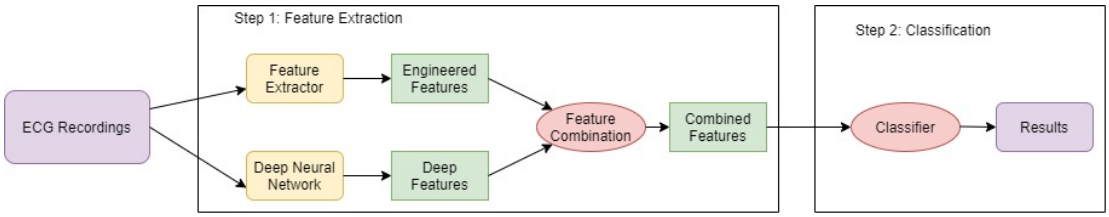


Fig. 6. The model proposed by Hong et al[10] for arrhythmia detection using single lead ECG

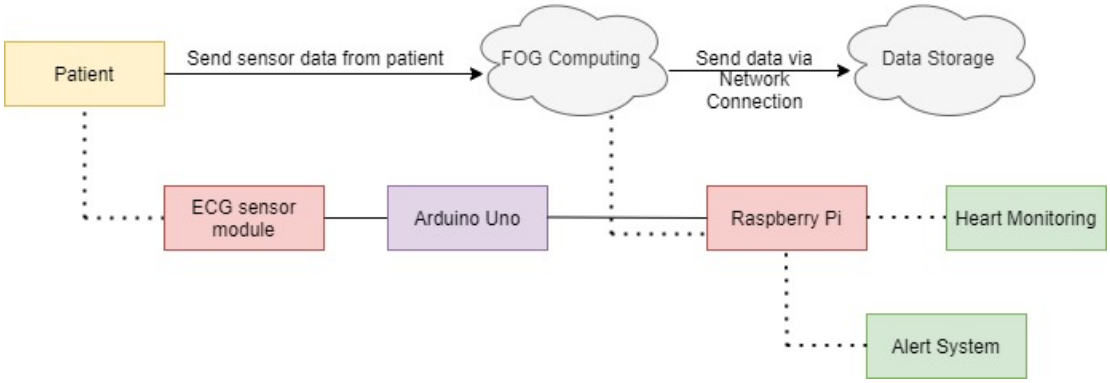


Fig. 7. Iot architecture used in the study done my Moghadas et al[15]

used to join input heart rhythms(RR-intervals) and the learned features. Next, a dense layer followed by a Softmax layer predicts the class to which the beats belong. The proposed DHCAF method has an accuracy of 91.4% and the proposed MCHCNN method has an accuracy of 93%.

Dimitra et al[2] proposed an algorithm for arrhythmia detection by ECG signal analysis and implemented it on an IoT-based platform. The dataset used is MIT-BIH Arrhythmia database. The IoT platform used is Intel's Galileo board, with Linux operating system, 400MHz processor and a RAM of 256MB. Application implemented on Galileo board takes a digitized ECG signal as input and these signals are read at 350 samples/second with a batch size of 3000 samples. First a bandwidth filter is written in C which is used to filter the signals at 1-50 Hz. Next, WFDB function wqrs() is implemented in C which gives back the starting point of QRS complex. Next, heartbeat segmentation is performed on the ECG signal. Next, Discrete Wavelet Transform method is used for feature extraction. After getting feature vectors a binary classifier is implemented which predicts beats as Normal and Abnormal. They used 104581 heartbeats from 45 selected ECG readings for classification. The LIBSVM library of Matlab interface also includes C implementation of svmpredict() which converts SVM model into format used in C implementation. The proposed signal analysis and classification architecture detected 10 heartbeats for every 3000 sample inputs, and in best case total time of execution is less than 9 seconds which shows that it can be used in real-time. The classification accuracy achieved by the proposed algorithm is above 97% while the best case accuracy is 98.9% for 16 feature vectors and 2493 support vectors.

Lakshmi et al[4] proposed an IoT-based cardiac arrhythmia detection system using ECG signals. The data set used for training and validation of the classifier is obtained from Physionet ATM database. ECG signals are collected from various data sets: eighteen ECG readings from MIT-BIH Normal Sinus Rhythm Database(NSRDB), twenty two ECG recordings from MIT-BIH Malignant Ventricular Ectopy database, thirty five recordings from CU Ventricular Tachycardia Database and seventy eight ECG signals from MIT-BIH Supraventricular Arrhythmia Database. To acquire ECG signals they used an Arduino controller which receives data from ECG sensor AD8232. The AD8232 sensor reads electrical signals from the heart and then the Arduino controller gets that data and transmits it to the processing system. To detect QRS complexes, the Pan Tompkins algorithm is used. Feature extraction of raw ECG signal gives statistical features and dynamic features. The model for classification is implemented in MATLAB and the classification app consists of various types of SVM classifiers. The proposed algorithm with only statistical features set has an effectiveness of 88.9% and the proposed algorithm with dynamic features has an effectiveness of 90.8% while a combination of these two features has an effectiveness of 92.2%.

IV. Conclusion and Future Works

In this paper we have reviewed various recent and impactful papers on arrhythmia detection using various artificial intelligence and IoT based techniques. As stated above heart diseases are one of the leading causes of human deaths and

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:29:02 UTC from IEEE Xplore. Restrictions apply.

TABLE I
SUMMARY OF WORKS DONE IN THE FIELD OF ARRHYTHMIA DETECTION

Author	Date of Publication	Technology Used	Dataset Used	Performance Measure
Rajpurkar et al[18]	2017	Convolutional Neural Network	Novel data set collected by the authors	Precision: 0.8, F1 score: 0.776
Yao et al[25]	2020	Attention based time incremental Convolution Neural Network	1st China physiological Signal challenge data set	Precision: 0.826, F1 score: 0.812
Hong et al[10]	2019	DNN + machine learning	PhysioNet/ computing in Cardiology Challenge data set	F1 score: 0.7605
Wang et al[23]	2020	Deep multi scale fusion Neural Network	physioNet/CiNC 2017 and CPSC_2018 data set	F1 score(CPSC): 0.828, F1 score(CINC): 0.841
Isin et al [11]	2017	Transfer Learning using AlexNet + Neural Network	MIT-BiH arrhythmia dataset	Accuracy: 92.4%
Warrick et al[24]	2017	Convolutional Neural Network + Long short Term Memory	PhysioNet Challenge 2017	F1 score: 0.83
Moghadas et al[15]	2019	Internet of Things + K-nearest neighbours	UCI dataset for arrhythmia	Accuracy: 94.44%
Yildirim et al[26]	2018	1D-Convolution Neural Network	MIT-BIH Arrhythmia data set	Accuracy: 91.33%
Miquel et al[1]	2019	Echo State Network based classifier	MIT-BIH AR dataset and MIT-BIH AHA dataset	Accuracy(AR): 98.6%, Accuracy(AHA): 97.8%
Sannino et al[19]	2018	Deep Neural Network	MIT-BIH Arrhythmia database	Sensitivity:98.55%, Specificity:99.52%
Petmezas et al [16]	2021	Hybrid Convolution Neural Network and Long Short Term Memory	MIT-BIH Arrhythmia database	Sensitivity:97.87%, Specificity:99.29%
Dimitra et al [2]	2016	Internet of Things + Support Vector Machines	MIT-BIH Arrhythmia dataset	Accuracy : 97%
Lakshmi et al [4]	2019	Internet of Things + Support Vector Machines	MIT-BIH arrhythmia data set	Accuracy: 92.2%
Jinyuan et al [9]	2020	Multi-channel Heartbeat Convolution Neural Network(MCHCNN) and Dynamic Heartbeat Classification with Adjusted Features(DHCAF)	MIT-BIH Arrhythmia dataset	Accuracy(DHCAF): 91.4%, Accuracy (MCHCNN): 93%

it is quintessential for computational systems to be able to accurately diagnose various heart diseases. With this paper we have attempted to give a glimpse of the major advancements in the field of arrhythmia detection in recent years. We have reviewed various papers on the subject matter and succinctly written about them in a technical manner. This paper will be useful for any researcher or academician venturing into the world of arrhythmia detection using recent computational technologies. The recent landscape of the subject matter mostly consists of deep neural networks such as deep convolutional neural networks, RNN or a hybrid deep neural network comprising various different neural networks. These networks have successfully been shown to give very accuracy in the field of arrhythmia detection. As the amount of ECG data grows and artificial intelligent technologies advance, it is sure that computational systems that are capable of 100% accuracy will be created in the near future. Perhaps one important application future researchers can work on is the early detection of arrhythmia, this is to be able to detect different types of arrhythmia even before it occurs in patients. The ultimate goal in the field of automatic arrhythmia diagnosis using ECG signals would be to create an end to end system that is capable of monitoring the heart beat of a patient all the time and give 100% accurate diagnosis of any little change that is happening in the heart of the patient. In the future artificially intelligent systems will probably make the role of a cardiologist obsolete as the system itself will correctly diagnose and treat a patient.

REFERENCES

[1] Miquel Alfaras, Miguel C Soriano, and Silvia Ortın. “A fast machine learning model for ECG-based heartbeat classification and arrhythmia detection”. In: *Frontiers in Physics* 7 (2019), p. 103.

[2] Dimitra Azariadi et al. “ECG signal analysis and arrhythmia detection on IoT wearable medical devices”. In: *2016 5th International conference on modern circuits and systems technologies (MOCAST)*. IEEE. 2016, pp. 1–4.

[3] *Cardiovascular diseases*. URL: https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:29:02 UTC from IEEE Xplore. Restrictions apply.

[4] R Lakshmi Devi and V Kalaivani. "Machine learning and IoT-based cardiac arrhythmia diagnosis using statistical and dynamic features of ECG". In: *The Journal of Supercomputing* (2019), pp. 1–12.

[5] Elgar Fleisch, Markus Weinberger, and Felix Wortmann. "Business models and the internet of things". In: *Interoperability and open-source solutions for the internet of things*. Springer, 2015, pp. 6–10.

[6] Awishkar Ghimire et al. "Accelerating business growth with big data and artificial intelligence". In: *Proc. 4th Int. Conf. IoT Soc. Mobile, Anal. Cloud, ISMAC 2020*. November. 2020, pp. 441–448. ISBN: 9781728154640. DOI: 10.1109/I-SMAC49090.2020.9243318.

[7] Awishkar Ghimire et al. "AI and IoT Solutions for Tackling COVID-19 Pandemic". In: *Proc. 4th Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2020*. November. 2020, pp. 1083–1092. ISBN: 9781728163871. DOI: 10.1109/ICECA49313.2020.9297454.

[8] Awishkar Ghimire et al. "Machine Learning Approach Based on Hybrid Features for Detection of Phishing URLs". In: *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*. IEEE. 2021, pp. 954–959.

[9] Jinyuan He et al. "A framework for cardiac arrhythmia detection from IoT-based ECGs". In: *World Wide Web* 23.5 (2020), pp. 2835–2850.

[10] Shenda Hong et al. "Combining deep neural networks and engineered features for cardiac arrhythmia detection from ECG recordings". In: *Physiological measurement* 40.5 (2019), p. 054009.

[11] Ali Isin and Selen Ozdalili. "ScienceDirect ScienceDirect Cardiac arrhythmia detection using deep learning". In: *Procedia Comput. Sci.* 120 (2018), pp. 268–275. ISSN: 1877-0509. DOI: 10.1016/j.procs.2017.11.238. URL: https://doi.org/10.1016/j.procs.2017.11.238.

[12] Avinash Kumar Jha et al. "A Review of AI for Urban Planning: Towards Building Sustainable Smart Cities". In: *Proc. 6th Int. Conf. Inven. Comput. Technol. ICICT 2021*. January. 2021, pp. 937–944. ISBN: 9781728185019. DOI: 10.1109/ICICT50816.2021.9358548.

[13] Arunish Kumar et al. "Crack Detection of structures using Deep learning frameworks ." In: 13 (2020), pp. 1–24. DOI: 10.1109/ICISS49785.2020.9315949.

[14] Shanthi Mendis et al. *Global atlas on cardiovascular disease prevention and control*. World Health Organization, 2011.

[15] Ehsan Moghadas, Javad Rezazadeh, and Reza Farahbakhsh. "An IoT Patient Monitoring based on Fog computing and Data Mining : Cardiac Arrhythmia Usecase". In: *Internet of Things* (2020), p. 100251. ISSN: 2542-6605. DOI: 10.1016/j.iot.2020.100251. URL: https://doi.org/10.1016/j.iot.2020.100251.

[16] Georgios Petmezas et al. "Automated Atrial Fibrillation Detection using a Hybrid CNN-LSTM Network on Imbalanced ECG Datasets". In: *Biomedical Signal Processing and Control* 63 (2021), p. 102194.

[17] Jennifer S Raj. "A Novel Information Processing in IoT Based Real Time Health Care Monitoring System". In: *Journal of Electronics* 2.03 (2020), pp. 188–196.

[18] Pranav Rajpurkar et al. *Cardiologist-level arrhythmia detection with convolutional neural networks*. 2017. arXiv: 1707.01836.

[19] Giovanna Sannino and Giuseppe De Pietro. "A deep learning approach for ECG-based heartbeat classification for arrhythmia detection". In: *Future Generation Computer Systems* 86 (2018), pp. 446–455.

[20] S Smys. "A Survey on Internet of Things (IoT) based Smart Systems". In: *Journal of ISMAC* 2.04 (2020), pp. 181–189.

[21] Surendrabikram Thapa et al. "Data-Driven Approach based on Feature Selection Technique for Early Diagnosis of Alzheimer's Disease". In: *Proc. Int. Jt. Conf. Neural Networks*. 2020. ISBN: 9781728169262. DOI: 10.1109/IJCNN48605.2020.9207359.

[22] Surendrabikram Thapa et al. "Feature Selection Based Twin-Support Vector Machine for the diagnosis of Parkinson's Disease". In: *2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC)*. 2020.

[23] Ruxin Wang, Jianping Fan, and Ye Li. "Deep Multi-Scale Fusion Neural Network for Multi-Class Arrhythmia Detection". In: *IEEE J. Biomed. Heal. Informatics* 24.9 (2020), pp. 2461–2472. ISSN: 21682208. DOI: 10.1109/JBHI.2020.2981526.

[24] Philip Warrick and Masun Nabhan Homsi. "Cardiac arrhythmia detection from ECG combining convolutional and long short-term memory networks". In: *Comput. Cardiol. (2010)*. Vol. 44. 2017, pp. 1–4. DOI: 10.22489/CinC.2017.161-460. arXiv: 1801.10033.

[25] Qihang Yao et al. "Multi-class Arrhythmia detection from 12-lead varied-length ECG using Attention-based Time-Incremental Convolutional Neural Network". In: *Inf. Fusion* 53.February 2019 (2020), pp. 174–182. ISSN: 15662535. DOI: 10.1016/j.inffus.2019.06.024.

[26] Özal Yıldırım et al. "Arrhythmia detection using deep convolutional neural network with long duration ECG signals". In: *Computers in biology and medicine* 102 (2018), pp. 411–420.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:29:02 UTC from IEEE Xplore. Restrictions apply.

2021 2nd International Conference for Emerging Technology (INCET)
Belgaum, India. May 21-23, 2021

Medical Image Fusion using Lifting based Transform Techniques

Anisha Gupta
Department of Electronics & Communication Engineering
Delhi Technological University
Shahbad Daulatpur, Bawana, Delhi –42
anishagupta772@gmail.com

Priyesh Chaurasia
Department of Electronics & Communication Engineering
Delhi Technological University
Shahbad Daulatpur, Bawana, Delhi –42

Sudipta Majumdar
Department of Electronics & Communication Engineering
Delhi Technological University
Shahbad Daulatpur, Bawana, Delhi –42

2021 2nd International Conference for Emerging Technology (INCET) | 978-1-7281-7029-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/INCET51464.2021.9456179

***Abstract*- Medical image fusion is the means to obtain an enhanced image with better quality and reduced redundancy by fusion of two or more images that may or may not belong to the same modalities. The imaging modalities such as Magnetic Imaging Resonance (MRI), Computed Tomography (CT), Positron Emission Tomography (PET), Single Positron Emission Tomography (SPECT) etc have found their application in diagnosis in the medical domain. The paper presents the use of these imaging modalities pairs (MRI-T1/T2; MRI-PET/ SPECT) to perform medical fusion in transform domain. The paper presents two approaches. 1) It uses the lifting step-based wavelet transform (LWT), a better algorithm that takes significantly less computational time, does distortion-free reconstruction for the Average-Max, Edge-Energy fusion rules. 2) It proposes the new fusion rule to obtain an improved fused image. The fusion rule is applied using the DWT and LWT, and comparisons are made accordingly on the image pairs. Entropy, Peak Signal to Noise Ratio (PSNR), and average gradient are used as performance metrics to assess the image quality.**

I. INTRODUCTION

Image fusion is a technique to derive desirable characteristics from the images used in amalgamation to form an image that contains accurate and more eminent data content. There are several Imaging Modalities like MRI, CT, PET, SPECT and X-rays used in medical diagnosis to provide a diagnosis but in a restricted domain. The X-rays helps in examinations of fractures or related bone-related dislocations and degeneration [1]; CT scan is an advanced version for X-ray that creates a detailed bone image, tendons, and soft-tissue; MRI gives a better diagnosis and perspective on the ailments concerning soft-tissues, joints; and PET and SPECT provide better pathological, metabolic functionality, unlike the other techniques.[1],[2].

Therefore, most of the time CT/MRI images are fused with PET/SPECT to get both the anatomical and metabolic characteristics that increase the applicability and the use of images in medical examination [1].

The Image fusion methodology includes either the transform domain or the spatial domain. The spatial domain techniques work with the pixel of images, manipulating them using logarithmic transform, histogram equalization, or simple averaging techniques [3]. The transform domain works in the frequency domain and is associated with the frequency content of the image [3]. The transform domain techniques convert images into multi-scale or sub-bands. Then, some operator is applied to them to enhance the quality before using the inverse operation.

Various methods over the years have been proposed for Medical image fusion. Bing Huang, Feng Yang, Mengxiao Yin, Xiaoying Mo, Cheng Zhong [4] presents the review for medical fusion in popular domains like non sampled shearlet transform (NSST), DWT, Nonsubsampled Contourlet Transform and (NSCT), NSST domain and different methods including pulse-coupled neural network (PCNN) based fusion, Frei-Chen Operator Medical Image Fusion and deep- learning-based methods. K. KoteswaraRao and K. Veera Swamy [5] used both the transform domain (NSCT, DWT) and the ESOP values in the spatial domain to fuse the images. B. Ashwanth and K. V. Swamy [2] proposed a new fusion algorithm to carry out Medical image fusion in DWT and SWT transform domains highlighting the comparisons between the usage of un-decimated and decimated wavelet transformation. Tan, W., Tiwari, P., Pandey, H.M. *et al.* [6] proposes the new algorithm based on PCNN and energy fusion in NSST domain tested on Alzheimer's disease and related datasets. M. P., T. Singh, R. Nayar and S. Kumar [7] uses CNN architecture for medical image fusion and different fusion methods for denoising the medical images. Jingming Xia, Yi Lu, Ling Tan [8] proposes the new fusion method in the NSST domain that employs the fusion of high-level frequency components using parameter adaptive PCNN and low-level frequency components using convolutional sparse representation (CSR). P.V.Ramaraju, D.Sudharani [9] proposes a new hybrid fusion methods for CT and MRI images in DWT and discrete stationary wavelet transform domain. P. Mathiyalagan [10] proposes Curvlet Transform for medical image fusion that intents to give sharpened image resolution of a fused image. K. Vanitha, D. Satyanarayana and M. N. G. Prasad [11] proposes hybrid l1-l0 decomposition based two scale fusion method and mean, standard deviation methods are employed to evaluate results.

In our paper, to perform the medical image fusion, we use LWT on pre-existing fusion rules like Average-Max and Edge-Fusion to draw comparisons. We also propose a new fusion algorithm that uses both the DWT, a first-generation-based transformation, and the LWT, a second-generation-based wavelet transformation.

DWT has its advantages like the preservation of the spatial features [2],[12], less distortion, and more efficiency in 1-D singularity [2]. The LWT preserves these advantages of DWT and improves the limitations by reducing the error, giving perfect reconstruction, decreasing the computational time and increasing the vanishing moments [13], [14].

II. WAVELET TRANSFORMS

A. Discrete Wavelet Transform (DWT)

The DWT is defined by [12]:

$$W\varphi(j_0,k) = \frac{1}{\sqrt{M}} \sum_x f(x)\varphi_{j_0,k}(x) \quad (1)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:31:41 UTC from IEEE Xplore. Restrictions apply.

$$W\psi(j,k) = \frac{1}{\sqrt{M}} \sum_k f(x)\psi_{j,k}(x) \quad (2)$$

The coefficient in the first equation is called the approximation coefficient and the coefficient in the second equation is called the detail coefficient.

The 2-D DWT is performing the 1-D wavelet transform two times on a 2-D image. Therefore, at an instance, only one dimension of an image is considered. When DWT is applied on a row, coarse row and detail row sub-bands are created. This process is repeated with the column dimension as well [15].

The sub-band created can either be due to low pass filter or high pass filter. If we produce a sub-band using a low pass filter on a row and a high pass filter on a column, it is denoted as LH. The decomposition of an image gives Low-Low (LL), Low-High (LH), High-Low (HL), High-High (HH) sub-band. Each sub-band gives different kind of information. The bands LH gives the horizontal features, HL gives the vertical features, HH gives the information about the diagonal features and the LL gives the approximation details.

TABLE I. SUB-BAND AND THE INFORMATION IT STORES

$Image_{LL\text{-}BAND}$ Approximation	$Image_{LH\text{-}BAND}$ Horizontal features
$Image_{HL\text{-}BAND}$ Vertical features	$Image_{HH\text{-}BAND}$ Diagonal features

B. Lifting Wavelet Transform (LWT)

The lifting wavelet transform is the second generation wavelet transform, an improvement over the first-generation transform that can perform wavelet mapping from Integer to Integer [13], provide perfect reconstruction, and can be applied on non-separable lattices.

Discrete Wavelet Transform incorporates the calculation of successive averages and differences of a signal. Contrary to this, Lifting Wavelet Transform applies these operations on a part of a signal. This is achieved by three lifting steps Split, Predict and Update [14].

Split - Generally, the signal is split into two disjoint subsets, mainly an odd subset and an even subset.

Predict - Usually, an even subset is used to predict the odd component (subset) via the Prediction operator P. The correlation between the subsets serves as the basis of this prediction. The difference between the actual odd-subset and the subset obtained from the prediction of the even component is computed. This is known as Residual Error(r) or the detail coefficient.[13]

Update - Then Update operator is employed on the r to give coarse approximation data. The data obtained when added to even-subset generates the information equivalent to the information obtained from the signal at low-resolution.

After these three steps, the first level signal upgrade is completed. This is equivalent to the single layer of wavelet-decomposition.[13]

These lifting steps reduce the arithmetic operations by the factor of 2 and significantly decreases the computational time.

III. METHODOLOGY

1. *Average-Max* – In this algorithm, after the decomposition of an image, the average operation is applied on the LL band and maximum operation on the other bands before the inverse operation is applied to create a fused image.[16]
2. *Max-Max* - In this algorithm, a max operation is applied on all the frequency bands, the low and high-frequency bands to construct a fused image.
3. *Edge-Energy* – This algorithm calculates the edge information using the Sobel operator in the LL band, that is low-level sub-band and using the energy fusion rule in the remaining LH, HH, and HL sub-bands to create a fused image [2]. The energy fusion rule for the band is:

$$E(j,k) = \sum_{j-1}^{j+1} \sum_{k-1}^{k+1} (LH/HL/HH(j,k)) \quad (3)$$

IV. PROPOSED METHOD

We use different image pairs to perform the image-fusion. In this paper, we have used the MRI_T1-MRI_T2, MRI-PET, and MRI-SPECT image modalities to evaluate the accuracy of both the approaches.

1) LWT using the Average-Max and Edge-Energy fusion rule.

2) LWT and DWT using the proposed fusion rule.

We preprocess the image by converting it into grayscale and making the size of both the images equal.

We have done the decomposition and reconstruction process using 'sym5' wavelet.

A. LWT on Edge-Energy and Avergae-Max

This paper uses the two-dimensional lifting decomposition to 'sym5' wavelet to construct the fused image. We use two fusion rules, Average-Max and Edge-Energy Fusion rules, and then we make comparisons with DWT using these same rules [3].

B. Proposed Hybrid Approach using DWT and LWT

Step 1: We first perform the Edge-Edge fusion on Input Image A and Input Image B to get the intermediary fused image. The edge-edge fusion rule calculates the edge information on all the sub-bands. In this particular edge-edge, we calculate edge information using the Sobel Operator.

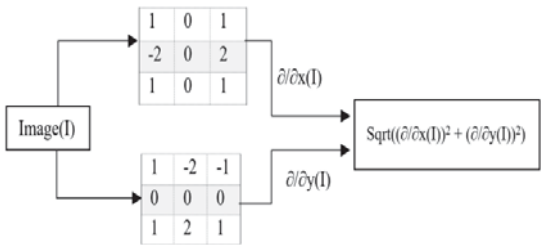


Fig. 1. Sobel Edge Detection

Step 2: We fuse the fused image obtained in step1 with the input image A to create a second fused image using the Average-Maximum fusion rule.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:31:41 UTC from IEEE Xplore. Restrictions apply.

Step 3: We fuse the fused image obtained in step1 with the input image B to create a third fused image using the Average-Maximum fusion rule.

Step 4: We create a final image using the images obtained in Step-2 and Step-3 using the Maximum-Maximum rule.

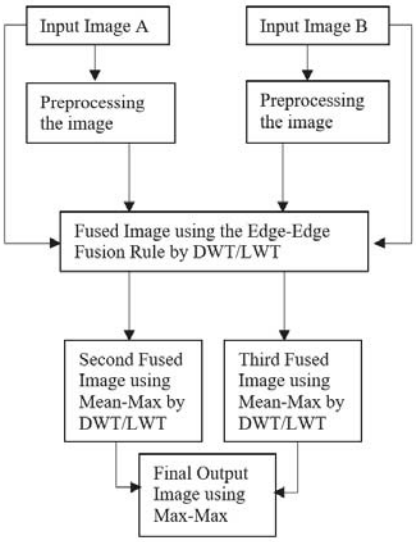


Fig. 2. Proposed Approach

V. RESULTS

A. Performance Metrics

1) Entropy – It is used to determine the information content in an image. Therefore, the greater entropy is desirable.

2) PSNR - It is defined as the ratio between the maximum signal power of the reference signal and the corrupting noise. It is expressed in the decibels (dB). If A be the reference signal/image and B be the fused/output image. The PSNR is given by

$10\log_{10}(N^2/\text{mean-square-error(A,B)})$.

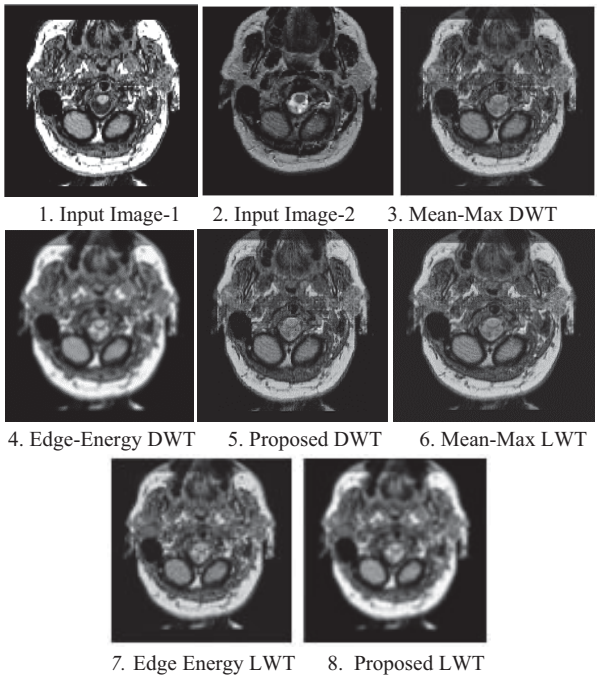
If error is less, PSNR value is high.

3) Average Gradient - We characterize the clarity and the quality of the image using the average gradient. Therefore, the greater the gradient, the better the quality of the fused image.[17]

$$g = \frac{1}{(M-1)(N-1)} \sum_{i=1}^{(M-1)(N-1)} \sqrt{\frac{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}{2}}$$

B. OUTPUT – Code and Simulations are done through MATLAB.

SET-1 – MRI_T1-MRI_T2



1. Input Image-1 2. Input Image-2 3. Mean-Max DWT

4. Edge-Energy DWT 5. Proposed DWT 6. Mean-Max LWT

7. Edge Energy LWT 8. Proposed LWT

Fig. 3. Output of Set-1
1)1st Input Image- MRI – T1, 2) 2nd Input Image – MRI – T2 3)Output with Mean-Max DWT, 4) Output Image with Edge-Energy DWT, 5) Proposed approach – DWT, 6) Output image with Mean-Max LWT, 7) Output with Edge-Energy LWT, 8) Output image with proposed approach - LWT

TABLE : II COMPARISON OF PREVIOUS APPROACHES WITH OUR NEW APPROACH USING DWT AND LWT.

		DWT			LWT		
		Mean-Max	Edge-Energy	Proposed	Mean-Max	Edge-Energy (Approach-1)	Proposed (Approach-2)
Entropy	Set-1	1.6558	1.7925	1.7363	1.7233	1.8771	1.9448
	Set-2	2.0153	2.0987	2.1165	2.0318	2.2205	2.1663
	Set-3	1.4686	1.4325	1.5119	1.5308	1.5204	1.6054
	Set-4	1.5405	1.6679	1.6681	1.5770	1.7733	1.8001
PSNR	Set-1	18.1520 18.2035	11.6834 12.9241	13.7846 13.7846	18.2036 18.1515	11.4756 11.4602	11.9433 18.7769
	Set-2	20.4311 20.4311	11.5390 14.6748	18.0022 19.0945	20.4309 20.4313	11.5396 14.6704	18.0018 19.0951
	Set-3	15.3647 15.3827	11.4913 8.5450	15.5894 13.9051	15.3831 15.3643	11.3644 8.4624	15.6091 13.8919
	Set-4	17.3891 17.4018	6.9787 9.7437	13.0504 5.2784	17.4015 17.3894	6.9912 9.7726	11.3609 15.9802
Average Gradient	Set-1	56.9549	39.9167	42.6392	57.3730	43.5732	43.1929
	Set-2	27.1991	21.4137	23.2190	27.2273	21.7885	26.2621
	Set-3	37.6370	25.6572	31.0977	37.9434	28.7622	31.3981
	Set-4	46.4641	37.0165	36.8755	46.6557	36.3363	37.1184

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:31:41 UTC from IEEE Xplore. Restrictions apply.

SET-2 MRI-SPECT

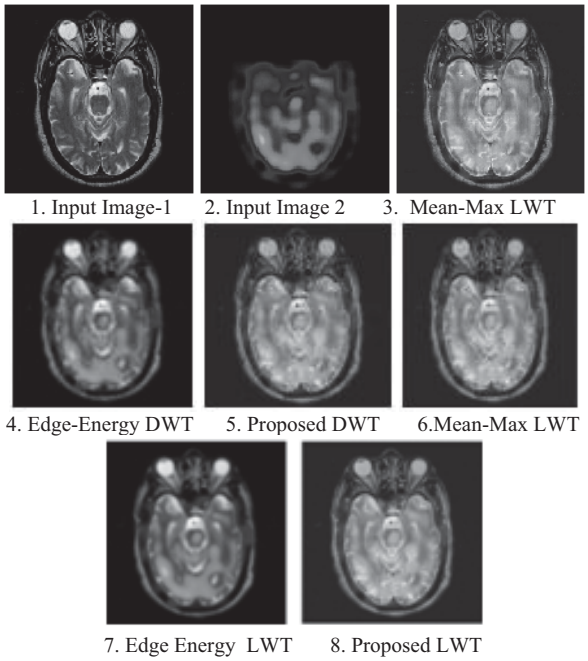


Fig. 4. Output of Set-2
1)1st Input Image- MRI , 2) 2nd Input Image – SPECT 3)Output with Mean-Max DWT, 4) Output Image with Edge-Energy DWT, 5) Proposed approach – DWT, 6) Output image with Mean-Max LWT, 7) Output with Edge-Energy LWT, 8) Output image with proposed approach - LWT

SET - 3 MRI-PET

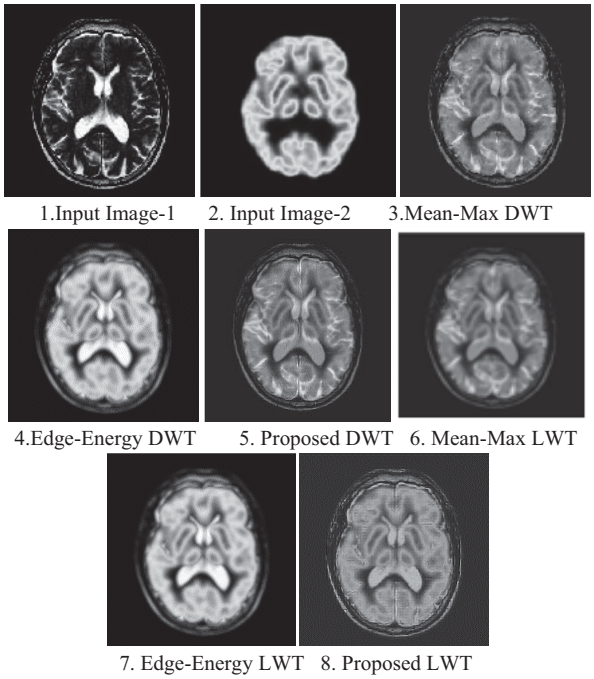


Fig. 5. Output of Set-3
1)1st Input Image- MRI , 2) 2nd Input Image – PET 3)Output with Mean-Max DWT, 4) Output Image with Edge-Energy DWT, 5) Proposed approach – DWT, 6) Output image with Mean-Max LWT, 7) Output with Edge-Energy LWT, 8) Output image with proposed approach - LWT

SET-4 MRI_T1-MRI_T2

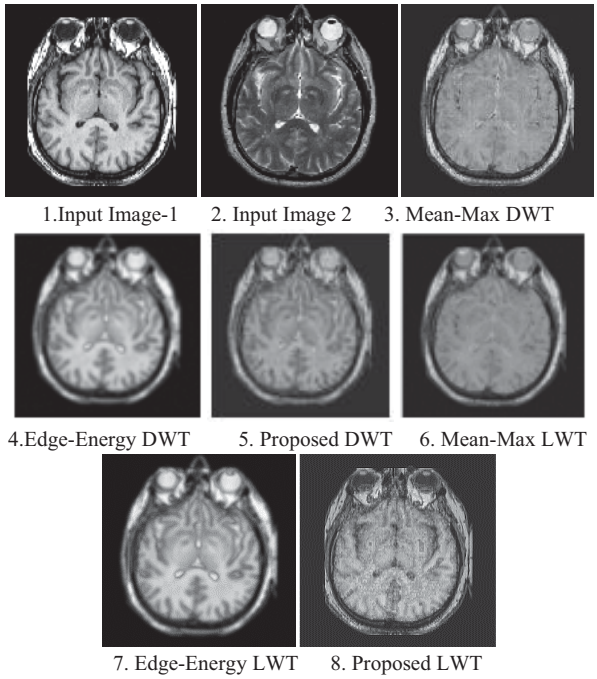


Fig. 6. Output of Set-4
1)1st Input Image- MRI-T1, 2) 2nd Input Image – MRI-T2 3)Output with Mean-Max DWT, 4) Output Image with Edge-Energy DWT, 5) Proposed approach – DWT, 6) Output image with Mean-Max LWT, 7) Output with Edge-Energy LWT, 8) Output image with proposed approach – LWT

VI. CONCLUSION

In terms of entropy, LWT performs better on all the image pairs than the DWT counterparts. Our proposed approach for LWT-Edge Energy works better than existing DWT-Edge Energy. Our proposed fusion algorithm (Approach-2) for LWT works best among all the scenarios.

Similarly, the average gradient calculated using the LWT performs better than the DWT counterparts in all the image-pairs.

The table shows the two values of PSNR, PSNR of fused image with Input Image-1 and PSNR of fused image with Input Image-2.

REFERENCES

[1] https://www.envrad.com/difference-between-x-ray-ct-scan-and-mri/

[2] B. Ashwanth and K. V. Swamy, "Medical Image Fusion using Transform Techniques," 2020 5th International Conference on Devices, Circuits and Systems (ICDCS), Coimbatore, India, 2020, pp. 303-306, doi: 10.1109/ICDCS48716.2020.243604.

[3] Snehal O.Mundhada, V. K.Shandilya, " Spatial and Transformation Domain Techniques for Image Enhancement ", International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 1, Issue 2, November 2012, ISSN: 2319-5967 ISO 9001:2008 Certified.

[4] Bing Huang, Feng Yang, Mengxiao Yin, Xiaoying Mo, Cheng Zhong, "A Review of Multimodal Medical Image Fusion Techniques", Computational and Mathematical Methods in Medicine, vol. 2020, Article ID 8279342, 16 pages, 2020. https://doi.org/10.1155/2020/8279342

[5] K. KoteswaraRao and K. Veera Swamy, " Multimodal Medical Image Fusion using NSCT and DWT Fusion Frame Work", Volume-9 Issue2, December 2019,pp3644-3645.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:31:41 UTC from IEEE Xplore. Restrictions apply.

[6] Tan, W., Tiwari, P., Pandey, H.M. et al. Multimodal medical image fusion algorithm in the era of big data. Neural Comput & Applic (2020). https://doi.org/10.1007/s00521-020-05173-2

[7] M. P., T. Singh, R. Nayar and S. Kumar, "Multi Modal Medical Image Fusion using Convolution Neural Network," 2019 Third International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 2019, pp. 351-357, doi: 10.1109/ICISC44355.2019.9036373

[8] Jingming Xia, Yi Lu, Ling Tan, "Research of Multimodal Medical Image Fusion Based on Parameter-Adaptive Pulse-Coupled Neural Network and Convolutional Sparse Representation", Computational and Mathematical Methods in Medicine, vol. 2020, Article ID 3290136, 13 pages, 2020. https://doi.org/10.1155/2020/3290136

[9] P.V.Ramaraju , D.Sudharani, "Medical Image Fusion using Discrete Wavelet Transform and Proposed Hybrid technique", International Journal for Research in Applied Science & Engineering Technology (IJRASET), ISSN: 2321-9653, Volume 6 Issue V, doi: http://doi.org/10.22214/ijraset.2018.6139.

[10] P. Mathiyalagan, "Multi-Modal Medical Image Fusion Using Curvelet Algorithm," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India, 2018, pp. 2453-2458, doi: 10.1109/ICACCI.2018.8554759.

[11] K. Vanitha, D. Satyanarayana and M. N. G. Prasad, "Multimodal Medical Image Fusion Based on Hybrid L1- L0 Layer Decomposition Method," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 2019, pp. 1-5, doi: 10.1109/ICCCNT45670.2019.8944896.

[12] (2008) Discrete Wavelet Transform (DWT). In: Furht B. (eds) Encyclopedia of Multimedia. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-78414-4_305.

[13] Yu Guang and Zhao Xue Jun, "The study of the second generation of wavelet transform," World Automation Congress 2012, Puerto Vallarta, Mexico, 2012, pp. 1-4.

[14] La Cour–Harbo, A., & Jensen, A. (2009). Wavelets and the Lifting Scheme. Encyclopedia of Complexity and Systems Science, 10007–10031. doi:10.1007/978-0-387-30440-3_588

[15] Mistry, Dr. (2013). DISCRETE WAVELET TRANSFORM USING MATLAB. International journal of Computer Engineering & Technology (IJCET). 4. 252-259.

[16] Meda Balachandra Mule and Padmavathi N.B,"Basic Medical Image Fusion Methods", 2015,pp1046-1048

[17] El-Hoseny, H. M., El.Rabaie, E.-S. M., Elrahman, W. A., & El-Samie, F. E. A. (2017). Medical image fusion techniques based on combined discrete transform domains. 2017 34th National Radio Science Conference (NRSC). doi:10.1109/nrsc.2017.7893518

[18] Jagalingam Pa and Arkal Vittal Hegdeb, " A Review of Quality Metrics for Fused Image",2015.135-137

[19] TianLan, ZheXiao, YiLi,Yi Ding, and Zhiguang, |"Medical image fusion techniques based on combined discrete transform domains",2016,pp788-881

[20] Swathi.P.S, Sheethal.M.S, Vince Paul, "Survey on Multimodal Medical Image Fusion Techniques", IJCSET, vol 6, Issue 1, 33-39, 2016.

[21] Z. Wang and A. Bovik, "A universal Image Quality Index," IEEE Signal Processing Letters, vol. 9, no. 3, pp. 81-84, 2002.

[22] C. Yang, J. Zhang, X. Wang, X. Liu, "A novel Similarity based Quality Metric for Image Fusion," Information Fusion, vol. 9, no. 2, pp. 156-160, 2008.

[23] Ijcaonline.org

[24] www.rroioj.com

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:31:41 UTC from IEEE Xplore. Restrictions apply.

Modified Sliding Singular Spectrum Analysis-Based Noncontact-Type Detection of Separated Instrument in Root Canal Therapy Using Low-Frequency Ultrasonic Pattern Transceiver Design

Rajiv Kapoor, Deepak Sharma, and Aarchishya Kapoor

***Abstract*— In root canal therapy, breakage of the instrument in the root canal is the most common problem which may lead to failure of the treatment. The live localization of the separated instrument from root canal is a major challenge. The radiographic imaging and cone beam computed tomography (CBCT)-based detection methods are proposed by various researchers. In this article, a novel approach (noncontact type) is proposed which uses ultrasonic signal for detection and continuous localization of the separated instrument in the root canal. The ultrasonic signal is transmitted in the root canal using 20–80-kHz ultrasonic pattern generator designed using dental gel without directly touching the ultrasound probe to the tooth (noncontact type). The signals from tooth are received by the designed ultrasonic sensor using dental gel again as interface and then analyzed using sliding singular spectrum analysis (SSSA) approach to extract the information about the position of the separated instrument. The detected position of the separated instrument is displayed in real-time on the screen in a doctor-friendly form. The design has been tested on extracted teeth and human jaws at Maulana Azad Institute of Dental Sciences, Delhi, India. This novel approach demonstrates an efficiency of approximately 87% and provides an ultraeconomical solution. The results show that this approach reduces the efforts to detect the position of the separated instrument without the loss of tooth anatomy and helps doctor extract the broken instrument out of root canal.**

***Index Terms*— Deep classifier network, piezoelectric sensor, root canal therapy (RCT), sliding singular spectrum analysis (SSSA), ultrasonic.**

I. Introduction

With the rise in dental diseases day by day, the dentists follow various dental diagnoses and treatment. The dentists preserve the badly damaged or infected tooth by root canal therapy (RCT) [1], [2]. The annoying fact during the RCT is breakage of bit which can inhibit proper cleaning and shaping of the root canal [3]–[5]. The implicit substantial properties, inappropriate use, deficient access, anatomy of root canal, and fabrication deformities are the main reasons of separation of instrument (bit) in the root canal [6]. The diagnosis of tooth having separated instrument in the root canal depends on the phase of the treatment before the breakage of the instrument and possibility of its removal/bypass [7], [8]. There are several attempts to be made for the removal of the separated instrument, and proper cleaning of the root canal is done after each attempt [9], [10]. Sometimes the separated instrument (bit) is removed by surgery as the bit is fixed in the root canal and blocks it [11], [12]. The frequently seen postsurgical efforts are papillae shrinkage, scar tissue formation, and gingival recession [13]. If the actual location of the separated instrument is known, then the separated instrument can be easily extracted with less risk.

Most of the detection methods of separated instruments are based on radiographic approach and cone beam computed tomography (CBCT). Patel *et al.* [14] discussed various applications of CBCT to handle various endodontic problems, whereas Venskutonis *et al.* [15] explained the use of CBCT particularly for the detection of the separated instrument in the root canal. Rosen *et al.* [16] proposed the radiographic identification of broken instrument in the apical third of the root canal and analyzed its ability to detect separated instrument of different materials like stainless steel (SS) or nickel titanium (NiTi). Ramos Brito *et al.* [17] proposed the method of finding broken endodontic instruments in the root channel using digital radiography system and CBCT with different resolutions. Ayatollahi *et al.* [18] evaluated the performance of periapical radiography (PR) and CBCT approaches for identification of separated instruments with the help of various parameters. Alemam *et al.* [19] and Baratto-Filho *et al.* [20] also evaluated CBCT and PR approaches and reported that 2-D radiography is better than CBCT for identification of separated instruments. Although ultrasonic imaging is one of the good choices, it is found in the related work that ultrasonic is used for the extraction of separated instrument from the root canal not for detection. Ward *et al.* [21] proposed a method to remove fractured NiTi instrument using ultrasonic and dental operating microscope. Souter and Messer [22] discussed the complication during retrieval of separated instrument using ultrasonic. Sokhi *et al.* [23] elaborated on the

Manuscript received March 25, 2021; revised May 25, 2021; accepted June 23, 2021. Date of publication July 5, 2021; date of current version July 20, 2021. This work was supported by the Department of Science and Technology, Ministry of Science and Technology, Government of India. The Associate Editor coordinating the review process was Dr. Tarikul Islam. *(Corresponding author: Rajiv Kapoor.)*

Rajiv Kapoor and Deepak Sharma are with the Department of Electronics and Communication Engineering, Delhi Technological University, New Delhi 110042, India (e-mail: rajivkapoor.dtu@gmail.com).

Aarchishya Kapoor is with the Department of Computer Science and Engineering, Thapar Institute of Engineering and Technology, Patiala, Punjab 147004, India.

Digital Object Identifier 10.1109/TIM.2021.3094635

1557-9662 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:18:29 UTC from IEEE Xplore. Restrictions apply.

retrieval of broken instrument in the coronal third of the root canal using ultrasonic and dental operating microscope. Cruz *et al.* [24] removed the separated instrument from the root canal using ultrasound and microtube microscope. Subrata and Hardini [25] and Madarati [26] proposed the use of ultrasonic generator tip to vibrate the broken instrument for its evacuation. For the accurate bit position, the knowledge of accurate shape of the root canal is required. Analoui *et al.* [27] proposed a geometric approach for modeling and measurement of the root canal of a tooth based on stereo digital radiography, whereas Hong and Dong [28] used both radiographic and endoscopic images to prepare a tooth model. Endo *et al.* [29] used ultrasonic imaging and proposed a fuzzy-logic-based root canal detection method. Lee *et al.* [30] used 3-D reconstruction software and micro current transformer (CT) images to detect the 3-D root canal shapes in the first molar. Willershausen *et al.* [31] used X-ray images, whereas van Soest *et al.* [32] proposed an approach based on optical coherence tomography. Germans *et al.* [33] elaborated the virtual-reality-based system for root canal detection. Benyo *et al.* [34] proposed a method to detect the root canal from micro-CT images based on *c*-means clustering [35] and region growing methods [36]. All these techniques either detected the separated instrument using existing ultrasound machines which are not suitable for tooth or used some mechanical instrument to physically extract the separated part using an ultrathin needle-like tool which itself is difficult to be handled by even expert doctors. In remote locations, it is difficult to get a good doctor who can read the ultrasound correctly because ultrasounds contain an extra noise along with signals. Hence, there is a need of techniques which could continuously show the location of separated instrument inside the root canal and that display should show the movement of the separated instrument inside the root canal. This was possible, if one-time X-ray is taken of the patient.

At first, blind source separation (BSS) strategies used independent component analysis (ICA) [37] to extract the wanted signal from multichannel information. Despite the fact that the singular spectrum analysis (SSA) [38] can be effectively applied to separate patterns from acoustics signals, it fails to act on the account of firmly modulated nonstationary signals. Another approach, that is, empirical mode decomposition (EMD) [39], additionally shares similar objectives as SSA. The sliding SSA (SSSA) approach [40] can undoubtedly manage the inadequacies of the past methodologies.

Nowadays, doctors need an assistance system that is capable of giving real-time location of the separated instrument in the root canal during retrieval of the separated instrument. The major drawback in the commonly used methods as described above is that none of the detection methods has the provision to display real-time location. Although ultrasonic has been used in endodontic processes before, it was used for extraction of separated instrument and not for detection purposes. There is a scope to propose a detection method which provides real-time information about the location of separated instrument. It will be very interesting to use ultrasonic for detection of location of the separated instrument in addition to extraction of the same. In this article, a novel ultrasonic-based approach is proposed in which the ultrasonic transmitter probe is placed along the tooth using dental gel (noncontact type) for applying signal to the root canal. Another novelty of the work is the design of the ultrasonic sensor probe for the detection of signals received from the separated instrument in the root canal. After reception of the signal from root canal, the next crucial step is extraction of useful information using modified SSSA [40] technique. Then, the data are given to the trained deep-network-based classifier [41] which indicates the actual position of the bit in the root canal. To show the separated instrument at the specific location in the root canal, a root canal identification method is also proposed in the article. The major contributions of the article are highlighted as follows.

1) The proposed ultrasonic-based detection approach provides the live location of the separated instrument in the root canal during extraction of separated instrument. Such instruments, which can track the location of the broken instrument in root canal, do not exist at present. The instrument design is comparatively economical and will be useful to people who cannot afford expensive tests.
2) The design of ultrasonic sensor is proposed to accurately detect the signal from the separated instrument in the root canal having various resolutions.
3) The modified SSSA approach has been used for the first time for signal decomposition to extract the location information from received signal in endodontic applications.
4) An efficient and less complex approach is applied for identification of the root canal to display the accurate location of separated instrument.

The article is organized as follows: the proposed method in detail including steps of data collection, detection of root canal, and deep network based classifier are discussed in Section II. Section III describes the results and discussions. Finally, the article is concluded in Section IV.

II. Proposed Method

The proposed technique is used to spot the separated instrument in the tooth during endodontic treatment in real-time using noncontact ultrasonic approach as shown in Fig. 1(a).The ultrasonic signal is applied through ultrasonic transmitter probe to all the teeth to record the received signals using designed ultrasonic sensor. Then, useful information about the separated instrument in the teeth is extracted from the received signal using modified SSSA. The knowledge base generated through SSSA, from the teeth database, has been used to train the deep classifier network. When a specimen tooth is used for detection of separated instrument in its root canal, the same process is used for acquisition of signal regarding broken instrument in root canal using ultrasonic sensor and SSSA is used to extract useful information about separated instrument. This information is applied to the trained deep classifier network to get the location of the separated instrument. At the same time, the X-ray image of the same tooth is used to find the root canal coordinates using the root canal detection algorithm. These coordinates are used to display the predicted location of the separated instrument in real-time. When the new location is predicted by the deep

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:18:29 UTC from IEEE Xplore. Restrictions apply.

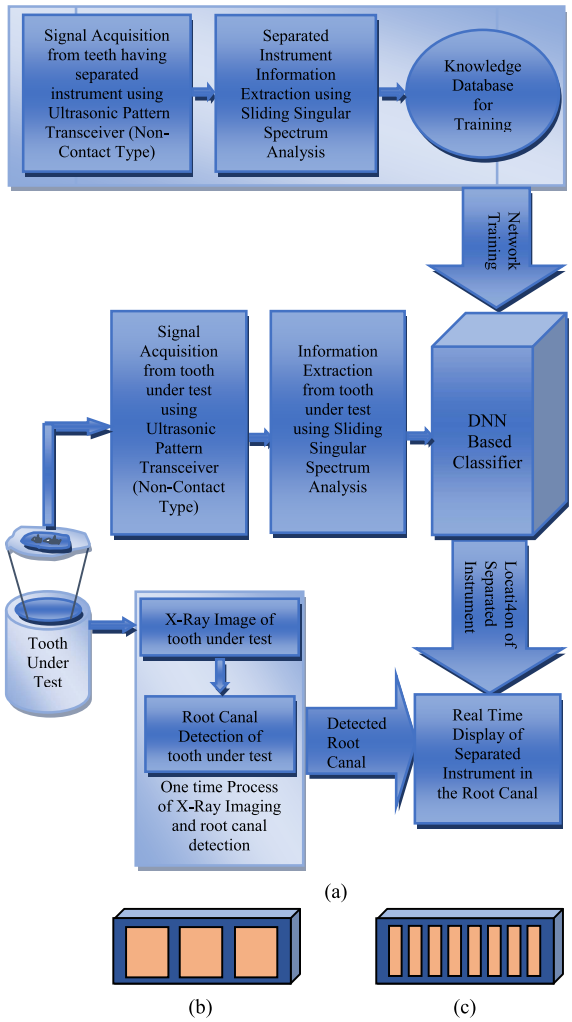


Fig. 1. (a) Basic block diagram of the noncontact-type proposed method used to detect the separated instrument in the tooth in endodontic therapy. (b) Three-crystal ultrasonic sensor. (c) Eight-crystal ultrasonic sensor.

TABLE I
DIMENSIONS OF ULTRASONIC SENSOR FOR DIFFERENT PIEZOELECTRIC CRYSTAL ARRAY SIZES

Piezoelectric crystal Array Size	Number of Piezoelectric Crystals used	Size of Piezo-electric crystal *(l x w)*	Separation Space between Crystals	Total Dimensions of Ultrasonic Sensor *(L x W)*
1*3	3	4 mm*4 mm	0.8 mm	15.2 mm * 5.6 mm
1*8	8	1.6 mm* 4 mm	0.3 mm	15.5 mm * 5.6 mm

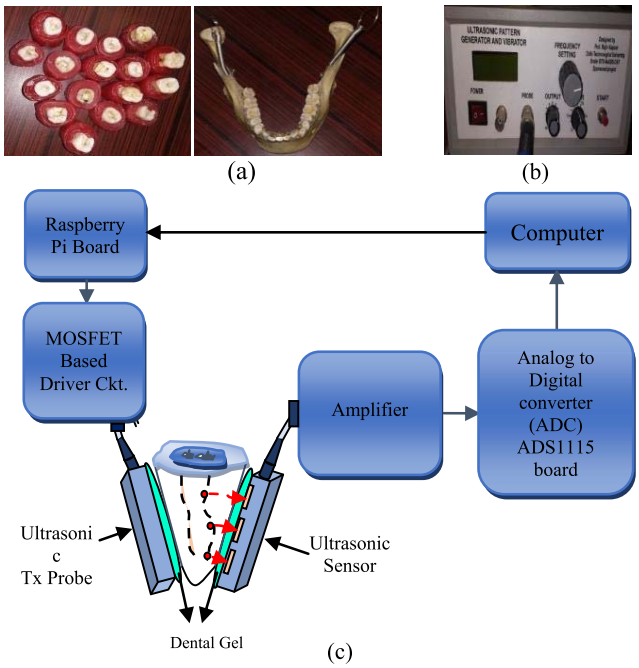


Fig. 2. (a) Teeth and jaw received from MAIDS, Delhi. (b) Ultrasonic pattern generator used to produce 20–80-kHz signal. (c) Experimental setup for signal acquisition from the tooth using ultrasonic transceiver.

classifier, the new location is updated on the screen with the same X-ray image and detected in the root canal. The different steps involved in the proposed work are as follows.

A. Ultrasonic Sensor Design

For data acquisition, the initial and important step is the design of ultrasonic sensor to sense the signal from the separated instrument in the root canal of the tooth. The ultrasonic sensor is designed using "PZT5A" piezoelectric crystals [42] having material composition of lead zirconate titanate (PZT) with a density of 7500 kg/m^3, piezoelectric constant 374 $\times$ 10^{-12} m/V, and coupling coefficient 0.710. The resolution of the ultrasonic sensor depends on the number of piezoelectric crystals used to design the sensor; the resolution can be further improved by increasing the number of piezoelectric sensors in the array. The ultrasonic sensor designed with crystal array of size 1 $\times$ 3 and 1 $\times$ 8 is shown in Fig. 1(b) and (c), respectively. The dimensions of the ultrasonic sensor are fixed to about 15 mm [43] as per the size of the tooth root canal. Therefore, the size of crystals and separation space varies as per the size of the ultrasonic sensor. Table I illustrates dimension details such as size of each piezoelectric crystal, thickness of epoxy separation, and the total size of the ultrasonic sensor for 1 $\times$ 3 and 1 $\times$ 8 crystal array sensor. The piezoelectric crystals are combined in the form of array using hard epoxy. The hard epoxy not only attaches the piezoelectric crystal but also prevents the crosstalk between them. This designed sensor probe is placed along the root canal of the tooth using dental gel to recognize the signal.

B. Data Acquisition

The information is acquired from 1020 teeth having separated instruments at various positions in the root canal. For experimental purpose, these teeth and jaws are provided by the Maulana Azad Institute of Dental Sciences (MAIDS), Delhi, as shown in Fig. 2(a). The ultrasonic [29] pattern generator produced ultrasonic signal of variable frequencies

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:18:29 UTC from IEEE Xplore. Restrictions apply.

20–80 kHz using a raspberry pi board as shown in Fig. 2(b). This signal is applied to the MOSFET driver circuit which is used to operate the ultrasonic transmitter probe. The ultrasonic pattern generator has the provision of varying the amplitude of the ultrasonic and to set the time period of the ultrasonic signal between 0 and 50 s. The noncontact-type experimental arrangement used for signal recording from the root canal of the tooth is presented in Fig. 2(c). The transmitter probe and sensor probe are placed along the root canal of the tooth using dental gel. The received signal is transformed by the sensor into an electric signal e which is given as

$$e = p \times s \times t \tag{1}$$

where p is the pressure applied on the piezoelectric crystal, s is the voltage sensitivity, and t is the thickness of the material. The output signal e from the crystal is applied to the voltage amplifier circuit and then the amplified signal is sent to the analog-to-digital converter (ADC) [44], ADS1115 board, which transforms this signal into a digital signal "u." The ADC converts the signal at the sampling rate of 100 MS/s in batch of 2000 samples.

C. Sliding SSA

The received information is a real-time data having gradually changing patterns with low-frequency components. The SSSA [40] method assists in separating the required information from the nonstationary natural occurring signals. The window of length "L" is slide on the input signal "u" of length "N." There are "p" positions of the window with the progression "Δ" given as

$$p = N - L + 1. \tag{2}$$

"c" and "l" are the middle index and the last index of the frame, respectively, and given as

$$c = p + \frac{L-1}{2} \tag{3}$$

$$l = p + L - 1. \tag{4}$$

Initially, the input signal "$u(p{:}l)$" is transformed into a trajectory matrix "Y" using embedding length "M." The eigen triplets $Y_i \leftarrow \sigma_i x_i y_i$ are determined by singular value decomposition (SVD) [45]. After that, diagonal averaging of each Y_i is performed to extricate $s(i)$ elementary components. Finally, "m" clusters are obtained from $s(i)$ $(i = 1, 2, \ldots, R)$ using agglomerative hierarchical clustering. The yield of automatic SSA $z(j)$ is given as

$$z^i = \text{autoSSA}(u, m, M) \tag{5}$$

where $j = 1, 2, \ldots, m$.

At the primary position of the sliding window (i.e., $p = 1$), the first "c" samples of the estimated components z_c^j are straightforwardly instated with the outcome as

$$\hat{z}_c^j \leftarrow z_c^j, \quad \forall c \in [1, C], \ \forall j \in [1, m]. \tag{6}$$

At that point for the following window positions $(p > 1)$, the current assessed components z_c^j furnished by auto SSA are matched with previous components $\hat{z}_{p-\Delta+c}^{j'}$ using distance metric (Euclidean distance). Every component should be associated only once, consequently to guarantee that the previously associated components "j" are stored in variable "B" (initially $B \leftarrow \Phi$) during matching procedure

$$j' \leftarrow \underset{k' \notin B}{\text{argmin}}\left(\text{dist}\left(z_c^j, \hat{z}_{p-\Delta+c}^{k}\right), \forall c \in [1, C]\right) \tag{7}$$

$$B \leftarrow \{B, j'\}. \tag{8}$$

For the window position $(p < N - L + 1)$

$$\hat{z}_{p-\Delta+c}^{j} \leftarrow z_c^j, \quad \forall c \in [c, c + \Delta - 1]. \tag{9}$$

At the last window position (i.e., $p = N-L+1$), all samples in the range $[c, L]$ of z_c^j are stored in their associated components. The last "M" clusters $\hat{z}_n^{j'}$ (where $j' = 1, 2, \ldots, M$ and $n = 1, 2, \ldots, N$) are toward the end of the algorithm. In the current work, the desired information are the low-frequency components of the received signal, and therefore, the number of clusters is set to $m = 2$, that is, the lower frequency component cluster and other remaining frequency clusters. The low-frequency component $\hat{z}_n^1$ is used as data about the separated instrument location.

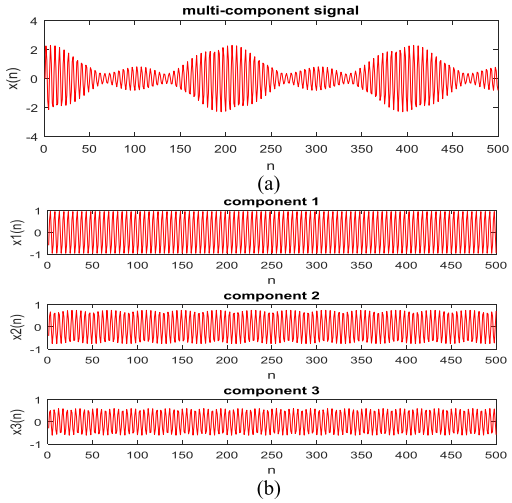


Fig. 3. (a) Multicomponent signal. (b) Signals decomposed from multicomponent signal.

Now, we illustrate how SSSA decomposes a multicomponent signal. Let us consider $N = 500$ samples of multicomponent signals displayed in Fig. 3(a) and expressed as $k_1 = 1.1$, $k_2 = 0.75$, $k_3 = 0.6$, $f_1 = 40$ kHz, $f_2 = 40.5$ kHz, and $f_3 = 41$ kHz. The SSSA method is applied on $x(n)$ to obtain three components. Fig. 3(b) shows the resulting signal components of the signal. These components are reconstructed using SSSA with $r = 3$, the embedding length $M = 40$, and

$$x(n) = x_1(n) + x_2(n) + x_3(n) \tag{10}$$

where

$$x_1(n) = k_1 \sin(2\pi f_1 n)$$

$$x_2(n) = k_2 \sin(2\pi f_2 n)$$

$$x_3(n) = k_3 \sin(2\pi f_3 n)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:18:29 UTC from IEEE Xplore. Restrictions apply.

and the sliding window $W = 181$. The window moves one sample at each position, $\Delta = 1$. The linkage step is based on the Euclidean distance between samples of the preceding and the current subcomponents. The SSSA is capable of decomposing the components having minimal frequency gap.

D. Deep Classifier Network

The data were acquired from the teeth from five particular classes such as top, center, and lower location of the separated instrument, root canal with no instrument, and no teeth conditions. The classifier is designed using deep neural network (DNN) [41] to discover the position of separated instrument in the tooth under test. The DNN is trained using data acquired from 1020 teeth ordered in four known categories. The known training information is obtained from predefined classes using SSSA approach having a size of 16×125, the input size of the DNN is 16, and 100 hidden layers are used. There are five fully connected layers in DNN. The output layer has size 5 according to the output conditions. The stochastic gradient decent with momentum (SGDM) [46] is used as solver for network training. The weight Δw is shown as

$$\Delta w = a\Delta w - n\nabla F_i(w). \tag{11}$$

The updated weight is given as

$$w = w + \Delta w \tag{12}$$

where "w" is the weight which minimizes the function "F," "n" is the learning rate, and "a" is the decay factor having a value somewhere in the range of 0 and 1 to portray the relationship between the contribution of current gradient and earlier gradient to change in weight [41]. The maximum epochs and mini batch size are set as 100 and 1, respectively. The information is shuffled after every epoch during training. At the point when the network is trained, the testing of the sample tooth data is applied to detect the location of the separated instrument in the root canal.

E. Detection of Root Canal

Once the location of the separated instrument is predicted, the next step is to monitor real-time location of the separated instrument in the root canal of the tooth on the screen. X-ray image is used to indicate the separated instrument location using the evaluated coordinates of the root canal. The following stage is to identify the root canal in X-ray image of the particular tooth. Different predicted locations of the separated instrument are mapped to different coordinate locations of the root canal. Then, the real-time location of the separated instrument is displayed in the root canal of the X-ray image at the corresponding coordinates of the root canal. X-ray imaging and root canal recognition is the one-time process for each tooth. Every new predicted location according to the real-time movement of the separated instrument in a particular tooth is indicated on the same X-ray image every time at different coordinates of the same root canal. The X-ray images of 1020 teeth are provided by MAIDS, Delhi, as shown in Fig. 4. The root canal is identified by applying the detection

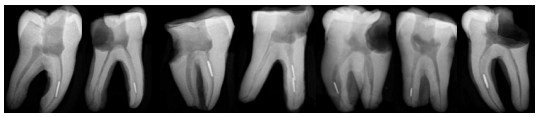


Fig. 4. X-ray images of teeth.

Algorithm 1 Root Canal Detection

Input: X-Ray Image of tooth under test.
Output: Root canal coordinates.
begin
1. Input.
2. Apply Gaussian filtering on the $\boldsymbol{I}$,

$$\boldsymbol{f}[\boldsymbol{i}, \boldsymbol{j}] = \boldsymbol{I}[\boldsymbol{i}, \boldsymbol{j}] * \boldsymbol{G}[\boldsymbol{i}, \boldsymbol{j}]$$

3. Perform the median filtering on Gaussian filter output f,

$$\boldsymbol{f}'[\boldsymbol{i}, \boldsymbol{j}] = \boldsymbol{f}[\boldsymbol{i}, \boldsymbol{j}] * \boldsymbol{m}[\boldsymbol{i}, \boldsymbol{j}]$$

4. Adjust the contrast of the output of median filter f' using CLAHE.
5. Segment the output image of CLAHE into clusters using *k-means* clustering.
6. Calculate the midpoint coordinates between segments vertically and assign pixel value '*255*'.
End

algorithm, as discussed in Algorithm 1. In this algorithm, the Gaussian filter [47] is used to make the small structures invisible and to smooth the root canal territory of the tooth. The function for the 2-D Gaussian filter is represented as

$$g(i, j) = \frac{1}{2\pi\sigma^2} e^{-\frac{i^2+j^2}{2\sigma^2}} \tag{13}$$

where $g(i, j)$ is the Gaussian filter response at the location coordinates (i, j) and σ is the standard deviation. As the value of "σ" increases, it results in additional blurring. The large kernel size $k \times k$ results in narrow distribution peak (less blurring) and vice versa for small kernel size. Even after Gaussian filtering, the image still has small dot-like patterns in the root canal area. The nature of these undesired spots resembles salt and pepper noise as discussed in the following

$$\eta = \begin{Bmatrix} 0 & \text{Peppernoise} \\ 255 & \text{Saltnoise} \end{Bmatrix}. \tag{14}$$

The leftover undesirable small dots' structures undetected by the Gaussian channel are additionally cleared by the median filter which is given as

$$m[i, j] = \text{median}\{f[i, j], (i, j) \in H\} \tag{15}$$

where $m[i, j]$ is the output of the median filter [48], and $f[i, j]$ is the input block to the median filter related to the neighborhood H of size $k \times k$. The tradeoff between noise removal and image quality is to be kept up during kernel size determination. The contrast-limited adaptive histogram equalization (CLAHE) [49] is applied to the output of the median filter to improve the contrast between tooth walls and root canal area that contains root canal veins. The assessment

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:18:29 UTC from IEEE Xplore. Restrictions apply.

of clip limit is acclimated to improve contrast without noise upgrade or dental X-ray images and the ideal estimation of the clip limit is 0.01 [50]. The output image of the CLAHE is additionally divided using k-means clustering [51]. It is a method for vector quantization that segments "M" perception into "N" clusters in which each perception belongs to the cluster having closest mean. The k-mean algorithm [52] focuses at limiting an objective function (squared error function) which is given as

$$K = \sum_{m=1}^{M} \sum_{n=1}^{N} (\|u_m - w_n\|)^2 \tag{16}$$

where $\|u_m - w_n\|$ is the Euclidean distance between a value u_m and a centroid w_n, iterated over all values in the mth cluster for N clusters. The clusters having best segmented regions are picked. The top four areas, that is, two on the either side of the root canal, are perceived from the selected cluster. The left-side edge column coordinates of the first and third regions and the right-side edge coordinates of the second and fourth regions are recognized. After that, the midpoint coordinates of all areas edges of same row are resolved.

III. Results and Discussion

A. Collection of Data and Classifier Training

The information is obtained from 1020 teeth using the setup as shown in Fig. 5(a). The signals from the teeth are sensed by the piezoelectric crystal and given to a voltage amplifier. The amplifier output is applied to ADC boardADS1115 to produce an output as given in Fig. 5(b). Likewise, the digital output signal of ADS1115 is recorded for every tooth of all the five cases. At that point, the modified SSSA algorithm is applied on these signals to find the useful data. There are different parameters which impact the performance of SSSA, that is, embedding length "M" and window length "L." The length of sign "N" is 2000. The embedded length "M" is varied, for example, 20, 40, 60, and 80; also, the window length "L" is likewise changed at various lengths such as 61, 91, 131, and 181. Each one of the five classification signals is analyzed using SSSA for various values of "M" and "L." Fig. 5(c) shows the received signal when there is no tooth available close to the ultrasonic sensor at the embedding length "$M = 40$" and window length "L" above 91. Next, the signal received from the tooth having no separated instrument in the root canal is shown in Fig. 5(d) along with the extracted data at "M" more than 40 and window length "L" around 181. Fig. 5(e) shows the received signal from the tooth having separated instrument at the top area of the root canal for "M" $= 40$ and "L" $= 91$. After this, the received signal and the extricated data of the tooth having separated instrument at the center of the root canal are shown in Fig. 5(f) at an embedding length of "M" $= 15$ and window length "L" $=$ 161 that has a delayed peak with lower amplitude than the top position peak. At that point finally, the received signal and the extricated data from the tooth having separated instrument at the bottom of the root canal are shown in Fig. 5(g). The extricated data peak at the bottom location has lower amplitude

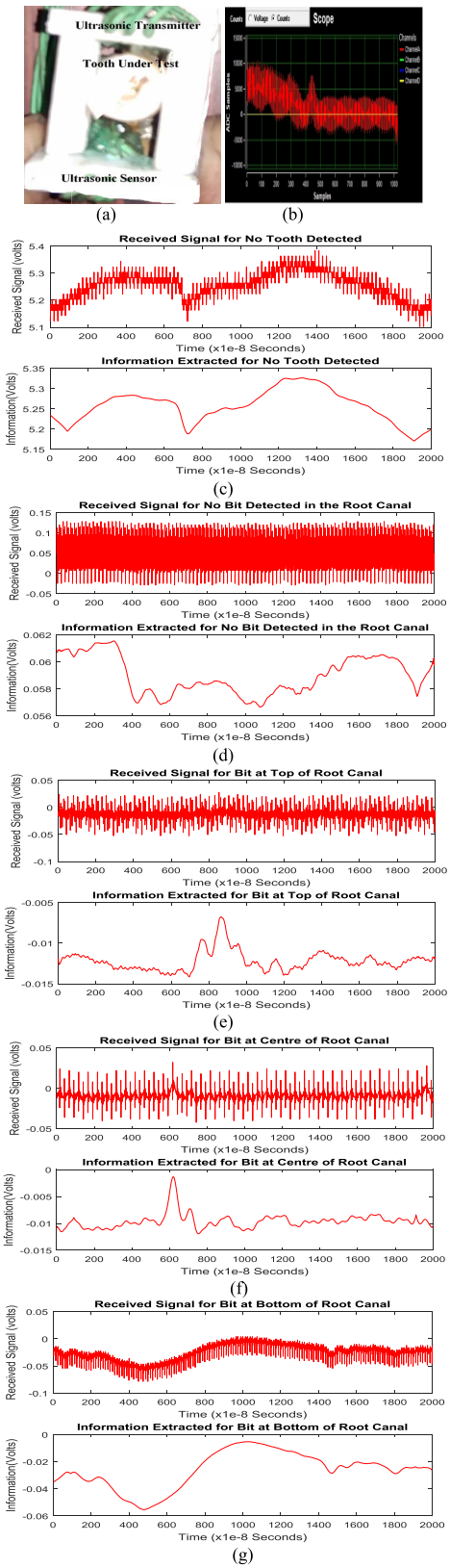


Fig. 5. (a) Data acquisition setup. (b) Received ultrasonic signal. (c) Received signal and extracted information for the no-tooth condition. (d) For no bit in the root canal condition. (e) For bit at the top location in the root canal. (f) For bit at the center location in the root canal. (g) For bit at the bottom location in the root canal.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:18:29 UTC from IEEE Xplore. Restrictions apply.

and delayed time instant than the peak of the center position at "M" = 40 and "L" = 91. Each extricated signal is of length 1 × 2000. The information is altered to measure 16 × 125 for every tooth. As the information from every tooth of the collection of teeth is obtained, 1020 datasets are framed. These datasets are used to train the deep network classifier as discussed in Section II.

The SGDM is used as a training solver and details such as input size, number of layers, mini batch size, and number of epochs are discussed in Section II which are used to train the network. At the outset iterations, when the accuracy is fluctuating, the loss has high value. In the higher iterations, as the accuracy in the preparation cycle increases, the loss decline tends to "0."

B. Detection of Root Canal

After the training of network, the root canal detection of each tooth is performed on X-ray image as per the methodology explained in Section II. Above all, the X-ray image is preprocessed to clean undesired subtleties from the root canal area of the teeth. Gaussian filtering at different kernel sizes and standard deviation are preceded. However, the output of the Gaussian filter at a kernel size of 3 × 3 and standard deviation "σ" = 1 is most appropriate for the smoothing of root canal zone. The output of the Gaussian filter is given to the median filter to remove the remaining little dab-like structures. The output of the median filter is examined at various kernel sizes, but the kernel size 5 × 5 is the most trading off an incentive for both smoothing and image quality viewpoint for this dataset. The contrast of the median filter image is improved using CLAHE at a clip limit of 0.01 [43]. Then, k-means clustering is applied to the preprocessed X-ray image with updated contrast to produce cluster images. The midpoint of gap is calculated and plotted as root canal.

C. Identification of Location of Separated Instrument

The data acquired from the tooth under test are applied to the trained classifier network that predicts the location of the separated instrument in the tooth. By then, its output is used to demonstrate the separated instrument location at the root canal of the tooth in the image. In this article, three locations of the separated instrument are recognized, that is, top, center, and bottom with two more categories, such as no bit in root canal and no tooth accessible for identification.

The separated instrument at the bottom location in the tooth is given in Fig. 6(a). The separated instrument is perceived at the center position due to two reasons: either the instrument is separated at the center of the root canal or it is moved up from the bottom location. The separated instrument identified at the center of the root canal is shown in Fig. 6(b). Likewise, the separated instrument is detected at the upper part of the root canal; either the instrument is separated at that location of root canal or is moved up from the center location during the extraction cycle of the separated instrument from the tooth. Fig. 6(c) shows the separated instrument at the top of the root canal. At the end, Fig. 6(d) and (e) shows the conditions when there is no separated instrument recognized in the root canal

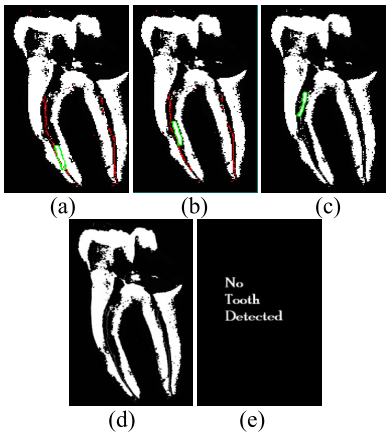


Fig. 6. (a) Detected separated instrument at the bottom location. (b) Detected separated instrument at the center location. (c) Detected separated instrument at the top location. (d) No bit detected in the root canal. (e) No teeth detected.

and there is no tooth detected, respectively. As discussed in Section II, the resolution for the proposed work depends on the number of crystals in the ultrasonic sensor probe. The ultrasonic sensor having three crystals has resolution equal to the length of root canal by 3, that is, 15 mm/3 mm = 5 mm. As the numbers of crystal in the sensor are increased to eight, the resolution improves to 15 mm/8 mm = 1.875 mm. The method shows the improvement in accuracy to 87% with an increase in number of crystal in the ultrasonic sensor. This design has been tested on the extracted teeth and human jaws at MAIDS, Delhi, India.

IV. Conclusion and Future Scope

In this article, we have proposed a novel ultrasonic-based methodology to identify the location of the separated instrument in the root canal of the tooth. The ultrasonic is applied in the root canal of tooth using noncontact-type method and the received signals are used to recognize the location of the separated instrument in the root canal of the tooth. The recognized location of the separated instrument is helpful for the extraction of the separated instrument from the tooth. Therefore, the proposed strategy can be exceptionally robust and helpful as it avoids harm to tooth during extraction of the separated instrument. This approach not only gives good recognition of the separated instrument in the tooth but also the instrument design is comparatively economical and will be useful to people who cannot afford costly tests. The design of ultrasonic sensor using 128 crystals is under development. In future, it will be interesting to recognize the isolated instrument using this ultrasonic sensor with improved resolution.

Acknowledgment

The authors would like to express their deep gratitude to the Department of Science and Technology for their financial assistance to the research project. They would also like to thank Maulana Azad Institute of Dental Sciences (MAIDS), Delhi, (especially Dr. Amit Malhotra, Dr. Ruchika Rungta Nawal, and Dr. Sangeeta Talwar) for providing valuable suggestions and dental data necessary for the research.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:18:29 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] H. Patil and S. A. More, "Dental image processing in endodontic treatment," *Int. J. Electron., Electr. Comput. Syst.*, vol. 6, no. 9, p. 8887, Sep. 2017.

[2] A. Gandevivala, B. Parekh, G. Poplai, and A. Sayed, "Surgical removal of fractured endodontic instrument in the periapex of mandibular first molar," *J. Int. Oral Health*, vol. 6, no. 4, p. 85, 2014.

[3] A. Kaufman and H. Neuman, "Iatrogenic damages caused by dental procedures. Foreign bodies in the oral cavity," *Quintessence Int. Dental Dig.*, vol. 14, no. 3, pp. 361–366, 1983.

[4] L. M. Lin, P. A. Rosenberg, and J. Lin, "Do procedural errors cause endodontic treatment failure?" *J. Amer. Dental Assoc.*, vol. 136, no. 2, pp. 187–193, Feb. 2005.

[5] A. A. Madarati, D. C. Watts, and A. J. Qualtrough, "Factors contributing to the separation of endodontic files," *Brit. Dental J.*, vol. 204, no. 5, pp. 241–245, 2008.

[6] B. Sattapan, G. J. Nervo, J. E. Palamara, and H. H. Messer, "Defects in rotary nickel-titanium files after clinical use," *J. Endodontic*, vol. 26, no. 3, pp. 161–165, 2000.

[7] A. Heydari, M. Rahmani, and M. Heydari, "Removal of a broken instrument from a tooth with apical periodontitis using a novel approach," *IEJ Iranian Endodontic J.*, vol. 11, no. 3, pp. 237–240, 2016.

[8] P. Spili, P. Parashos, and H. H. Messer, "The impact of instrument fracture on outcome of endodontic treatment," *J. Endodontic*, vol. 31, no. 12, pp. 845–850, 2005.

[9] O. Nagai, N. Tani, Y. Kayaba, S. Kodama, and T. Osada, "Ultrasonic removal of broken instruments in root canals," *Int. Endodontic J.*, vol. 19, no. 6, pp. 298–304, Nov. 1986.

[10] J. L. Saunders, P. D. Eleazer, P. Zhang, and S. Michalek, "Effect of a separated instrument on bacterial penetration of obturated root canals," *J. Endodontic*, vol. 30, no. 3, pp. 177–179, 2004.

[11] M. Hulsmann, "Methods for removing metal obstructions from the root canal," *Endodontic Dental Traumatol.*, vol. 9, no. 6, pp. 223–237, 1999.

[12] B. Suter, A. Lussi, and P. Sequeira, "Probability of removing fractured instruments from root canals," *Int. Endodontic J.*, vol. 38, no. 2, pp. 112–123, 2005.

[13] T. von Arx, "Apical surgery: A review of current techniques and outcome," *Saudi Dental J.*, vol. 23, no. 1, pp. 9–15, Jan. 2011.

[14] S. Patel, A. Dawood, T. P. Ford, and E. Whaites, "The potential applications of cone beam computed tomography in the management of endodontic problems," *Int. Endodontic J.*, vol. 40, no. 10, pp. 818–830, Oct. 2007.

[15] T. Venskutonis, G. Plotino, G. Juodzbalys, and L. Mickeviciene, "The importance of conebeam computed tomography in the management of endodontic problems: A review of the literature," *J Endod*, vol. 40, pp. 1895–1901, 2014.

[16] E. Rosen, H. Azizi, C. Friedlander, S. Taschieri, and I. Tsesis, "Radiographic identification of separated instruments retained in the apical third of root Canal–filled teeth," *J. Endodontics*, vol. 40, no. 10, pp. 1549–1552, Oct. 2014.

[17] A. C. Ramos Brito *et al.*, "Detection of fractured endodontic instruments in root canals: Comparison between different digital radiography systems and cone-beam computed tomography," *J. Endodontics*, vol. 43, no. 4, pp. 544–549, Apr. 2017.

[18] F. Ayatollahi, M. Tabrizizadeh, H. Razavi, and M. Mowji, "Diagnostic value of cone-beam computed tomography and digital periapical radiography in detection of separated instruments," *Iran Endod. J.*, vol. 14, pp. 14–17, Jan. 2019.

[19] S. Alemam, S. Abuelsadat, S. Saber, and T. Elsewify, "Accuracy, sensitivity and specificity of three imaging modalities in detection of separated intracanal instruments," *G. Ital. Endod.*, vol. 34, pp. 97–103, Jun. 2020.

[20] F. Baratto-Filho, J. V. de Freitas, F. S. F. Tomazinho, M. C. L. Gabardo, F. Mazzi-Chaves, and M. D. Sousa-Neto, "Cone-beam computed tomography detection of separated endodontic instruments," *J Endod.*, vol. 2399, no. 20, pp. 1776–1781, 2020.

[21] J. R. Ward, P. Parashos, and H. H. Messer, "Evaluation of an ultrasonic technique to remove fractured rotary nickel-titanium endodontic instruments from root canals: Clinical cases," *J Endod.*, vol. 29, no. 11, pp. 764–767, 2003.

[22] N. J. Souter and H. H. Messer, "Complications associated with fractured file removal using an ultrasonic technique," *J Endod.*, vol. 31, pp. 450–452, Jun. 2005.

[23] V. Shenoy and R. Sokhi, "Retrieval of separated instrument using ultrasonics in a permanent mandibular second molar: A case report," *J. Contemp. Dentistry*, vol. 4, no. 1, pp. 41–45, 2014.

[24] A. Cruz, C. G. Mercado-Soto, I. Ceja, L. G. Gascon, P. Cholico, and C. A. Palafox-Sanchez, "Removal of an instrument fractured by ultrasound and the instrument removal system under visual magnification," *J Contemp Dent Pract.*, vol. 16, no. 3, pp. 238–242, 2015.

[25] A. Subrata and N. Hardini, "Removing a fractured instrument from the root canal using ultrasonic tips," *Sci. Dental J.*, vol. 3, no. 3, p. 95, 2019.

[26] A. A. Madarati, "Retrieval of multiple separated endodontic instruments using ultrasonic vibration: Case report," *J. Taibah Univ. Med. Sci.*, vol. 11, no. 3, pp. 268–273, Jun. 2016.

[27] M. Analoui, S. Krishnamurthy, and C. Brown, "Modeling and measurement of root canal using stereo digital radiography," *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 3976, pp. 306–314, Apr. 2000.

[28] S. Y. Hong and J. Dong, "3-D root canal modeling for advanced endodontic treatment," *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 4702, pp. 321–330, Jun. 2002.

[29] M. Endo, S. Kobashi, K. Kondo, and Y. Hata, "Dentistry support ultrasonic system for root canal treatment aided by fuzzy logic," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, vol. 2, Oct. 2005, pp. 1494–1499.

[30] J.-K. Lee, B.-H. Ha, J.-H. Choi, S.-M. Heo, and H. Perinpanayagam, "Quantitative three-dimensional analysis of root canal curvature in maxillary first molars using micro-computed tomography," *J. Endodontics*, vol. 32, no. 10, pp. 941–945, Oct. 2006.

[31] B. Willershausen, A. Kasaj, B. Röhrig, and B. B. Marroquin, "Radiographic investigation of frequency and location of root canal curvatures in human mandibular anterior incisors *in vitro*," *J. Endodontics*, vol. 34, no. 2, pp. 152–156, Feb. 2008.

[32] G. van Soest, H. Shemesh, M. K. Wu, L. W. M. van der Sluis, and P. R. Wesselink, "Optical coherence tomography for endodontic imaging," *Proc. SPIE Prog. Biomed. Opt. Imag.*, vol. 6843, Feb. 2008, Art. no. 68430F.

[33] D. M. Germans, H. J. W. Spoelder, L. Renambot, H. E. Bal, S. van Daatselaar, and P. van der Stelt, "Measuring in virtual reality: A case study in dentistry," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 6, pp. 1177–1184, Jun. 2008.

[34] B. Benyo, L. Szilagyi, T. Haidegger, L. Kovacs, and C. Nagy-Dobo, "Detection of the root canal's centerline from dental micro-CT records," in *Proc. Annu. Int. Conf. Eng. Med. Biol. Soc.*, Minneapolis, MN, USA, Sep. 2009, pp. 3517–3520.

[35] M. Gong, Y. Liang, J. Shi, W. Ma, and J. Ma, "Fuzzy C-means clustering with local information and kernel metric for image segmentation," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 573–584, Feb. 2013.

[36] S. A. Hojjatoleslami and J. Kittler, "Region growing: A new approach," *IEEE Trans. Image Process.*, vol. 7, no. 7, pp. 1079–1084, Jul. 1998.

[37] E. Oja, A. Hyvarinen, and J. Karhunen, *Independent Component Analysis*. Hoboken, NJ, USA: Wiley, 2001.

[38] R. Vautard and M. Ghil, "Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series," *Phys. D, Nonlinear Phenomena*, vol. 35, no. 3, pp. 395–424, May 1989.

[39] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London. Ser. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[40] J. Harmouche, D. Fourer, F. Auger, P. Borgnat, and P. Flandrin, "The sliding singular spectrum analysis: A data-driven nonstationary signal decomposition tool," *IEEE Trans. Signal Process.*, vol. 66, no. 1, pp. 251–263, Jan. 2018.

[41] G. P. Zhang, "Neural networks for classification: A survey," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 30, no. 4, pp. 451–462, Nov. 2000.

[42] Q. Zhou, K. H. Lam, H. Zheng, W. Qiu, and K. K. Shung, *Piezoelectric Single Crystals for Ultrasonic Transducers in Biomedical Applications*. Amsterdam, The Netherlands: Elsevier, 2014.

[43] S. Y. Kim, S. H. Lim, S. N. Gang, and H. J. Kim, "Crown and root lengths of incisors, canines, and premolars measured by cone-beam computed tomography in patients with malocclusions," *Korean J. Orthodontics*, vol. 43, no. 6, pp. 271–278, 2013.

[44] S. Boe and K. Kristoffersen, "Predictive analog to digital conversion of Doppler ultrasound signals," *IEEE Trans. Biomed. Eng.*, vol. 42, no. 3, pp. 260–268, Mar. 1995.

[45] G. W. Stewart, *Introduction to Matrix Computations*. New York, NY, USA: Academic, 1973.

[46] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:18:29 UTC from IEEE Xplore. Restrictions apply.

[47] J. Flusser, S. Farokhi, C. Hoschl, T. Suk, B. Zitova, and M. Pedone, "Recognition of images degraded by Gaussian blur," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 790–806, Feb. 2016.
[48] G. George, R. M. Oommen, S. Shelly, S. S. Philipose, and A. M. Varghese, "A survey on various median filtering techniques for removal of impulse noise from digital image," in *Proc. Conf. Emerg. Devices Smart Syst. (ICEDSS)*, Mar. 2018, pp. 235–238.
[49] K. Singh, D. K. Vishwakarma, G. S. Walia, and R. Kapoor, "Contrast enhancement via texture region based histogram equalization," *J. Modern Opt.*, vol. 63, no. 15, pp. 1444–1450, Aug. 2016.
[50] J. Joseph, J. Sivaraman, R. Periyasamy, and V. R. Simi, "An objective method to identify optimum clip-limit and histogram specification of contrast limited adaptive histogram equalization for MR images," *Biocybern. Biomed. Eng.*, vol. 37, no. 3, pp. 489–497, 2017.
[51] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. Piatko, R. Silverman, and A. Y. Wu, "The analysis of a simple K-means clustering algorithm," in *Proc. 16th Annu. ACM Symp. Comput. Geometry*, May 2000, pp. 100–109.
[52] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient K-means clustering algorithm: Analysis and implementation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 881–892, Jul. 2002.

Rajiv Kapoor received the B.E. degree in electronics and communication engineering from Jamia Millia Islamia University, New Delhi, India, in 1988, the M.E. degree in electronics and communication engineering from the Delhi College of Engineering, Delhi University, New Delhi, in 1993, and the Ph.D. degree in electronics and communication engineering from Punjab Engineering College, Panjab University, Chandigarh, India, in 2004.

He is currently a Professor with the Department of Electronics and Communication Engineering, Delhi Technological University (DTU), New Delhi, India. He is currently working on multiple industry-sponsored projects. He has extensive industry and academic experience. He has published seven books with a total experience of more than 30 years. He has authored or coauthored more than 200 research articles in various renowned international journals and conferences and has an H-index of 24. His research interests include computer vision, machine learning, signal and image processing, sensor design, and so on.

Deepak Sharma received the B.Tech. degree in electronics and instrumentation engineering from Kurushetra University, Kurukshetra, India, in 2005, and the M.Tech. degree in electronics and communication engineering from Maharishi Markandeshwar (Deemed to be University), Ambala, India, in 2009.

He is currently a Junior Research Fellow with the Department of Electronics and Communication Engineering, Delhi Technological University (DTU), New Delhi, India, and has a total experience of more than 15 years. He has research articles in renowned international journals and conferences. His research interests include image processing, computer vision, and machine learning.

Aarchishya Kapoor is currently pursuing the B.Tech. degree in computer science and engineering with the Thapar Institute of Engineering and Technology, Patiala, Punjab, India.

She holds four patents and two SCI Journal articles and also owns a Startup registered with NIT Jalandhar. She is working in the field of artificial intelligence, machine learning, and fusion algorithms.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:18:29 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Trends in Electronics and Informatics (ICOEI).
IEEE Xplore Part Number:CFP21J32-ART; ISBN:978-1-6654-1571-2

Movie genres and beyond: An analytical survey of classification techniques

Avleen Singh Mehal
Dept. of Software Engineering
Delhi Technological University
New Delhi, India
avleenmehal@gmail.com

Kalpana Meena
Dept. of Software Engineering
Delhi Technological University
New Delhi, India
kalpanameena_2k17se52@dtu.ac.in

Rijul Bir Singh
Dept. of Software Engineering
Delhi Technological University
New Delhi, India
rijulbirsingh_2k17se93@dtu.ac.in

Prashant Giridhar Shambharkar
Dept. of Computer Engineering
Delhi Technological University
New Delhi, India
prashant.shambharkar@dtu.ac.in

2021 5th International Conference on Trends in Electronics and Informatics (ICOEI) | 978-1-6654-1571-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICOEI51242.2021.9453021

***Abstract*— In today's world with a plethora of options available on movie content, it is very important to have models which could classify these large contents based on their genres. Movies are the most widely entertaining source all over the world. Every person has his own taste of movies, some feel action movies are better, some would say romantic. So the online platforms must have the collection of movies according to a particular category i.e. based on genres. This paper has proposed an extensive and exhaustive survey for various models and techniques, which could be applied to solve the challenges involved in performing the genre based classification. This paper summarizes and compares all those techniques by classifying them into four broad domains -through text mining, poster detection, audio and video features of the movies. These domains included various classifying techniques: Bi-directional LSTM, gram layer, scene categorization and CNN respectively. With such an extensive amount of work going in the field in the last not many years, this review paper will help specialists in the field of movie genre classification.**

I. INTRODUCTION

Video based content assume a major function in satisfying individuals' delight needs. On account of Internet innovation headways, content consumers approach a remarkable measure of motion pictures from different online administrations.//With increasing number of users using the internet in India and around the world have shot up multiple folds in recent past, thanks to free internet and lower rate/cheap internet packs. Because of this there has been a sudden increase in number of people engaging in consumption of online entertainment channels having web shows, web series, movies, daily soaps etc. This has made the need for programmed content-driven moving-picture show suggestions of genres before a user starts watching it.

Number of interesting models has been already proposed by researchers in the recent past. In [1] algorithm of SVM has een extensively used along with text containing information about the video which helps in the classification process. Main focus in this model is given to the three basic steps for text based classification i.e. word segmentation, feature engineering, text representation, and designing followed by implementation of a movie classification system based on SVM algorithm[1] follows the basic framework of classification as shown below:

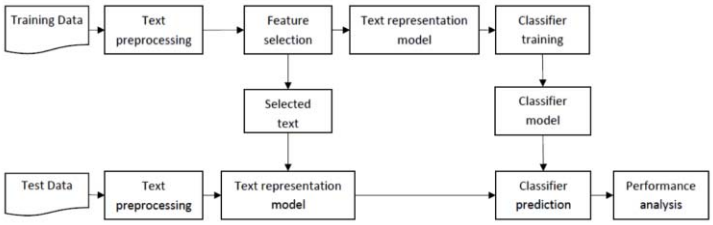


Fig. 1. General architecture for classification

This survey paper summarizes, discusses and compares the recent work of researchers, journal authors and data scientists working on novel ideas of classifying video content published on the web based on their genres. The work discussed in this survey paper is done in the last 3-4 years. This survey paper consists of 6 sections which are presented as follow. Section 2 discusses the background: presented various techniques used by different authors on the classifying problem of video content. Section 3 consists of extensive discussion on existing methodologies. Firstly the technique of bi directional LSTM is discussed, which uses plots of the video to classify its genre. Secondly, the method of Deep Neural Network (DNN) has been discussed for performing classification based on the poster of a movie video followed by a brief discussion on CNN layers classification technique. Section 4 lists the information about famous datasets that are publicly available. Section 5 summarizes the various methods categorized by us.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:24:51 UTC from IEEE Xplore. Restrictions apply.

Section 6 discusses the future scope and challenges in this domain and concludes this survey paper.

II. BACKGROUND

Research papers by different authors can be classified on the basis of approaches they have adopted.

Samuel S. & Chokshi [2] together dissected the film classification by utilizing banners of motion pictures by a profound learning model that distinguishes the film sort passed on by a film banner. To track down an ideal model, they executed modified variants of three standard profound learning structures for picture order: ResNet-50, VGG-16, and DenseNet-169. DenseNet-169 gave us the best outcomes, beating the other two models and models planned in past endeavours at the problem. They expressed that, the utilization of an item identifier calculation like YOLO, assortment of better information, and use of other convolutional neural organization (CNN) designs to the issue.

On comparable lines Gabriel and Kris [3] effectively portrayed that ResNet34 and a custom engineering were sensibly fruitful at foreseeing a banner's classifications. This explains that there is some steadiness in movie visual parts remembered for film banners of a given sort.

Yadav, A. and Vishwakarma, D. K [4] investigated a bound together framework of deep networks for type characterization utilizing film trailers. Their work meant to break down the connection between the feelings inspired by the film trailers and how they contribute in settling the multi-mark type grouping problem. The proposed novel system is approved by performing cross-dataset testing for three huge scope datasets, to be specific LMTD-9, MMTF14K, and ML-25M datasets. Extensive analyses showed that the proposed calculation beat all the best in class strategies.

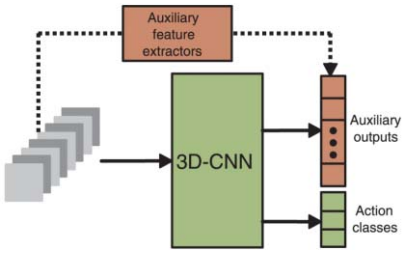


Fig. 2. Explanation of 3D CNN model

Shambharkar, P. G., and Doja, M. N. [5] did the film trailer grouping utilizing deer chasing advancement based profound convolutional neural organization in video sequences. Features was prepared and ordered with the assistance of upgraded deep convolutional neural network (DCNN) for the film trailer classifications. The consequences of the proposed strategy were contrasted and separating measure and without sifting measure in which 95.23% of precision were accomplished for the recommended approach with sifting and 90.91% of exactness was accomplished for the proposed approach without separating measure.

Varshit, Vishal, Rohit, Krishna Reddy, Mamidi [6] anticipated the Genre and Rating of a Movie Based on its Synopsis. They made the Multi-Language Movie Review Dataset (MLMRD). The dataset comprises classification, rating, and summary of a film across various dialects, to be specific Hindi, Telugu, Tamil, Malayalam, Korean, French, and Japanese. In this work, they give different model designs that can be utilized to anticipate the class and the rating of a film across different dialects present in our dataset dependent on the outline.

K Sivaraman and Gautam Somappa [7] dealt with distinguishing the class of a film by investigating simply the visual highlights of its trailer. In this they made a vision system that can do likewise, precisely. They talked about the methodologies they took and exploratory perceptions. The commitments of this work are: (1) proposes a neural organization (in view of VGG) that can characterize film trailers dependent on their sorts; (2) release a curated dataset, called YouTube-Trailer Dataset, which has more than 800 film trailers crossing more than 4 types, accomplishing a precision of 80.1% with the spatial highlights, and 85% with utilizing LSTM and set these outcomes as the benchmark for this dataset.

Y.F. Huang and S.H. Wang [8] thought of a methodology that utilizes the SAHS (Self Adaptive Harmony Search) calculation in choosing highlights for various film types. After this, a Support Vector Machine is stacked with a bunch of highlights which comprises of 277 highlights separated from each film trailer. After this, they utilized a greater part casting a ballot method to gauge the class of the film. They utilize both visual just as sound qualities in this method and utilize a limit of 25 highlights to recognize two distinct types of motion pictures.

S. Ji, W. Xu, Kai Yu [9] chipped away at paper that thinks about the computerized acknowledgment of human activities in observation recordings. In this paper, they built up a novel 3D CNN model for activity acknowledgment. This model concentrates highlights from both the spatial and the transient measurements by performing 3D convolutions, subsequently catching the movement data encoded in various nearby frames.

III. EXISTING METHODOLOGY

A. Bi-Directional LSTM

One major domain to classify movie genre is by taking plot summaries as a dataset and training different models on them

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:24:51 UTC from IEEE Xplore. Restrictions apply.

TABLE I. COMPARISON OF BENCHMARK TECHNIQUES

Year of publication	Dataset used	Technique	Observation	Accuracy
2020[10]	Approx. 21,000 movie posters dating from 1913 to 2019	Gram layer	Genre classification of video content based on poster classification using Inter-Channel features utilises Gram layers, identifying the characteristics of posters and the correlations among every genre, thus engendering better accuracy multi-genre classification. Gram layer brings better focus on videos based on content features of posters than on the attention mechanism, which focuses on basic highlights of photos.	Fscore-0.466 Accuracy-0.45 Precision-0.57 Recall-0.45
2020 (ICICCS)[11]	Movie trailers from Youtube playlists	3D Convulational Neural Networks	The benefit of 3D CNN is its capacity to catch the spatial just as the fleeting data present in a film trailer. 3D CNN had an accuracy of 82.14% which is more noteworthy than different models which use simply 2D CNNs	Accuracy -82.14%
2020 (ICICT)[12]	Deap and Decaf datasets containing movie clips	EEG, MEG, MCA	Identification based on the feelings that are seen by the watchers while watching media cuts is accomplished. The analysis demonstrated that the coupling of magnetoencephalographic signs and multimedia media signals gives higher performance idicator. The high exactness is acquired with this combination alone is on the grounds that the MEG signals have low commotion, accordingly, giving more spatial granularity of brain exercises	RF-0.722 SVM-0.55 KNN-0.59 NB-0.46
2018 (IEEE)[13]	MovieLens dataset of plot summaries	Bidirectional LSTM (Bi-LSTM).	Preparing Bi-LSTM network subsequent to separating the plot outlines into their sentences and intertwining the forecasts for singular sentences beat preparing the organization with the entire plot synopses with the restricted measure of information.	Precision- 67.75% Recall-67.61% F-score-67.68%
2019 (IEEE)[14]	ImageNet dataset	Image based Deep learning model, CNN, LSTM	In the picture cycle, a progression of key-outline highlights got by utilizing the CNN so the LSTM used to become familiar with the unique highlights. Results demonstrated that the deep learning model displays unrivaled execution than the customary techniques. The model based on Deep learning improves the RR% about 2% than the deep learning model introduced in 2017.	CNN+LSTM-77.8% Proposed method without CNN-78.8% , and with CNN 79.7%
2018 (Springer)	IMDB dataset for plot summaries	Topological Data Analysis, NLP	Various general research models that are widely used such as such as Logistic Regression for binary content characterization, Multinomial Naive Bayes and furthermore can perform better and beat further developed procedures like neural organizations when in multi-label text grouping. On the errand of characterization of videos as indicated by four most regular classes, this acquired a Jaccard score of 54.8%, F-score of 71.88%, and hit pace of 83.3%	F-score 71.88% Jacquard index 54.8%
2018 (Springer)	Wikipedia Movie Plots	Text Mining	This paper proposes a framework which takes film plot to be an information and naturally gives a small amount of types present in that film as a yield. Application of modelling procedures, like EM algorithms and LDA, performance scores are boosted. Insight of group theory as film audit can be utilized as a dataset to anticipate the film type data.	Error 7.69%
2010 (ACM international conference)[15]	Movie Trailers	Shot Boundary Detection, Scene Categorization	Genre Classification by means of Scene Categorization utilizes a methodology wherein every trailer is decayed into an assortment of key frames by the means of the model i.e. shot boundary analysis. From these key frames, we use best in class scene indicators and descriptors to extricate highlights. This permits to address trailers applying a bow model with shot classes as vocabularies. The genre grouping task was drawn closer by planning bow transiently organized trailer highlights to four significant level film sorts: action, comedy, drama or thrillers	GIST-71.8% CENTRIST-74.7% W-CENTRIST-73.2%
2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)[16]	Small dataset of movie clip samples	Gaussian kernel SVM prediction model with VC dimension	While implementing support vector machine (SVM) arrangement model and a logistic regression (LR) model was set up to extricate highlights from test information and their test mistake in-example was compared. The accuracy of the calculation expanded to 92% with a more modest VC measurement and less over fitting.	Prediction accuracy of around 92% and 91% were obtained

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:24:51 UTC from IEEE Xplore. Restrictions apply.

Augmented Human Research (2020)[17]	30,000 images From IMDB using a web crawler	Neural networks with knowledge transfer learning	A methodology for the age of a multi-label film sort forecast from its banner utilizing neural organizations are used in the paper. The methodology chips away at two fronts; one is pointed toward making an enormous, various and adjusted dataset for film type forecast. The subsequent front includes re-examining the issue to more straightforward single-name multi-class characterization and producing a multi-label forecast on a given film banner as input.	
2020[18]	500,000 subtitles in Dutch and English for a broad variety of movies.	AdaBoost, C4.5, Naive Bayes, Random Forest, and SVM	Text-based video genre classification concluded that Random Forest and SVM were found to be the strongest algorithms amongst all. Both algorithms are known to be effective on larger datasets. The experimental results showed that of the three feature types, the content-specific features result in the most accurate classifier.	F1-score Linear SVM-0.88 Random Forest-0.83
ASC 2017[19]	LMTD dataset for movie trailers	Multi-Label approach based on Convolutions through Time, multi-label classification	A new model based on neural architecture is proposed relying on convolutional neural networks to perform multilabel video genre classification. It even extends itself to compare this approach with the current state-of-the-art methods for video genre classification. The proposed architecture's performance outplays all the compared models and architectures by the means of various comparison techniques and hence makes the classification even more efficient.	
2019 (ICIT)[1]	Public Film Database	TF-IDF and SVM	It centers around the different advances that should be finished during the process of classification, for example, text word division, including designing, text portrayal, and so on, and plans and executes a film grouping framework dependent on SVM calculation.	

to get results. Several papers are already published on this area of work. Ali Mert Ertugrul et al [13] proposed a classification model called the bi directional LSTM model which showed promising results and accuracy. Plot summaries which are taken as the primary dataset consists of various sentences and assuming that these summaries reflect genre of movie are separated into different sentences. Each sentence is trained with this model and genre of each sentence is generated. In this model dataset as movie names is retrieved from MovieLens[20]. Movie plot summaries are sent in as input and genres are retrieved as the output. This process is done through OMDB API[21]. After data is collected it is pre-processed and represented as semantic and systematic relation with the words.as the length of each lot summaries is different, using data embedding and skip-gram model each input is presented in a vector representation having a dimension of 300.

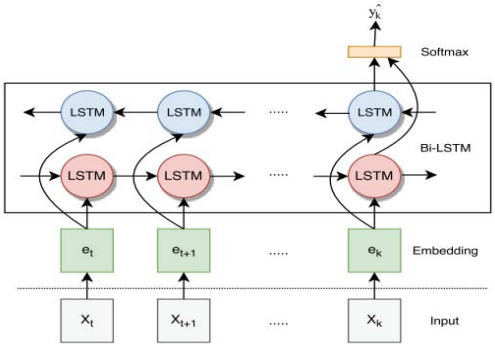


Fig. 3. Bi-Directional LSTM model

BI-LSTM is basically an RNN model which improves and learns complex interdependencies with passage of time. It basically has two neural networks. These are: the forward neural layer which handles and models the previous/preceding text and the other is backward neural network which models the upcoming/following context of plot summaries. These layers are shown in this fig.

Another interesting technique used to classify movie genre based on text summaries is proposed by Quan Hoang et al [22]. In this paper, in contrast to previous techniques discussed, pure, generic machine learning models were implemented as: Naive Bayes, Word2Vec+XGBoost and Recurrent Neural Networks (RNN).This paper discussed these techniques with taking dataset of around 250,000 movie videos and eve achieved f-score of 0.5 with accuracy of around 80 percent.

B. Deep Neural Networks (DNN)

In this section we discuss some work done in classifying movie videos based on genres through Poster images. Various people [10][23][14]have taken up this approach and implemented different models to improve accuracy and efficiency. Classification of genres with Deep neural network is state -of-the-art work done in this field by Wei-Ta Chu [23] et al.

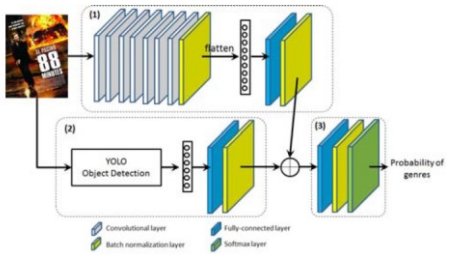


Fig. 4. DNN model using YOLO algorithm

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:24:51 UTC from IEEE Xplore. Restrictions apply.

In this paper authors discuss classification based on DNN model along with object identifier algorithm YOLO[24] as depicted in Fig 3. An overview may explain that large dataset was taken into account along with the metadata of movie posters from IMDB website while crawling across western movies from the era of '80s to 2015. A fixed resolution ranging from 89 × 132 to 300 × 581 was taken into account for poster images. DNN model makes a training set of all the images as D = {A,B }, where A = (a1, a2, ..., aN) is the set of n poster images and B0 = (b1 ,b2 , ...,bN) is the corresponding genre information. The vector ci = (ci'1, ci'2, ..., ci'M) is a binary vector, where ci,j = 1 shows what genre is connected with which of its corresponding poster image.

One problem that is generally encountered in classifying genre through poster images is that a single image may portray more than one genre, which is also called the multi label image genre classification problem.

Apart from DNN Samuel Sung et al. [2] worked on movie genre classification with poster images using other deep learning models such as: ResNet-50, VGG-16, and DenseNet-169. DenseNet-169.and in conclusion DenseNet-169 was the best performing model as it achieved the highest F1 score amongst others.

C. Convolutional Neural Networks (CNN)

Convolutional Neural Network is a very unique and special tool used for picture classification which can be also extended to movie frames classification and is widely used by various tech giants such as Facebook and Amazon for segregating images for their clients.

In CNN the main task is to accept an image as an image and create or define its class. A multidimensional vector is created every time an image is taken as an argument by the neural layer. For instance if a resolution of an image is 250X 250, then the dimension of the vector created will be of 250X250X3 (3 as RGB values).

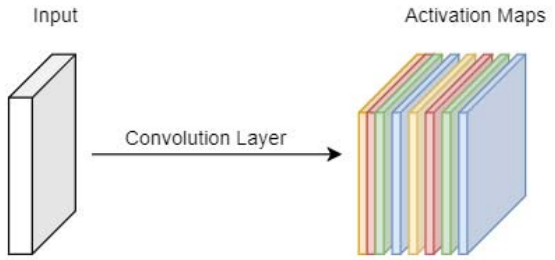


Fig. 5. 3D CNN model

Based on CNN various papers have been already published which attempts to classify movie genres. First paper we discuss here is by Gabriel S. Simoes et al [25]. Borrowing ideas from image processing concepts, this paper took a dataset of around 3500 known movie videos. Subsequently in the latter half of the paper the CNN model produced results are compared with feature extraction models such as GIST CENTRIST and W CENTRIST.

These feature extraction models which are used in various papers are summed up as follows:

- GIST: The GIST model gives a singular, comprehensive feature descriptor for the input image that encodes semantic data illustrating features of the image scene. The semantic attributes of interest illustrates high level properties of the scene, like its roughness, ruggedness, and openness.

- CENTRIST: The CENsus TRansform hISTogram, is a visual feature extracted created for recognition of the semantic category of original scenes and indoor atmosphere, e.g. coasts, forests, bedrooms, streets, living rooms, etc. They demonstrated that CENTRIST creates outstanding results.

- W-CENTRIST: Both GIST and CENTRIST models deny hues indicator on creating descriptors. Therefore, W-CENTRIST is a model of the CENTRIST feature extract which collects colour data in addition to intensity stats.

Furthermore, In [15] a file of around 144 new trailers which were not the part of dataset used for training the data was tested on the same model of scene categorization. With, original parameter settings model was tested and an accuracy of around 72% was measured which more than the accuracy was obtained on the training set.

In [25] the authors also introduced a novel method In this paper supported by a CNN architecture to classify movie trailers according to their genres, namely CNN-Motion (Convolutional Neural Networks for Movie Trailer Classification). Finally the accuracy with this model comes out to be around 74% with this mode.

IV. DATASET

A variety of summaries, plots, posters, trailers are used in this research to depict the movie genre precisely in the distinct categories of 'romance, action, comedy, horror' etc. Dataset used in this research topics differs with the different approaches used in predicting the movie genre, may it be through the features of movie trailers, movie posters or using the descriptive movie plots which tell the do and about of the movies.

- For the movie trailers, the data has been extracted through the playlist provided in Youtube which gives us over 1000+ trailers for each category.
- For plots, the main dataset used is the IMDB Film Review Dataset which contains over 50,000 movie reviews of all the wide variety of genres.
- For posters, the dataset is either extracted from the IMDB website or using the Full MoiveLens dataset from Kaggle consisting of data from TMDB and groupLens for over 45,466 movies which have the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:24:51 UTC from IEEE Xplore. Restrictions apply.

data of their genres and posters of 500 x 750 resolutions in size which perfectly fits.

V. CONCLUSION AND FUTURE WORK

Today we live in a world where having access to numerous video based content is not only easily accessible but also abundant. This video content can vary in genres and forms. Analysing them and operating different techniques, models and algorithms helps to put large data into separate 'buckets' of genres so that it is easier for the user to segregate and chose a video based on his/her choice.

In our paper, we conducted a thorough survey of the latest papers in the field of classification of videos and analysed state of the art techniques and methods applied by different researchers and scientists to solve this problem of genre classification. We also categorized these papers into four subcategories on the basis of type of parameter a model takes up to classify for instance: Plot of the video, poster of the video, audio and video features etc. Apart from algorithms and methods, we also analyse the popular and publicly available datasets and list them in our paper.

There has been a rise in the research for new state of the art classifications in the last few years and it is only going to increase in the future. There have been lots of opportunities and challenges in the problem of making visual content easily accessible to the users considering that there is plethora of options available to consume. In the last few years a lot of research has been conducted in classification models and techniques and it will be interesting to explore the intersection of genre classification models and classification of videos based on other parameters such as adult content or false/rumor content, and also classification based on type of music used in a movie, grouping dataset based on message a video is trying to portray and many more. Basically having more parameters to classify video contents opens up lot of space to explore in future .There are also a lot of challenges that need to be tackled such as scalability on real world networks of content, along with numerous opportunities and areas to be explored while going in this direction of area of work.

References

[1] N. Fei, "Movie genre classi fi cation using TF-IDF and SVM," pp. 3–8.

[2] R. Chokshi and S. Sung, "Classification of Movie Posters to Movie Genres," p. 2020, 2020.

[3] G. Barney and K. Kaya, "Predicting Genre from Movie Posters," *Semant. Sch.*, 2019.

[4] A. Yadav and D. K. Vishwakarma, "A unified framework of deep networks for genre classification using movie trailer," *Appl. Soft Comput. J.*, vol. 96, p. 106624, 2020, doi: 10.1016/j.asoc.2020.106624.

[5] P. G. Shambharkar and M. N. Doja, "Movie trailer classification using deer hunting optimization based deep convolutional neural network in video sequences," *Multimed. Tools Appl.*, vol. 79, no. 29–30, pp. 21197–21222, 2020, doi: 10.1007/s11042-020-08922-6.

[6] V. Battu, V. Batchu, R. R. Reddy, M. K. Reddy, and R. Mamidi, "Predicting the genre and rating of a movie based on its synopsis," *Proc. 32nd Pacific Asia Conf. Lang. Inf. Comput. PACLIC 2018*, no. December, pp. 52–62, 2018.

[7] S. K. S, G. Somappa, K. S. Sivaraman, and G. Somappa, "MovieScope : Movie trailer classification using Deep Neural Networks," pp. 2–8, [Online]. Available: https://virginia.app.box.com/s/hhe7xeq96d99yuqn1nr1q9yxdhs7ey9r.

[8] Y. F. Huang and S. H. Wang, "Movie genre classification using SVM with audio and video features," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7669 LNCS, no. December, pp. 1–10, 2012, doi: 10.1007/978-3-642-35236-2_1.

[9] S. Ji, W. Xu, M. Yang, and K. Yu, "3D Convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, 2013, doi: 10.1109/TPAMI.2012.59.

[10] J. A. Wi, S. Jang, and Y. Kim, "Poster-Based Multiple Movie Genre Classification Using Inter-Channel Features," *IEEE Access*, vol. 8, pp. 66615–66624, 2020, doi: 10.1109/ACCESS.2020.2986055.

[11] P. G. Shambharkar, P. Thakur, S. Imadoddin, S. Chauhan, and M. N. Doja, "Genre Classification of Movie Trailers using 3D Convolutional Neural Networks," *Proc. Int. Conf. Intell. Comput. Control Syst. ICICCS 2020*, no. Iciccs, pp. 850–858, 2020, doi: 10.1109/ICICCS48265.2020.9121148.

[12] H. Pothina, G. Gopakumar, and A. Narayanan, "Multimedia Genre Prediction by Analysing Neurophysiological Signals," *Proc. 5th Int. Conf. Inven. Comput. Technol. ICICT 2020*, pp. 364–368, 2020, doi: 10.1109/ICICT48043.2020.9112475.

[13] A. M. Ertugrul and P. Karagoz, "Movie Genre Classification from Plot Summaries Using Bidirectional LSTM," *Proc. - 12th IEEE Int. Conf. Semant. Comput. ICSC 2018*, vol. 2018-Janua, pp. 248–251, 2018, doi: 10.1109/ICSC.2018.00043.

[14] C. H. Chou and P. C. Jen, "Image based deep learning model for movie trailer genre classification," *2019 IEEE 8th Glob. Conf. Consum. Electron. GCCE 2019*, pp. 1064–1066, 2019, doi: 10.1109/GCCE46687.2019.9015293.

[15] H. Zhou, T. Hermans, A. V. Karandikar, and J. M. Rehg, "Movie genre classification via scene categorization," *MM'10 - Proc. ACM Multimed. 2010 Int. Conf.*, pp. 747–750, 2010, doi: 10.1145/1873951.1874068.

[16] H. Wang and H. Zhang, "Movie genre preference prediction using machine learning for customer-based information," *2018 IEEE 8th Annu. Comput. Commun. Work. Conf. CCWC 2018*, vol. 2018-Janua, pp. 110–116, 2018, doi: 10.1109/CCWC.2018.8301647.

[17] K. Kundalia, Y. Patel, and M. Shah, "Multi-label Movie Genre Detection from a Movie Poster Using Knowledge Transfer Learning," *Augment. Hum. Res.*, vol. 5, no. 1, 2020, doi: 10.1007/s41133-019-0029-y.

[18] A. Pal, A. Barigidad, and A. Mustafi, "Identifying movie genre compositions using neural networks and introducing GenRec-a recommender system based on audience genre perception," *Proc. 2020 Int. Conf. Comput. Commun. Secur. ICCCS 2020*, 2020, doi: 10.1109/ICCCS49678.2020.9276893.

[19] J. Wehrmann and R. C. Barros, "Movie genre classification: A multi-label approach based on convolutions through time," *Appl. Soft Comput. J.*, vol. 61, pp. 973–982, 2017, doi: 10.1016/j.asoc.2017.08.029.

[20] "https://grouplens.org/datasets/movielens/." .

[21] "http://www.omdbapi.com/." .

[22] Q. Hoang, "Predicting Movie Genres Based on Plot Summaries," *arXiv*, pp. 1–12, 2018.

[23] W. T. Chu and H. J. Guo, "Movie genre classification based on poster images with deep neural networks," *MUSA2 2017 - Proc. Work. Multimodal Underst. Soc. Affect. Subj. Attrib. co-located with MM 2017*, pp. 39–45, 2017, doi: 10.1145/3132515.3132516.

[24] "https://appsilon.com/object-detection-yolo-algorithm/." .

[25] G. S. Simões, J. Wehrmann, R. C. Barros, and D. D. Ruiz, "Movie genre classification with Convolutional Neural Networks," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2016-Octob, pp. 259–266, 2016, doi: 10.1109/IJCNN.2016.7727207.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:24:51 UTC from IEEE Xplore. Restrictions apply.

Multi Label Toxic Comment Classification using Machine Learning Algorithms

Abhishek Aggarwal, Atul Tiwari

Abstract*: *Toxic comments are the comments found in the online forums that are rude, offensive, or unfair and usually cause many users to exit the conversation. The threat of bullying and abuse on the internet obstructs the free exchange of ideas by limiting people's opposing viewpoints. Most of the Websites fail to successfully facilitate healthy conversations, leading them to either restrict or disable user comments entirely. This paper would explore the scope of online abuse and categorize them into different labels to assess the toxicity as accurately as possible using machine learning algorithms.

Keywords*: *Accuracy, Multilabel Classification, Machine Learning Algorithms, Toxic Comments

I. INTRODUCTION

One of the best inventions of the twenty-first century is that one person can connect with another person anywhere in the world using only a smartphone and the internet, thanks to the rapid growth of computer science and technology.

People only used email to communicate with one another in the early days of the internet, and it was flooded with spam. It was difficult to distinguish between authentic and spam emails back then. However, communication and data flow across the internet have evolved dramatically over time, especially with the introduction of social networking websites like Facebook and Reddit. Hence, it is becoming critical to categorize posts as positive or negative to avoid societal damage and save individuals from engaging in antisocial behavior.

Authorities have recently made several arrests as a result of people's toxic and dangerous online posts. Last Year, the Vadodara police arrested a popular YouTube figure, Shubham Mishra, after he uploaded a video threatening the stand-up comedian Agrima Joshua to his thousands of YouTube subscribers. Moreover, in January of 2021, after allegedly inciting the Washington riots, Donald J. Trump was banned from nearly all social networking sites. As a result, there is a troubling situation, and it is critical to spot certain material before it is uploaded. And these harmful contents are making the internet a dangerous environment for users.

Manuscript received on May 07, 2021.
Revised Manuscript received on May 15, 2021.
Manuscript published on May 30, 2021.
* Correspondence Author
Abhishek Aggarwal, Bachelor of Technology in Electrical Engineering, Delhi Technological University, Delhi, India.
Atul Tiwari*, Bachelor of Technology in Electrical Engineering, Delhi Technological University, Delhi, India.

Let's say somebody makes a statement on the internet," Bullshit? Back up wanker. I'll have your account terminated". The derogatory connotation of terms like 'Bullshit' and 'Back up wanker' is clearly toxic. However, this statement will first undergo a specific process known as pre-processing and after which a classification algorithm will be used in order to extract the toxicity.

We will use various classification techniques and machine learning algorithms on the dataset to solve the toxic comment classification problem and compare them based on hamming loss, log-loss, and accuracy

II. RELATED WORK

Toxic comment classification has been extensively studied in recent years, especially in the context of social media, where researchers have used various machine learning algorithms to classify toxic comments found on social media forums into different toxic classes.

In [1], authors have used supervised learning for identifying harassment. To train a model for detecting toxic posts in chat rooms and discussion forums, authors combined local features, contextual features, and sentiment features.

In [2], Ravi used machine learning algorithms to find the toxicity of comments found on social media and achieved an accuracy of 82 percent using the WEKA machine learning toolkit.

In [3], authors used a semi-supervised approach to detect profanity-related offensive content on Twitter. They achieved a 75.1 percent TP rate with Logistic Regression and a 69.7 percent TP rate with popular keyword matching baseline. The false-positive rate was identical for both at about 3.77 percent.

In [4] authors used an automatic flame detection system that applies multilevel classification and extracts features at various conceptual levels.

In [5], the authors used SVM and Naïve Bayes classifiers for detecting abusive text messages and images on social networking websites. However, they were not able to detect offensive videos or audios on social networking websites.

In paper [6], For the classification, authors have used the neural network-based approach and the non-neural

network-based approach. They used the Naïve Bayes algorithm combined with logistic regression for the non-neural based approach. They achieved good accuracy, but the F1 score for this technique is very low. However, the neural networkbased model (RNN stacked and bidirectional) outperformed the non-neural based model in terms of F1 score and accuracy.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication
© Copyright: All rights reserved.

In paper [7], authors used Convolutional Neural Networks (CNN) over the traditional Bag of words (BoW) text

classification technique. The result showed that the use of CNN with word embedding outperformed the standard BoW text classification model, which employs SVM, KNN, NB, and LDA methods. Furthermore, CNN achieved an accuracy of more than 90 percent.

In paper [8], To protect teenagers from online harassment on YouTube, they used lexical and parser feature together to detect the toxicity in the comments section of YouTube.

Unfortunately, the prevalence of toxicity on the internet is having a negative impact on people's lives [9]. As a result, we need to find a method for detecting the toxicity of comments efficiently. In our paper, we will use techniques to break the multi-label problem into several single-label problems, allowing us to use existing single-label machine learning algorithms

III. PROPOSED METHODOLOGY

A. Type of Classification

In this paper, we will classify the given dataset (comments written by a user in an online forum) provided by Kaggle in six labels, i.e., toxic, obscene, identity hate, severe toxic, threat, or insult.

The next step is to determine if the given data (comment) belongs to one or more than one or none of the mentioned six labels. For example, the given comment can be toxic and insulting, hence falling into more than one label, but the comment can also be non-toxic and not fall into any of the six labels.

Before we begin, we should first understand the difference between multi-class and multi-label classification.

The classes in multi-class problems are mutually exclusive, meaning that each input is assigned to only one label. So, for example, your phone operating system could be android or iOS, but it cannot run both simultaneously.

However, Multi-label classification assigns each input a set of target labels, i.e., the input may be assigned to several labels at the same time.

So, we may conclude that toxic comment classification is a multi-label classification problem.

B. Exploratory Data Analysis

Exploratory data analysis is a crucial step in the data analysis process. The main aim of EDA is to gain a better understanding of the given data and to analyze their key characteristics. This is achieved by using data visualization techniques.

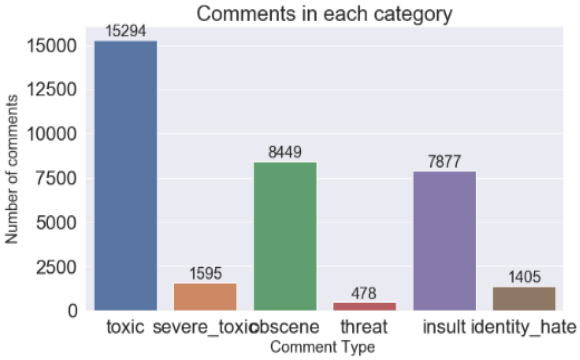


Fig. 1. Plot 1

Plot 1 depicts the number of comments that fall under each label. It can be observed that the bulk of the comments fall into the toxic category, and the threat category has the least number of comments.

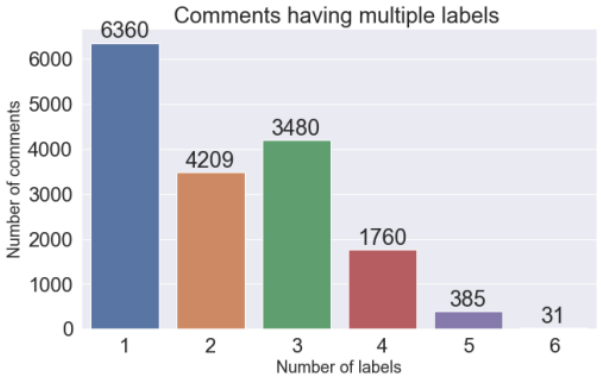


Fig. 2. Plot 2

Plot 2 shows the number of comments having multiple labels.

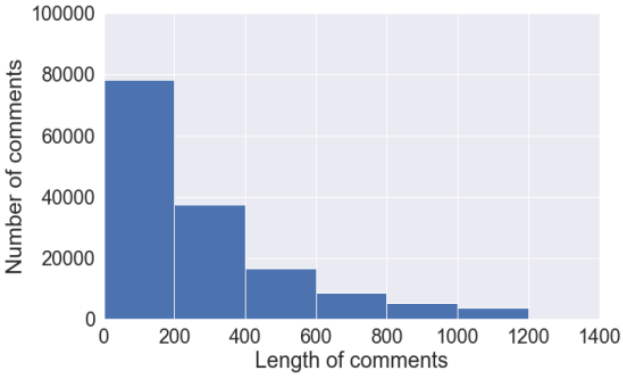


Fig. 3. Plot 3

Plot 3 depicts the number of comments of different lengths. It can be observed that the length of the comments ranged from less than 100 characters to more than 1200 characters. However, the length of most comments is less than 200 characters.

After conducting the exploratory data analysis, we have concluded that for pre-processing, we select comments with less than 400 words.

C. Data Pre-Processing

Data pre-processing is a technique used to transform the raw data into an understandable and readable format to make it suitable for building and training Machine Learning models. For our dataset, this can be achieved in 2 stages: (1) Data Cleaning (Removal of unnecessary elements from our text); (2) Feature Engineering (extracting features from data and transforming them into formats that are suitable for Machine Learning algorithms).

Retrieval Number: 100.1/ijrte.A58140510121
DOI: 10.35940/ijrte.A5814.0510121

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication
© Copyright: All rights reserved.

Steps for Data (Text) Cleaning:
• Removing Punctuations and other special/ non-ASCII characters.
• Splitting the comments into individual words.
• Removing Stop Words.
• Stemming and Lemmatising.
• Stemming and Lemmatising.

Hence, we get the comments as lists of clean tokens, and now we need to convert each of those comments into a vector through feature engineering to make them suitable for the SciKit Learn's algorithms.

The next step is to extract features using two techniques: **Count Vectorization and TF-IDF Transformation.**

	Word 1 Count	Word 2 Count	...	Word N Count
Message 1	0	1	...	0
Message 2	0	0	...	0
...	1	2	...	0
Message N	0	1	...	1

Fig. 4. Bag of words model using Count Vectorizer

Now our dataset is ready for train-test split and we can run it in any suitable Machine Learning model.

D. Finalizing Evaluation Metrics

Now that our data is pre-processed, the next step is to apply machine learning algorithms to them. But before applying any algorithms, we must first decide the proper evaluation metrics. Because machine learning algorithms' effectiveness is calculated using evaluation metrics. There are two main types of metrics for multi-label classification:

Label-based metrics: These are evaluated separately for each of the labels and then averaged for all of them without considering the labels' relationships. E.g., one-error, average precision, etc.

Example-based metrics: These are calculated for each example and then averaged across the test set. E.g., accuracy, log-loss, hamming-loss, etc.

An important observation is that our data is skewed; that is, the majority of the comments in our dataset are non-toxic. So, we cannot use accuracy as our only measure. For example, 92 percent of the comments in our dataset are non-toxic, which means even if we apply a basic machine learning algorithm that predicts the non-toxic value for all the comments would also result in 92 percent accuracy. As a result, selecting the metric that will determine the loss will be a safer alternative. Hence, we will use Hamming-Loss and Log-Loss along with accuracy as our evaluation metrics to compare the performance of various models in our machine learning algorithms.

E. Applying Multi Label Classification Techniques

The majority of conventional machine learning algorithms are designed for classification problems with single-label. Hence, we'll use techniques to break the multi-label problem into several single-label problems, allowing us to use the existing conventional machine learning algorithms.

1. **Binary Relevance Method:** The interdependence of labels is not taken into account in this process. Each label is solved separately, like a single-label classification problem.

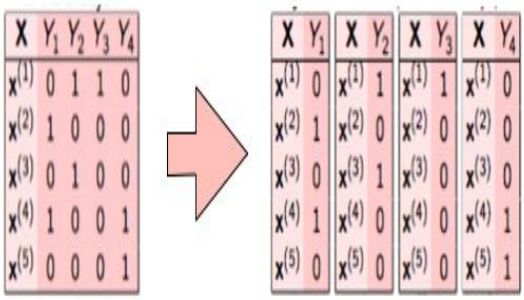


Fig. 5. Binary Relevance Method

2. **Classifier Chain Method:** We train the first classifier on the given data in this method, followed by each subsequent classifier being trained on the previous classifier and the input space, and so on. Hence, this approach considers the interdependence of labels and input data. Some classifiers may show dependence, such as toxic and severely toxic.

Fig. 6. Classifier Chain Method

3. **Label Power Set Method:** This approach takes all possible label combinations into account. As a result, any specific combination can be used as a label, breaking our multi-label problem into a multi-class classification problem.

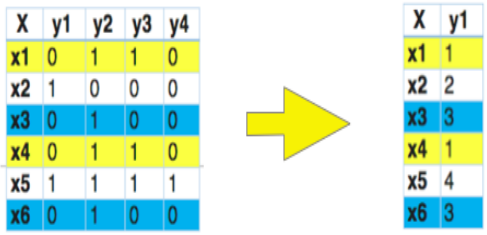


Fig. 7. Label Power Set Method

With each of these methods, we will use five machine algorithms to get optimal results.
1) Multinomial Naïve Bayes
2) Random Forest Classifier
3) Bernoulli Naïve Bayes
4) Nearest Centroid
5) Ridge Classifier

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication
© Copyright: All rights reserved.

IV. RESULT AND ANALYSIS

We used three methods, i.e., Binary Relevance, Classifier chain, and Label power set for each of the machine learning algorithms and the results for each machine learning algorithm are shown below with Log-Loss, Accuracy, and Hamming-Loss as our evaluation metrics.

	Hamming_loss	Accuracy	Log_loss
Binary Relevance	3.861695	88.290936	1.901773
Classifier Chain	3.616800	88.695013	1.374349
Label Powerset	4.167815	88.300119	0.536172

Fig. 8. Multinomial NB

	Hamming_loss	Accuracy	Log_loss
Binary Relevance	3.838736	88.235834	2.174770
Classifier Chain	3.798941	88.281752	2.014295
Label Powerset	4.317813	88.143999	0.538714

Fig. 9. Bernoulli NB

	Hamming_loss	Accuracy	Log_loss
Binary Relevance	2.398445	90.118468	1.982613
Classifier Chain	2.401506	90.146019	1.878685
Label Powerset	2.687728	89.925613	1.241680

Fig. 10. Random Forest Classifier

	Hamming_loss	Accuracy	Log_loss
Binary Relevance	3.453026	87.987878	1.336500
Classifier Chain	6.027490	87.060336	0.465341
Label Powerset	2.687728	89.925613	1.241680

Fig. 11. Nearest Centroid

	Hamming_loss	Accuracy	Log_loss
Binary Relevance	2.546913	90.118468	1.866874
Classifier Chain	2.528546	90.210304	1.536788
Label Powerset	3.093336	89.484801	1.226348

Fig. 12. Ridge Classifier

V. CONCLUSION

This paper has discussed three approaches to implement various machine learning algorithms and compared their Log-Loss, accuracy, and Hamming-Loss. After proper review, we may conclude there is no single best approach to solve the problem. Instead, each algorithm has got its own best approach for optimal results. However, if we look at the time complexity of the algorithms, Random forest is not suitable for this data set as other algorithms give almost the same results in lesser time.

In further research, we can use algorithm adaptation methods that transform the algorithms to perform multi-label classification directly. Furthermore, we can also experiment with more complex deep learning algorithms like CNN (convolutional neural network), MLP (multilayer perceptron), and RNN (Recurrent neural networks) in the near future as we believe our approach could reach the top performance when combined with deep learning models.

REFERENCES

1. Yin, Dawei, Xue, Zhenzhen, Hong, Liangjie, Davison, Brian, Edwards, April, Edwards, Lynne. (2009), "*Detection of harassment on Web 2.0*"
2. Ravi, P. (2012), "*Detecting Insults in Social Commentary*".
3. Guang Xiang, Bin Fan, Ling Wang, Jason Hong, and Carolyn Rose. 2012, "*Detecting offensive tweets via topical feature discovery over a large scale twitter corpus*". In Proceedings of the 21st ACM international conference on Information and knowledge management (CIKM '12). Association for Computing Machinery, New York, NY, USA, 1980–1984. DOI: https://doi.org/10.1145/2396761.2398556.
4. Razavi, A.H., Inkpen, D., Uritsky, S., and Matwin, S. (2010), "*Offensive Language Detection Using Multi-level Classification*". Canadian Conference on AI.
5. Kansara, Krishna B. and N. Shekokar. "*A Framework for Cyberbullying Detection in Social Network*." (2015).
6. Maxime Rivet and Mael Tran, "*Toxic comments classification*", Stanford University journal Year [2016].
7. Spiros V. Georgakopoulos et al. "*Convolutional Neural Networks for Toxic Comment Classification*", Cornell University arXiv:1802.09957 Year 2018.
8. Y. Chen and S. Zhu, "*Detecting Offensive Language in Social Media to Protect Adolescents*," [Online]. Available: http://www.cse.psu.edu/ sxz16/papers/S ocialCom2012.pdf
9. M. Duggan, "*Online harassment 2017*," Pew Res., pp. 1–85, 2017, doi: 202.419.4372

AUTHORS PROFILE

Abhishek Aggarwal, is currently pursuing his Bachelor of Technology in Electrical Engineering from Delhi Technological University. He has worked on multiple machine learning projects on synthetic datasets using python.

Atul Tiwari, is currently pursuing his Bachelor of Technology in Electrical Engineering from Delhi Technological University. He has worked on multiple machine learning projects and has built several websites which are currently live on the internet using the MERN stack.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication
© Copyright: All rights reserved.

MultiCraft

International Journal of Engineering, Science and Technology
Vol. 13, No. 1, 2021, pp. 1-9

INTERNATIONAL JOURNAL OF ENGINEERING, SCIENCE AND TECHNOLOGY

www.ijest1-ng.com
www.ajol.info/index.php/ijest
© 2021 MultiCraft Limited. All rights reserved

Multi-objective optimization of magnetic abrasive finishing by desirability function with genetic algorithm

Rajneesh Kumar Singh[1*], Swati Gangwar[2], D.K. Singh[2], Shadab Ahmad[3]

[1]Department of Mechanical Engineering, Meerut Institute of Engineering and Technology, Meerut, INDIA*
[2]Department of Mechanical Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur, INDIA
[3]Mechanical Engineering Department, Delhi Technological University, New Delhi, India
[*]Corresponding Author: e-mail: rajneeshsingh_rsme@mmmut.ac.in, Tel +91-9918985312
ORCID iDs: https://orcid.org/0000-0002-5823-0939(rajneesh), https://orcid.org/0000-0001-9792-0409(swati)

Abstract

Thermal stability and Surface hardness of the super-finished surface is a very important aspect to preserve the surface texture of workpiece in the MAF process. In this present study, the multi-objective optimization of EN-31 finished through the MAF process. “Increase in Temperature” and “Increase in Hardness” are considered for optimization to diminish their impact on the super-finished surface of EN-31. In present work Desirability function analysis (DFA) has been used to optimize the desired responses of the MAF process. Experiments were designed according to Taguchi L9 orthogonal array for the finishing of EN-31. The experiment results are processed using DFA and Desirability fitness function is established to convert the single response to multi-response. Genetic Algorithm (GA) is used to enhance the results of DFA and the regression model was developed to obtain the objective function of Genetic algorithm. Smaller-the-best criteria were used for ‘Increase in Temperature’ and ‘Increase in Hardness’ for obtaining favorable process parameters. The best optimal parametric combination is obtained by using the GA-DFA hybrid approach is at 2.5 mm (working gap), 20 gm (abrasive weight), and 2.0 A (Current) and 300 rpm (rotational speed).

Keywords: Magnetic abrasive particles, Temperature, Hardness, Genetic algorithm, Desirability function analysis.

DOI: http://dx.doi.org/10.4314/ijest.v13i1.1S

Cite this article as:
Singh R.K., Gangwar S., Singh D.K., Ahmad S. 2021. Multi-objective optimization of magnetic abrasive finishing by desirability function with genetic algorithm, *International Journal of Engineering, Science and Technology*, Vol. 13, No. 1, pp. 1-9. doi: 10.4314/ijest.v13i1.1S

Received: December 1, 2019; Accepted: February 5, 2021; Final acceptance in revised form: March 31, 2021

This paper was earlier presented at the International Conference on Energy, Environment & Material Sciences (ICE2M), 1-3 December 2019 and substantially improved for this Special Issue. Guest Editor: Dr. Sri Niwas Singh, Professor (HAG), Department of Electrical Engineering, Indian Institute of Technology Kanpur, 208016 (U.P.) India, former Vice-Chancellor, Madan Mohan Malviya University of Technology Gorakhpur (April 2017 to July 2020).

1. Introduction

The magnetic abrasive finishing process was familiarized to enhance the surface texture of advanced matters such as superalloys, ceramics and composites (Fox et al., 1994). This process applies the magnetic flux to regulate the magnetic-abrasive brush (MAB), through its material removal from the desired surface take place. MAB is the composition of magnetic particles (iron particles) and abrasive particles (SiC). Finishing is accomplished when these magnetic abrasive particles (MAPs) are put into the machining gap sandwiched between the processed surface and magnetic tool (Singh et al., 2005). The magnetic flux assembles the MAPs in the form of MAB and its stiffness depends on the magnetic flux intensity in the machining gap. Super-finishing is

accomplished by relative motion between the processed surface and MAB (Ahmad et al., 2017; Singh et al., 2004). This MAB acts as multi-point cutting head that remove the material in the form of micro-chips from the processed surface (Yamaguchi et al., 2012). During super-finishing a considerable quantity of heat is produced on machined surface due to "plastic distortion" and "frictional heat" (Mishra et al., 2014).

The produced heat affects the surface integrity of the machined surface. There have been many attempts made by researchers to predict the temperature and its effect on the target surface (Kumar and Yadav, 2009; Mulik et al., 2012). Surface hardness is other property that affects the surface-integrity of the machined surface (Hashimoto et al., 2016). In this finishing, temperature generation causes alteration of target surface hardness, and an increased hardness is beneficial for tribological application thus and a controlled operation is necessary (Levashov et al., 2011). It is impossible to avoid heat generation so in the process of minimizing the heat generation, benefits of hardness improvement have been exploited in this paper.

Bhagavatula and Komanduri (1996) explained that during super-finishing, the solid-phase reaction took place between workpiece and abrasive due to extreme pressure and temperature that is primarily produced by frictional heat during super-finishing. This phase reaction creates a chemical matter on the processed surface, altering its surface integrity. Further, Kumar and Yadav (2009) developed a finite element model to study the surface temperature effect on the processed surface. Inferences were made that temperature raised due to raise in MAB velocity & magnetic flux. Further, Singh et al. performed an experimental analysis of temperature on the FMAB-workpiece boundary and took aluminum 6060 as a workpiece for performing experiments in MAF process. Major outcome was surface temperature clearly depends on abrasive weight, voltage and working gap (Mulik et al., 2012).

To show the effect of the MAF process on the surface hardness of the processed surface, M.naif explored the influence on the surface hardness of the brass plate, finished via MAF process (M.naif et al., 2019). He performed a regression analysis to forecast the utmost substantial parameters affecting the response parameters. Hardness raised with raised in powder level and coil current and it reduced with a raise in rotational speed and working gap. Further, Ahmad et al. carry out research to examine the impact of abrasive size on the roughness of the finished surface (Ahmad et al., 2020). They employed the ANOVA to reveal the most influential process parameters to the response parameters. They also compared Taguchi and ANN-GA for best optimization for MAF process.

To examine the role of optimization in finishing process for the evaluation of their performance, various investigations were performed. The optimizations demonstrated improve in the capability of the manufacturing procedure, could very well implemented in industrial applications. Singh et al. combined the Simulated annealing with GRA to improve the surface texture of Aluminum 6060 workpiece finished via MAF. Their hybrid method had efficiently optimized the MAF performance (Kumar et al., 2020). Azhiri et al established a hybrid optimization method by merging "neuro-fuzzy" system with GRA to optimize the "surface roughness and cutting velocity" of Aluminium composite machined by WEDM. Their investigations were based on Taguchi's method and ANOVA revealed the most substantial process parameters (Bagherian Azhiri et al., 2014). Lately, Singh et al. had established a hybrid optimization technique by joining the Artificial neural network with Moth flame optimization. Their new methodology had successfully optimized the MAF process. they key findings were voltage, and the working gap should be kept at a minimum to achieve a better surface finish and hardness (Singh et al., 2019).

Ahmad et al. optimized five responses of the MAF utilizing the ANN-GA for multi-objective optimization. They used titanium to analyze the performance attributes between 5 parameters to 3 responses. Finally, their method positively optimized process parameters to enhance productivity of MAF (Ahmad et al., n.d.). Singh et al. applied DFA to optimize the responses of magnetic abrasive finishing. They used Aluminum 6060 that highly needed in the industries and importance of parameters were optimized by employing the PSO (Singh et al., 2020).

Pasandideh et al. studied that the simulation, generates requisite input data set from a simulated system with the desirability function analysis that can model the multi-response statistical data effectively, and the genetic algorithm effectively optimizes the model. This technique incorporates four strategies that di□er from one another in controlling the arbitrariness of the issue (Pasandideh and Niaki, 2006). In other studies, the authors have found a simultaneously optimizing multiple responses approach considers the ambiguity related to the fitted response surface model. The projected method has exceptionality because of its capability of considering all values intermediate to the chosen confidence interval rather than choosing only predicted values corresponding to a particular response and then defines the sturdiness measure for the conventional desirability function using the pessimistic case. A hybrid GA-DFA is developed to find the robust optima (He et al., 2012).

Table 1. Chemical composition of EN-31

Elements	C	Mn	S	P	Cr	Fe
Percentage	1.10	9.02	0.04	0.04	1.3	Rest

Literature review enlightens that so far, few multi-objective optimizations of response had been performed for the MAF process. The present study focuses on the Desirability function analysis of EN-31 finished through the MAF process and to enhance the results of DFA, a metaheuristic-based Genetic algorithm is used. In this research work, a hybrid optimization of the MAF process is developed by combining DFA with GA i.e., DFA-GA approach. This method is used to optimize the response measures of the

EN-31 surface when finished using the MAF process. 'Increase in temperature' and 'Increase in Hardness' are selected as a response for optimization and their influence has been minimizing to get the best desired performance.

The novelty of this paper is to exploit the benefits of hardness improvement in the process of minimizing heat generation. The desirability function is used to model the multi-response data, the simulation method produces the essential input data from a simulated system, and lastly, metaheuristic genetic algorithm has been implemented for multi-objective optimization of the model.

2. Experimentation of EN-31 by Magnetic abrasive finishing

The surface temperature experiments were done to acquire the 'Increase in temperature' of the target surface of EN-31. To register any alteration in hardness on the workpiece surface, pre-finishing and post-finishing hardness testing were done. The major responses that were selected for multi-objective optimization were 'Increase in Temperature' and 'Increase in Hardness' of EN-31 workpiece. The experiments were conducted on the EN-31 workpiece of dimension 100x100x10 mm^3 and the chemical composition was obtained by conducting the EDS test that shown in table 1. The selected process parameters are shown in table 2. The experimentation range of process parameters was selected on MAF set-up design considerations and constrains.

Table 2. Selected process parameters range

S. No	Process parameters		Taguchi L_9 OA			Units
			Level 1	Level 2	Level 3	
A	P1	Working gap	1.0	1.5	2.0	mm
B	P2	Abrasive weight	20	25	30	g
C	P3	Current	3	4.5	6	A
D	P4	Rotational speed	100	200	300	rpm

Initially, a Surface hardness test was done employing Vicker's test before the super-finishing of EN-31. Hardness was measured by selecting five different positions on samples before super-finishing via MAF process. 30 kgf load was employed to extract the pre-finishing hardness of EN-31. An average value was taken to determine the pre-hardness. The average Vickers hardness of 9 test samples was found in-between '131- 136' HV. After the super-finishing of EN-31, for the measurement hardness again Vickers test was performed for 9 test samples at five different locations.

Then, Surface temperature experiments of EN-31 were conducted using the MAF process shown in figure (1) to access the Increase in temperature. At first pilot experiments were performed to study the preliminary behavior of the MAF process. While performing pilot surface temperature experiments, it was seen that an increase in surface temperature had become constant approximately after 20 minutes and main surface temperature experiments were carried out for 20 minutes only. After that, Taguchi L9 orthogonal array was applied for systematic experiments. A total of nine experiments was conducted using different processing conditions shown in table 3. The experimental results showed the maximum surface temperature in EN-31 was found to be 41^o C with respect to the ambient temperature of 23 ^{0}C. Table 4 shows the 'Increase in Temperature' (ΔT) during the finishing of EN-31 for all experiments.

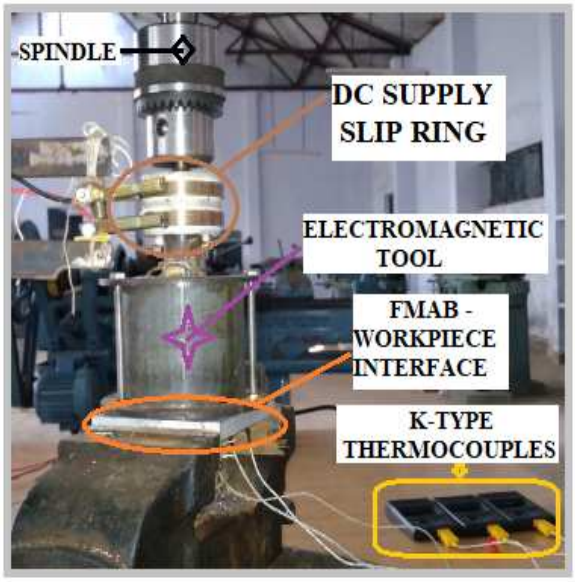


Figure. 1. Magnetic abrasive finishing experimental set-up

To examine any changes in hardness on the finished surface after conducting surface temperature experiment of EN-31, Vickers's hardness test was performed. Again, 30 kgf load was applied to find post super-finishing hardness of EN-31 at the different location. It was found that average final hardness of EN-31 is between '135.5-142.6' HV. The post super-finishing hardness was Increased by '4-7' HV approx. of all the nine test samples. Table 4 shows the 'Increase in hardness'(ΔH) of pre-finishing and post-finishing of EN-31. Experimental results concluded that, there was Increase in both temperature and hardness during the finishing of EN-31 through MAF process that shown in table 4. Maximum Increase in temperature of 18^{o} C was register for processing condition 3 and maximum alteration of hardness of 6.6 HV was achieved for processing condition 3. For examining the combined effect of 'Increase in Temperature' and 'Increase in Hardness' on finishing of EN-31 using MAF process, desirability function coupled with genetic algorithm is used for multi-objective optimization of process parameters.

Table 3. Surface temperature experimental results

TRIALS	P_1	P_2	P_3	P_4	ΔT	ΔH
1	1.0	20	3.0	100	07	4.5
2	1.0	25	4.5	200	12	5.6
3	1.0	30	6.0	300	18	6.6
4	1.5	20	4.5	300	09	5.1
5	1.5	25	6.0	100	14	5.6
6	1.5	30	3.0	200	11	5.4
7	2.0	20	6.0	200	10	5.0
8	2.0	25	3.0	300	07	4.3
9	2.0	30	4.5	100	08	6.0

3. Implementation of Desirability function

Derringer and Suich presented the DFA to translate multi-response process into a single response in the form of desirability index (d_i). They proposed three basic criteria for desirability functions. These functions were "Nominal-the-best", "Smaller-the-best", and "Larger-the-best". Where the desirability function y essentially lives amongst 0 to 1.

$$d_i = \begin{cases} 1, & \hat{y} \le y_{\min} \\ \left(\dfrac{\hat{y} - y_{\max}}{y_{\min} - y_{\max}} \right), & y_{\min} \le \hat{y} \le y_{\max}, \quad r \ge 0 \\ 0, & \hat{y} \ge y_{\max} \end{cases} \tag{1}$$

Step 1: The "single desirability index" is computed for "Increase in Temperature" and "Increase in Hardness" obtained after the experimentations of MAF process, provide in table 3. Aim is to minimalize both the chosen responses, therefore smaller-the-best desirability function is applied obtain desired characteristics. Single desirability index of "Increase in Temperature" and "Increase in Hardness" are listed in the table 4, which is computed by equation (1).

Table 4. Single desirability of experimental responses

TRIALS	ΔT	ΔH	d_G	DFF
1	1	0.913	0.955	0.511
2	0.545	0.434	0.486	0.672
3	0	0	0	1
4	0.818	0.652	0.730	0.577
5	0.363	0.434	0.397	0.715
6	0.636	0.521	0.576	0.634
7	0.727	0.695	0.711	0.584
8	1	1	1	0.5
9	0.909	0.260	0.486	0.672

Step 2: Then "composite desirability" (d_G) which a single variable is computed by merging the "single desirability index" of both the responses via following equation-

$$d_G = \sqrt[w]{\left(d_1^{w1} * d_2^{w2} \ldots\ldots * d_i^{wi} \right)} \tag{2}$$

here di is the "single desirability index", w_i is the weight of the "single desirability index" of the responses and w is the summation of the single weights. The "composite desirability value" is calculated by merging the "single desirability index" of "Increase in Temperature" and "Increase in Hardness" employing the equation (2) is listed in table 4. Weight is a vital feature of "composite

desirability" and it can be between 1 to 5, according to the significance of the responses. Here, similar weight is given to both the responses and weight of 5 is assigned to them.

Step 3: Lastly, the "desirability fitness function" (DFF) is estimated to get the optimal response. This DFF is calculated applying the following equation-

$$DFF = \frac{1}{1 + d_G} \quad (3)$$

DFF is determined using equation (3) and anticipated optimal level of the responses has selected on the DFF. Lower the DFF value, improved the surface finishing result. The minimum DFF is obtained for experiment 8 and maximum for experiment 3 as shown in table 4. DFF for experiment 8 is obtain the optimal level for finishing result, along with the optimal value based on the least DFF that are A3B2C1D4. Further, DFF are applied with meta-heuristic Genetic algorithm (GA) to acquire best parametric sets to further enhance the finishing capability of MAF process.

3.1 Regression Analysis for Desirability Fitness Function

An empirical regression model between desirability fitness function of responses with respect to the process parameters i.e., "working gap, abrasive weight, current and rotational speed" is developed. Montgomery (2000) had explained the implications and advantages of regression model and empirical model for this study is developed on MINITAB 18 statistical software. The regression results are listed in the table 5 and adequacy of DFF model is established by T-test which confirms the validity of developed DFF model. This model is competent as T-value is substantially small for all the coefficients that lie between -0.336 to 0.238. DFF regression model equation is provided below.

$$DFF = -0.049 - 0.1423\ P1 + 0.02111\ P2 + 0.07267\ P3 + 0.000298 P4$$

Analysis of variance (ANOVA) is sophisticated method in determining the significance of process parameters with respect to desirability fitness function. ANOVA has forecasted the percentage contribution of the parameters to the DFF. Table 5 shows that maximum percentage contribution is done by current which is 39.78%. followed by abrasive weight 37.30%, working gap with 16.96%.

Table 5. ANOVA forecast

Source	DF	Seq SS	Contribution	Adj SS	P-Value
Regression	4	0.173	96.99%	0.1738	0.003
P1	1	0.030	16.96%	0.0303	0.009
P2	1	0.066	37.30%	0.0668	0.002
P3	1	0.071	39.78%	0.0712	0.002
P4	1	0.005	2.96%	0.0053	0.118
Error	4	0.005	3.01%	0.0053	
Total	8	0.179	100.00%		

3.2 Implementation of Genetic Algorithm

Genetic algorithm (GA) is applied to build a metaheuristic GA-DFF forecasting model, for better forecasting of the best setting of finishing parameters. Genetic algorithm (GA) comes in class of meta-heuristic evolutionary algorithm. It is inspired by the logical revelation of natural selection. This method is used to generate impeccable solutions in field of optimization and pattern search problem rely on biological operatives such as selection, mutation and cross over. Genetic Algorithm optimization is implemented to process the results of desirability fitness function to improve the result of multi-response optimization. GA is employed using Optimization toolbox in "MATLAB". Objective function formulated applying regression equation, that is again utilized to optimize the DFF response. In order to achieve the best parametric optimal level for desirable surface finish of EN-31 meta-heuristic approach is further applied.

Table 6. Comparison between DFF and GA-DFF

TRIALS	P_1	P_2	P_3	P_4	DFF	GA-DFF	%ERROR
1	1.0	20	3.0	100	0.511	0.520	1.877
2	1.0	25	4.5	200	0.672	0.695	3.479
3	1.0	30	6.0	300	1.000	0.967	-3.260
4	1.5	20	4.5	300	0.577	0.577	-0.034
5	1.5	25	6.0	100	0.715	0.731	2.180
6	1.5	30	3.0	200	0.634	0.648	2.206
7	2.0	20	6.0	200	0.584	0.584	0.034
8	2.0	25	3.0	300	0.500	0.495	-0.960
9	2.0	30	4.5	100	0.672	0.656	-2.379

GA is employed by utilizing objective function that is based on the DFF regression equation. This objective function is to discover the global minima result of DFF. Objective function is fed in optimization toolbox. It begins with, assigning upper-lower bound of the objective function, selected from the different finishing settings listed in table 6. 500 population size and Double vector population type along with feasible populational creation function are taken. Then "rank scaling function, stochastic uniform selection function, adaptive feasible mutation function" is decided. At last heuristic crossover and forward migration are opted.

3.3 Genetic algorithm – Desirability function analysis

The GA-DFA model, simulated results with an % error amongst -3.2600 % to +3.4795%, which is a satisfactory forecast rate. GA-DFF global optimal result for all nine experiments was found, i.e., listed in table 6. GA-DFF is lowest for experiment-8 i.e., 0.4952, shown in figure (2) depicting the best optimal condition for "Increase in Temperature" and "Increase in Hardness" for finding desirable capability for finishing of EN-31 via MAF process. Maximum GA-DFF value is found at experiment-3 i.e., 0.9674, that is the poorest optimal state for finishing. Table 6 displays the error amongst the DFF and GA-DFF for nine trials.

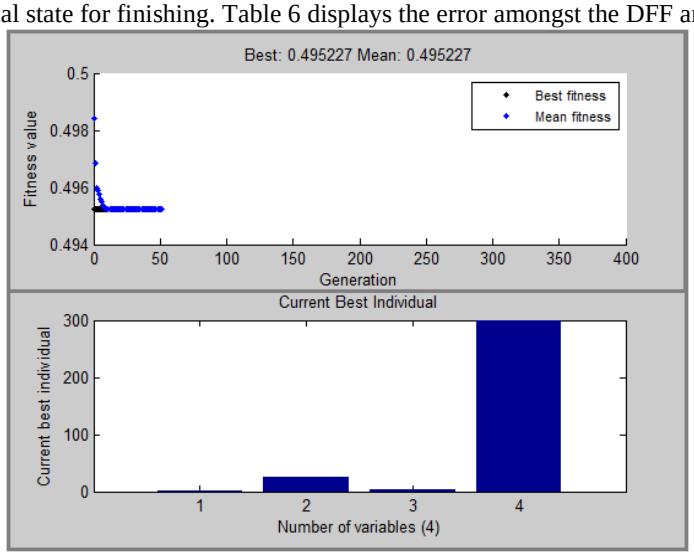


Figure. 2. Best fitness value of GA- DFA for experiment 8

3.4 GA-DFF for process parameters

The best parametric levels of Taguchi-based Desirability function (A3B2C1D4) were predicted via meta-heuristic genetic algorithm. The projected output of GA-DFA, post simulation was 0.4952. To acquire the best parameter level using GA module, one parameter was changed into different level, while other parameters were fixed. The parameter A (P1, working gap) level was changed from "1.5 mm to 2.5 mm", while others were fixed i.e., B2C1D4 level. For parameter B (P2, abrasive weight), the level was changed from "20g to 30g", while others were fixed at A3C1D4 level. Last, parameter C (P3, current) level was changed from 2A to 4A, while others were fixed at A3B1D4 level. The forecasted results for parameters A, B and C with respect to DFF are displays in Figure 3 (a-c) correspondingly.

Table 7. Comparative results

Method	**Optimal parametric setting**				**DFF**	**% IMPROVEMENT IN DFF**
	P_1	**P_2**	**P_3**	**P_4**		
Taguchi-GRA	2 .0 mm	25 g	3 A	300 rpm	0.5000	
Taguchi-GA-DFF (predicted)	2 .0 mm	25 g	3 A	300 rpm	0.4952	0.96%
Taguchi-GA-DFF (optimized)	2.5 mm	20 g	2 A	300 rpm	0.4155	16.93%

4. Validation tests

The confirmation test was done to establish the optimal setting projected by L_9 OA based GA-DFF with the experimental results. Simulation built on the forecasted parameters through Taguchi was performed. Final forecasted GA-DFF value was 0.4155. Table 7 displays the experimental results that were found by utilizing the optimal parameters forecasted by L_9OA of Taguchi-GRA model and L_9OA of Taguchi-based GA-DFF model. The forecasted DFF of parameters via L_9OA of Taguchi-based GA-DFF model has been enhanced by 16.93% which (Table 7) shows the agreeably performance of optimization model. Hence, it

is found that “Increase in Temperature” and “Increase in Hardness” can be concurrently optimized using a L_9OA of Taguchi-based GA-DFA model. Furthermore, the vigorous optimization can be implemented through the assistance of genetic algorithm model, forecasting the values beyond the defined parametric level.

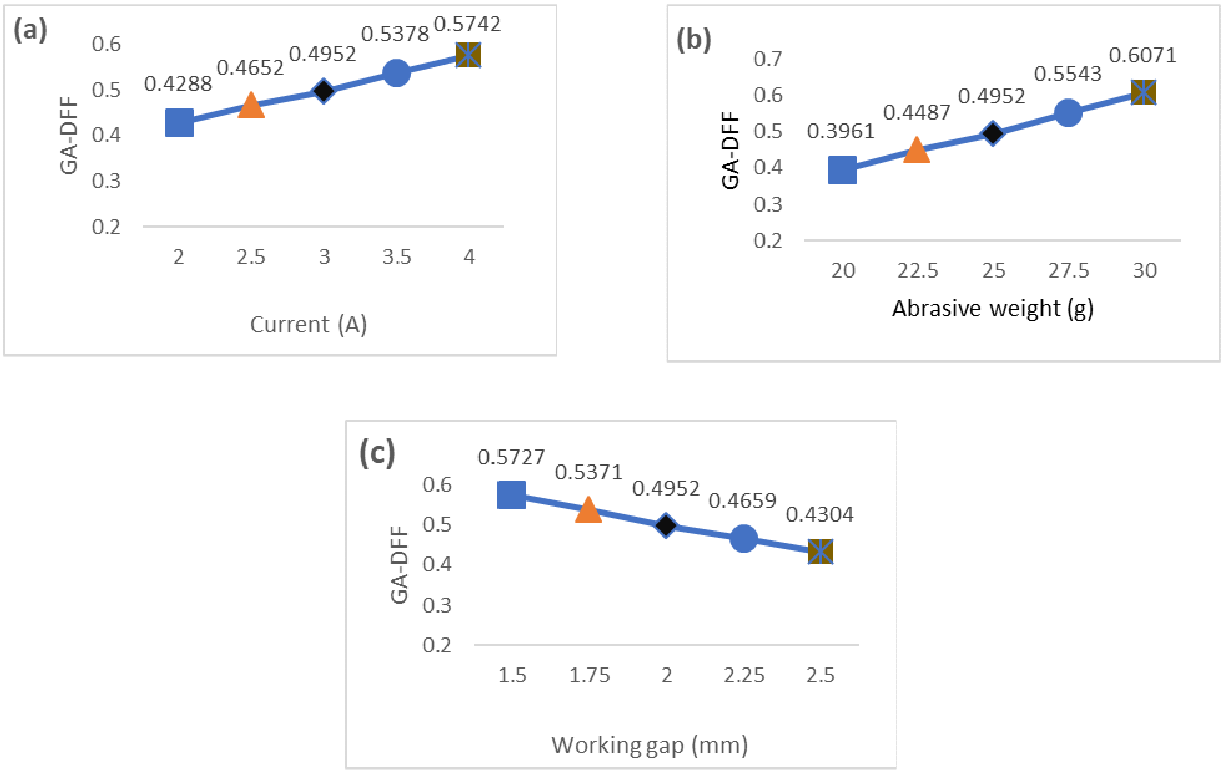


Figure.3. Simulation Results for Multiple Levels of (a) current, (b) abrasive weight, (c) working gap

5. Conclusion

GA-DFA Multi-objective optimization for “Increase in Temperature” and “Increase in Hardness” has successfully enhanced the performance of the MAF process. “Increase in Temperature” and “Increase in Hardness” are the responses that impact the surface quality of finished EN-31. Any modification in desired surface texture would propagate failures in manufactured parts with EN-31 and this failure would terminate the running machinery. To diminish the effects of “Increase in Temperature” and “Increase in Hardness” of finished surface in EN-31, DFA is done to study the combine role of responses. DFA and Genetic algorithm are combined to form a hybrid optimization approach i.e., GA-DFA. The advantage of GA-DFA technique over meta-heuristic approaches, that it can be applied for continuous quality assessment and on-line quality directive for manufacturing product. Hence following conclusions are drawn,

- DFA is applied to reduce the influence of “Increase in Temperature” and “Increase in Hardness” on the finished surface of EN-31 using MAF process. DFA is coupled with genetic algorithm to predict the best optimal level to produce the best surface quality in EN-31.
- Regression analysis is done to establish relationship between process parameters and Desirability fitness function of “Increase in Temperature” and “Increase in Hardness”. The most significant process parameters are current and abrasive weight.
- ANOVA provided the percentage contribution of each parameters in respect with DFF. “Current” contributed most with 39.78%, trailed by “abrasive weight” with 37.30% and “working gap” with 16.96%. The least contribution is done by Rotational speed with 2.96%.
- Genetic algorithm is employed to enhance the DFA results, as DFF forecasted the optimal parametric setting. At trial 8, GA forecasted lowest GA-DFF i.e., 0.4952 with A3B2C1D4 parametric setting and this sequence of parameters will produce best surface finish on EN-31.

- The simulation using meta-heuristic GA-DFA had forecasted best parametric level i.e., “working gap” (2.5mm), “abrasive weight” (20 g) and “current” (2), produced the uppermost GRG to amplify the capability of MAF process in order to finish EN-31 surface.
- Predicted DFF of parameters via L_9 OA of Taguchi-based GA-DFA is enhanced by 16.93%, that displays a acceptable performance of the above-stated optimization method.

Nomenclature

DFA	Desirability function analysis
GA	Genetic Algorithm
MAB	magnetic-abrasive brush
d_G	composite desirability
DFF	desirability fitness function
ANOVA	Analysis of variance
L_9OA	L_9Orthogonal array

Acknowledgement

The authors are very grateful to the TEQIP-II, the “Council of Scientific and Industrial Research (CSIR) - Government of India,” for financial support in the form of SRF.

References

Ahmad, S., Gangwar, S., Yadav, P.C., Singh, D.K., 2017. Optimization of process parameters affecting surface roughness in magnetic abrasive finishing process. *Materials and Manufacturing Processes*, Vol. 32, pp. 1723–1729. https://doi.org/10.1080/10426914.2017.1279307

Ahmad, S., Singari, R.M., Mishra, R.S., 2020. Modelling and optimisation of magnetic abrasive finishing process based on a non-orthogonal array with ANN-GA approach. *Transactions of the IMF*, Vol. 98, pp. 186–198. https://doi.org/10.1080/00202967.2020.1776966

Ahmad, S., Singari, R.M., Mishra, R.S., n.d. Tri-objective constrained optimization of pulsating DC sourced magnetic abrasive finishing process parameters using artificial neural network and genetic algorithm. *Materials and Manufacturing Processes*. https://doi.org/Manuscript DOI: 10.1080/10426914.2020.1866196

Bagherian Azhiri, R., Teimouri, R., Ghasemi Baboly, M., Leseman, Z., 2014. Application of Taguchi, ANFIS and grey relational analysis for studying, modeling and optimization of wire EDM process while using gaseous media. *International Journal of Advanced Manufacturing Technology*, Vol. 71, pp. 279–295. https://doi.org/10.1007/s00170-013-5467-y

Bhagavatula, S.R., Komanduri, R., 1996. On chemomechanical polishing of Si_3N_4 with Cr_2O_3. *Philosophical Magazine A*, Vol. 74, pp. 1003–1017. https://doi.org/10.1080/01418619608242173

Fox, M., Agrawal, K., Shinmura, T., Komanduri, R., 1994. Magnetic abrasive finishing of rollers. *CIRP Annals*, Vol. 43, pp. 181–184. https://doi.org/10.1016/S0007-8506(07)62191-X

Hashimoto, F., Yamaguchi, H., Krajnik, P., Wegener, K., Chaudhari, R., Hoffmeister, H.-W., Kuster, F., 2016. Abrasive fine-finishing technology. *CIRP Annals*, Vol. 65, pp. 597–620. https://doi.org/10.1016/j.cirp.2016.06.003

He, Z., Zhu, P.-F., Park, S.-H., 2012. A robust desirability function method for multi-response surface optimization considering model uncertainty. *European Journal of Operational Research*, Vol. 221, pp. 241–247. https://doi.org/10.1016/j.ejor.2012.03.009

Kumar, G., Yadav, V., 2009. Temperature distribution in the workpiece due to plane magnetic abrasive finishing using FEM. *International Journal of Advanced Manufacturing Technology*, Vol. 41, pp. 1051–1058. https://doi.org/10.1007/s00170-008-1557-7

Kumar, S., Gangwar, S., Singh, D.K., 2020. Exploration of GRA based multiobjective optimization of magnetic abrasive finishing process using simulated annealing. *FME Transactions*, Vol. 48, pp. 195–203. https://doi.org/10.5937/fmet2001195R

Levashov, E.A., Petrzhik, M.I., Tyurina, M.Ya., Kiryukhantsev-Korneev, F.V., Tsygankov, P.A., Rogachev, A.S., 2011. Multilayer nanostructured heat-generating coatings. Preparation and certification of mechanical and tribological properties. *Metallurgist*, Vol. 54, pp. 623–634. https://doi.org/10.1007/s11015-011-9350-5

Mishra, V., Goel, H., Mulik, R.S., Pandey, P.M., 2014. Determining work-brush interface temperature in magnetic abrasive finishing process. *Journal of Manufacturing Processes*, Vol. 16, pp. 248–256. https://doi.org/10.1016/j.jmapro.2013.10.004

Mulik, R.S., Srivastava, V., Pandey, P.M., 2012. Experimental investigations and modeling of temperature in the work-brush interface during ultrasonic assisted magnetic abrasive finishing process. *Materials and Manufacturing Processes*, Vol. 27, pp. 1–9. https://doi.org/10.1080/10426914.2010.515647

Nazar, M. 2012. Study on the parameter optimization in magnetic abrasive polishing for brass Cuzn33 plate using taguchi method. *Iraqi Journal for Mechanical and Materials Engineering*, Vol. 12, pp. 596–615.

Pasandideh, S.H.R., Niaki, S.T.A., 2006. Multi-response simulation optimization using genetic algorithm within desirability function framework. *Applied Mathematics and Computation*, Vol..175, pp. 366–382. https://doi.org/10.1016/j.amc.2005.07.023

Singh, D.K., Jain, V.K., Raghuram, V., 2004. Parametric study of magnetic abrasive finishing process. *Journal of Materials Processing Technology*, Vol. 149, pp. 22–29. https://doi.org/10.1016/j.jmatprotec.2003.10.030

Singh, D.K., Jain, V.K., Raghuram, V., Komanduri, R., 2005. Analysis of surface texture generated by a flexible magnetic abrasive brush. *Wear*, Vol. 259, pp. 1254–1261. https://doi.org/10.1016/j.wear.2005.02.030

Singh, R.K., Gangwar, S., Singh, D.K., 2020. Exploration of desirability function analysis-based particle swarm optimization for magnetic abrasive finishing of mild steel. *Journal of Advanced Manufacturing Systems*, Vol. 19, pp. 607–627. https://doi.org/10.1142/S0219686720500298

Singh, R.K., Gangwar, S., Singh, D.K., Pathak, V.K., 2019. A novel hybridization of artificial neural network and moth-flame optimization (ANN–MFO) for multi-objective optimization in magnetic abrasive finishing of aluminium 6060. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, Vol. 41, 270. https://doi.org/10.1007/s40430-019-1778-8

Yamaguchi, H., Srivastava, A.K., Tan, M.A., Riveros, R.E., Hashimoto, F., 2012. Magnetic abrasive finishing of cutting tools for machining of titanium alloys. *CIRP Annals*, Vol. 61, pp. 311–314. https://doi.org/10.1016/j.cirp.2012.03.066

Biographical notes

Dr. Rajneesh Kumar Singh is working as an Assistant professor in Mechanical Engineering Department at Meerut Institute of Engineering and Technology, Meerut, Uttar Pradesh (India) since July 2020. He received his Ph.D. (Advanced Manufacturing Processes) from Madan Mohan Malaviya University of Technology, Gorakhpur, Uttar Pradesh (India). After graduation in Mechanical Engineering from Shri Ramswaroop College of Engineering and Management, Lucknow, Uttar Pradesh INDIA, he joined MMMEC Gorakhpur, Uttar Pradesh for her Post graduation in Mechanical Engineering with specialization of CIM. He has published more than 10 papers in peer reviewed Journals and International/ national conferences. He has authored a book chapter in Wiley Scrivener Journal. His area of interest is, Manufacturing Science, Advance Manufacturing, Computer Aided Manufacturing, Finite Element Method and Optimization.

Dr. Swati Gangwar is working as an Assistant Professor in Department of Mechanical Engineering at Madan Mohan Malaviya University of Technology, Gorakhpur, Uttar Pradesh (India) since January 2015. After graduation in Mechanical Engineering from Shri Ram Murti Smarak College of Engineering and Technology, Bareilly, Uttar Pradesh INDIA, she joined NIT Hamirpur, Himachal Pradesh for her Post graduation in Mechanical Engineering with specialization of CAD/CAM. After that she had obtained Degree of Philosophy of Doctorate from MNIT Jaipur, Rajasthan in 2016. She has published more than 80 papers in peer reviewed Journals and International/ national conferences within very short span of her carrier. She has authored a book chapter in Wiley Scrivener Journal and 5 book chapter in national and international conference. She has filed a Patent with patent no.1319/DEL/2015 A. Her areas of interest are Micro/Nano Composite, Bio-composites material, Tribology, Materials Characterizations, and Advance Manufacturing.

Prof D.K. Singh is working as a Professor of Mechanical Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur (01, January 2008 to till date). Ex-Director, Rajkiya Engineering College (Formerly Manyawar Kanshi Ram Engineering College of Information Technology), Azamgarh (UP), Affiliated College of Dr.A.P.J. Abdul Kalam Technical University, Lucknow (Formerly U.P. Technical University) from 20 January 2015 to 19 January 2018. Head, Department of Mechanical Engineering, Madan Mohan Malaviya University of Technology (Formerly Madan Mohan Malaviya Engineering College), Gorakhpur (28thFebruary 2013 to 19th January 2015) and 31-12-2018 to till date. Assistant Professor of Mechanical Engineering, Madan Mohan Malaviya Engineering College, Gorakhpur (01, January 2000 to 31-12-2007). Assistant Professor, Mechanical Engineering Department, University of Garyounis, Libya, North Africa (October 2006 to July 2007). Lecturer of Mechanical Engineering, Madan Mohan Malaviya Engineering College, Gorakhpur (February 1988 to December 1999). Research Scholar, Mechanical Engineering Department, Indian Institute of Technology Kanpur (July 2000 to February 2005). Master Fellow, Mechanical & Industrial Engineering Department, University of Roorkee, Roorkee, (July 1995 to February 1997). He has published more than 60 papers in peer reviewed Journals and International/ national conferences. He has authored and edited various a book chapter. He has filed a Patent on Indian Patent on "A Method for Magnetic Abrasive Finishing using a Pulsating Flexible Magnetic Abrasive Brush and a Magnetic Abrasive Finishing Device"Patent Application No. 1990/DEL/2005A filed Dated 27-07-2005 and Published on Patent Journal on 13-07-2007 International Classification: B24B 31/00. His areas of interest are Advanced Manufacturing Technology, Computer Controlled Manufacturing Processes, Machining Science, Computer Aided Process Planning (CAPP) of Manufacturing Processes, (Conventional/Unconventional), Application of Robust Design & Design of Experiments in Engineering, Quality Engineering.

Shadab Ahmad is a research fellow at Mechanical Engineering Department, Delhi Technological University, Gorakhpur, India. He received his M Tech (Engineering) from MMM University of Tech. in Gorakhpur, India in 2015. He is actively interested in research in the fields of Advanced manufacturing processes, Magnetic assisted finishing, Composite material, Electro-Magneto-Thermo-Mechanics, Finite Element Analysis, Isogeometric Analysis, Optimization, Soft computing. He is also research coordinator at Center for advanced production and industrial engineering research (CAPIER), DTU Delhi. He worked as joint treasurer and editor of book of abstract in ICAPIE-2018.

Arabian Journal for Science and Engineering
https://doi.org/10.1007/s13369-021-05924-w

RESEARCH ARTICLE-MECHANICAL ENGINEERING

Multi-objective Optimization of Thermodynamic and Economic Performances of Natural Refrigerants for Cascade Refrigeration

Kaushalendra Kumar Singh[1,2] · Rajesh Kumar[1] · Anjana Gupta[3]

Received: 5 February 2021 / Accepted: 23 June 2021
© King Fahd University of Petroleum & Minerals 2021

Abstract

In this paper, thermo-economic optimization and comparative analysis of a cascade refrigeration system configured with flash gas removal in its high-temperature cycle (HTC) and flash intercooling with indirect subcooling in lower temperature cycle (LTC) using different natural refrigerant pairs is performed. Thermo-economic optimization is carried out to maximize the exergetic efficiency and minimize the overall cost rate. The optimization model involves six design variables which include subcooling and de-superheating parameters, LTC evaporation and condensation temperatures, HTC condenser temperature and cascade temperature difference. The comparative analysis of twenty-two natural refrigerant pairs based on the results of thermodynamic and economic optimizations reveals that R717-R290 is most efficient pair and R290-R1150 is least efficient refrigerant pair thermodynamically whereas R717-R1270 is the best and R600a-R290 is the worst pair economically. Seven potential refrigerant pairs are chosen via the thermodynamic and economic optimization results and they are further compared based on their performances obtained through multi-objective optimization (maximization of exergetic efficiency and minimization of total cost rate). Multi-objective genetic algorithm is used for optimization which results in seventy non-dominated Pareto optimal solutions where the TOPSIS method is used to select a unique solution for each refrigerant pair. A comparison of refrigerant pairs using these unique solutions shows that R717-R1270 is the best refrigerant pair for the cascade system under consideration. It is also found that R717-R1270 results in 7.77% rise in COP and 5.32% reduction in overall cost when compared with NH_3–CO_2 refrigerant pair working under identical operating conditions.

Keywords Cascade refrigeration system · Flash intercooler · Natural refrigerants · Multi-objective optimization · Comparative analysis

List of symbols

CFC	Chlorofluorocarbon
COP	Coefficient of performance
F	Correction factor
GWP	Global warming potential
h	Specific enthalpy (kJ/kg)
HCFC	Hydrochlorofluorocarbon
HFC	Hydrofluorocarbon
$\dot{m}$	Mass flow rate (kg/s)
ODP	Ozone depletion potential
p	Pressure (kPa)
$\dot{Q}$	Heat flow rate (kW)
s	Specific entropy (kJ/kgK)
T	Temperature (°C)
ΔT_{lm}	Logarithmic mean temperature difference (K)
U	Overall heat transfer coefficient (kW/m^{2}K)
$\dot{V}$	Volume flow rate (m^3/s)
$\dot{W}$	Work input (kW)
Z	Capital cost ($)
$\dot{Z}$	Capital cost rate ($/year

Greek Symbols

α_{el}	Energy charge ($/kWh)
ϕ	Maintenance factor
η	Energy efficiency (%)
ψ	Exergy efficiency (%)

✉ Rajesh Kumar
dr.rajeshmits@gmail.com

Kaushalendra Kumar Singh
kshlndra03@gmail.com

Anjana Gupta
anjanagupta@dce.ac.in

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

[2] Department of Mechanical Engineering, G L Bajaj Institute of Technology and Management, Greater Noida, UP, India

[3] Department of Applied Mathematics, Delhi Technological University, Delhi, India

Published online: 07 July 2021

δ	Pressure ratio
Δ	Difference
$	US dollar

Subscripts

0	Ambient state
ca, i	Condenser air inlet
ca, e	Condenser air exit
cas	Cascade
cc	Cascade condenser
cond	Condenser
comp	Compressor
cs	Cold space
e	Exit
ea, i	Evaporator air inlet
ea, e	Evaporator air exit
el	Electrical
env	Environmental
evap	Evaporator
i	Inlet
k	*k*th component
m	Mechanical
op	Operational
s	Isentropic
sup	Superheat

1 Introduction

With the increase in demand of efficient, economical and safer cooling systems for rapidly growing frozen food and chemical industry, optimization of low-temperature refrigeration systems using environment-friendly refrigerants has nowadays become one of the most important objectives to be achieved. Due to the involvement of high-temperature lift in these applications, single-stage refrigeration systems fail to cater the requirements of these industries because of its limitations on pressure ratio and thermophysical properties of a single refrigerant [1]. These limitations have been prevailed over by cascaded and multistage refrigeration systems using synthetic refrigerants as working fluids owing to their superior thermal efficiency, durability and safety of operation [2]. But the use of these refrigerants (CFC, HCFC and HFCs) reasoned the significant detrimental environmental impacts like depletion of the stratospheric ozone layer and global warming. CFCs were found to be responsible for depleting the stratospheric ozone layer and consequently banned under Montreal Protocol (1987). Later on, HCFCs, alternatives to CFCs, were also identified as hazardous to ozone layer and possessing high global warming potential and were scheduled the phase-out by 2020–2030 under Kyoto protocol (1997). Time-bound permission and interim solution to use HFC (up to 2040), also alternatives to CFCs, was also given under Kyoto Protocol. As the deadline is approaching, the search of environment-friendly alternative refrigerants has become a challenge for the researchers working in this area. Natural refrigerants are one of the most desirable alternatives because they inherently possess zero ODP as well as negligible GWP and were the backbone of HVACR from 1800 to 1930 until the evolution of synthetic refrigerants [3]. They include ammonia, CO_2, N_2O and hydrocarbons which eventuate in nature's chemical and biological cycle without human intervention. Ammonia possessing zero ODP and GWP shows excellent miscibility with water and largely used in the food and beverage industry. CO_2 being abundantly available in the air has zero ODP and one GWP and shows higher COP than HFC-based systems in supermarket applications. The hydrocarbons like ethane and ethylene have extremely low boiling points of − 88.84 °C and − 104 °C, respectively, and hence they are used for ultra-low-temperature refrigeration applications (− 80 °C). Propane and propylene are effective alternatives of R12 and R22 synthetic refrigerants, respectively [4]. Although some natural refrigerants are doing well but still most of them do not meet the potential of synthetic refrigerants to appear as environment-friendly alternative solutions. Extensive investigations are in progress regarding theoretical as well as experimental analysis and optimization of low-temperature cascaded and multistage refrigeration systems using different natural refrigerants.

Bingming et al. [5] carried out experimental investigations on the cascade refrigeration system using CO_2–NH_3 refrigerant pair and reported that COP increases with the decrease in temperature lift and remains almost unaltered with superheating. Bhattacharya et al. [6] determined an optimum intermediate pressure for N_2O–CO_2 cascade refrigeration system to maximize COP. In another study, Lee et al. [7] carried out thermodynamic optimization to find cascade condenser temperature to maximize system COP and minimize total exergy destruction. Bhattacharya et al. [8] maximized the cooling load by determining an optimal intermediate temperature for a two-stage internally reversible cascade system. Getu and Bansal [9] focussed on the maximization of COP of a CO_2–NH_3 cascade refrigeration system by tuning a combination of evaporator temperature and HTC to LTC mass flow ratio. Mosaffa et al. [10] compared two different configurations of CO_2–NH_3 cascade refrigeration systems employed with flash tank and flash intercooler and concluded that a configuration with a flash tank in HTC and flash intercooler in LTC is better than the cascade system with flash tanks in both the cycles (HTC and LTC). Dopazo et al. [11, 12] determined CO_2 condensing temperature through practical and theoretical analyses of CO_2–NH_3 cascade system in two different studies. Yarmohammadi et al. [13] carried out experimental evaluations and multi-objective optimization to find optimum combinations

of pressure drop and heat transfer coefficient during evaporation of R404a inside corrugated tubes of the evaporator and found that the corrugated tubes enhance both the parameters simultaneously.

The main focus of optimization works done was to maximize the thermodynamic performances of systems which might cause a hike in overall system cost therefore a need for a parallel economic optimization was felt. In this regard, Rezayan and Behbahaninia [14] carried out a thermo-economic optimization with an interest of minimizing the capital and running cost of a CO_2–NH_3 cascade system concurrently. Aminyaveri et al. [15] performed a multi-objective optimization using evolutionary NSGA-II algorithm on a similar system with exergetic efficiency and total system cost as the two conflicting objective functions. Eini et al. [16] performed a three-dimensional multi-objective optimization with risk function caused by leakage of NH_3 and propane in CO_2–NH_3 and CO_2-propane refrigerant pairs, respectively, used in cascade refrigeration systems. Nasruddin et al. [17] performed multi-objective optimization on cascade refrigeration system using refrigerant C_3H_8 in HTC and mixture of C_2H_6/CO_2 in LTC to minimize total annual cost as well as the total exergy destruction of the system.

The works mentioned so far concentrate on optimization of a cascade refrigeration system using a single refrigerant pair but the possibility of other superior refrigerant pair could be revealed through the comparative analysis of different refrigerants pair. Studies on comparative analyses of refrigerants include Kilicarslan and Hosoz [18] who compared different refrigerant pairs in a cascade refrigeration system based on COP and total irreversibility and reported that R717-R23 is the best choice. Sun et al. [19] evaluated the thermodynamic performance of low GWP refrigerant pairs in the cascade system and found that R41 is an effective replacement of R23 as HTC refrigerant when paired with R404a. Sun et al. [20] carried out a comparative study on twenty-eight low GWP refrigerant pairs working in the cascade system and reported that R41-R161 is the most suitable low GWP refrigerant pair from both the energy and exergy point of views. In another study Sun et al. [21] investigated the performance of low GWP refrigerants in a three-stage cascade refrigeration system which involved a low, medium and high-temperature cycles, and they recommended to use R1150 in the low-temperature cycle, R170 in a medium temperature cycle and R717 in the high-temperature cycle. Sarkar et al. [22] presented a comparative analysis of natural refrigerants working in a cascade refrigeration system with fifty-six possible refrigerant pairs and concluded that NH_3-propylene shows maximum COP and Propylene-CO_2 results in maximum volumetric efficiency, respectively. Singh et al. [23] presented a comprehensive comparative analysis of natural refrigerants working in a cascade refrigeration system incorporated with a flash tank in its HTC and a flash intercooler with indirect subcooler in LTC. They found that R744-R290 is the best refrigerant pair from energy, exergy and economic points of views. Roy and Mandal [24] performed thermo-economic analysis and multi-objective optimization on the cascade refrigeration system using four refrigerant pairs and found that R41-R161 and R170-R161 are better than R41-R404a refrigerant pair. In another study, Roy and Mandal [25] carried out a numerical comparative study on R41-R404A and R170-R161 refrigerant pairs working in a cascade refrigeration system and reported an effective reduction in exergy destruction and compressor power consumption using R170-R161 pair compared to R41-R404A. Paradeshi et al. [26] experimentally investigated the possibility of R290 being a drop in substitute of R22 in a direct expansion solar-assisted heat pump system and reported an increase in exergy efficiency of the system with R290 compared to R22. Shovon et al. [27] performed a thermodynamic comparative analysis on a solar ejector system using six environmental benign refrigerants and found that R717 gives maximum COP as well as the cooling load. Saleh et al. [28] worked on process optimization of refrigeration system using different environment-friendly refrigerants focussing on minimization of compressor power with the help of Aspen HYSYS software and suggested useful strategies for achieving most efficient and economical system. Turgut et al. [29] performed a thermo-economic comparative analysis on a cascade vapour compression–absorption refrigeration system using four low GWP refrigerants on the basis of the results of single- and multi-objective optimizations. They found that R717 gives the lowest cost whereas R290 gives the maximum second law efficiency. Sun et al. [30] worked on thermo-economic optimization of operational parameters and different configurations of integrating a compression–absorption cascade system with Organic Rankine Cycle. They aimed at minimization of destroyed exergy and annual cost simultaneously with multi-objective optimization technique and found the optimal configuration as well as the operational parameters for the system. In another study based on multistage refrigeration system with intercooler cum subcooler, Bakeem et al. [31] investigated the optimum thermo-economic performances of eight different refrigerants and found that ammonia is the best alternative from the thermodynamic and economic point of views.

Most of the studies are based on basic refrigeration cycles like cascade vapour compression refrigeration cycle, cascade vapour absorption cycle and multistage cycles with different synthetic and natural refrigerants. However, a comprehensive comparative study involving only natural refrigerant pairs based on thermo-economic multi-objective optimization is very rare so far in the reviewed literature. That is why, the previous studies commonly resulted in recommendation of synthetic refrigerants which are going to be phased out as per their scheduled deadlines in the very interest of

environment and health. In this situation, it is imperative to investigate the feasibility of using natural refrigerants through an optimization approach to explore their best possible results which can be instrumental in searching future environment-friendly alternative refrigerants. But most of the natural refrigerants show less thermodynamic performances relative to synthetic ones; therefore, a cascade refrigeration system with a two-stage LTC configuration is chosen for the analysis to overcome this draw back up to some extent. This system is a hybrid one which combines cascade system with multistage system making it more capable than the individual systems and so, is one of the suitable choices for the analysis of natural refrigerants. Although the practical implementation of this system is still to be explored, it can be used for low-temperature applications like frozen food industry, supermarket applications, fish cold storages and chemical industries. Here thermo-economic optimization of this system is done using twenty-two natural refrigerant pairs and TOPSIS decision-making technique has been employed to select the most efficient solution from Pareto optimal front and assign the order of preference for different refrigerant pairs. This study is an endeavour towards seeking alternative environment-friendly natural refrigerants for low-temperature applications through a comprehensive comparative analysis based on thermo-economic optimization.

2 System Description

Figure 1a shows the block diagram of a cascade system equipped with a flash tank in HTC and a flash intercooler cum indirect subcooler in LTC. Air from cold space at a temperature (T_{cs}) enters the evaporator (EVAP) at a temperature ($T_{ea,i}$) through evaporator fan (EF) and after transferring heat ($\dot{Q}_{evap}$) to the LTC refrigerant, it leaves at a $T_{ea,e}$. The LTC refrigerant undergoes evaporation (8–1) and enters the flash intercooler with indirect subcooler (FIIS) at superheated state (2) after a non-isentropic compression (1–2) in LTC compressor I (LTC comp I). A de-superheated stream (3) leaves the FIIS and enters the cascade condenser

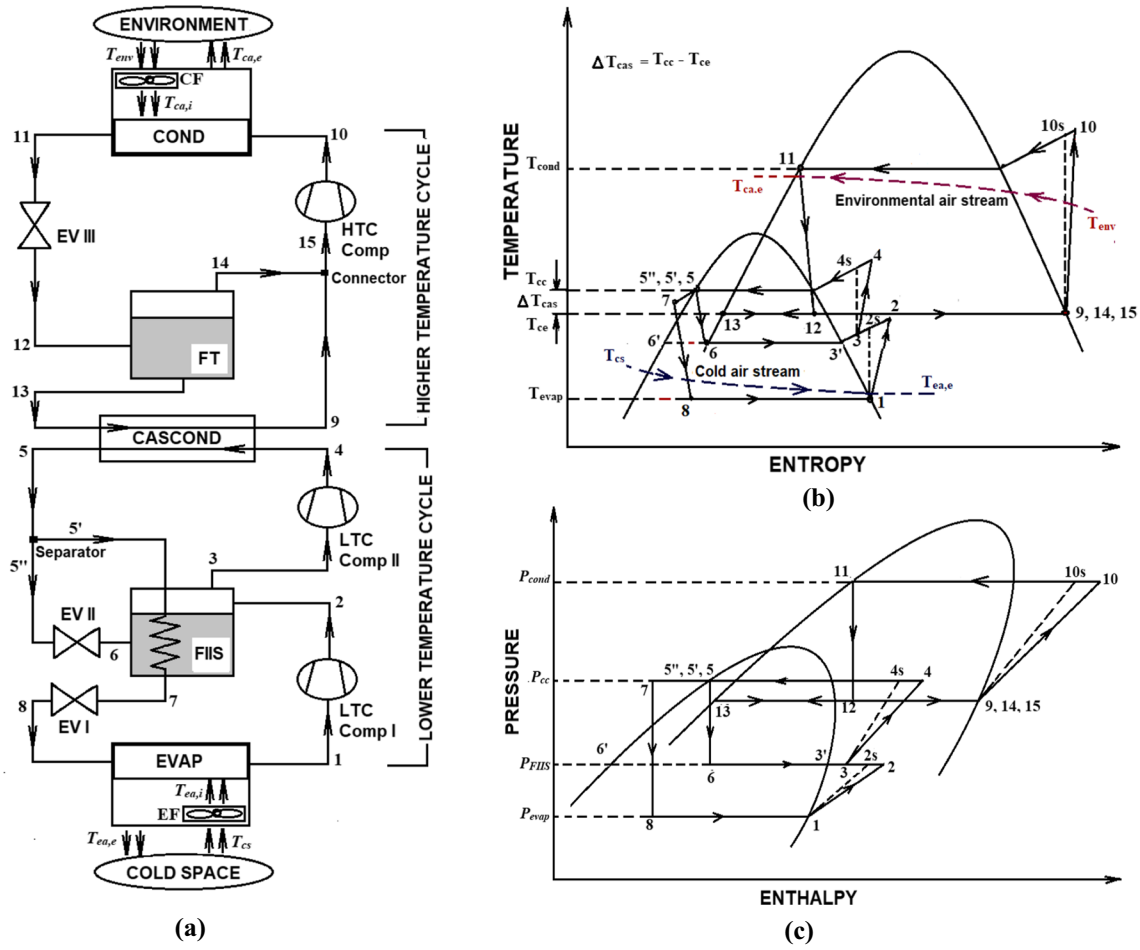


Fig. 1 **a** Schematic diagram of cascade system equipped with flash tank (in HTC) and flash intercooler cum subcooler (in LTC), **b** T-s diagram of cycles and **c** P–h diagram of cycles

(CASCOND) after second-stage compression (3–4) in LTC compressor II (LTC comp II). In CASCOND, the LTC refrigerant undergoes condensation (4–5) to a saturated vapour state (5) by transferring heat to the HTC refrigerant and then divided into two streams with the help of the separator. One stream (5″) serves as a makeup supply to the FIIS after getting expanded in expansion valve II (EV II) whereas another one passes through the subcooler tube of FIIS and gets subcooled to a state (7) and fed to the EVAP for extracting heat of the cold space, after going through an expansion (7–8) in expansion valve I (EV I). The HTC refrigerant receives heat from LTC refrigerant in CASCOND and undergoes evaporation (13–9) to saturated vapour state (9). This saturated vapour stream joins the saturated vapour stream (14) leaving the flash tank (FT) and enters the HTC compressor (HTC comp) for non-isentropic compression (15–10) to a superheated vapour state (10) and then gets condensed (10–11) to the saturated liquid state (11) in the condenser (COND) by transferring heat to the environmental air stream (T_{env}) which enters the condenser at a temperature $T_{ca,i}$ with the help of condenser fan (CF) and leaves it at $T_{ca,e}$. The saturated liquid stream exiting the condenser is expanded (11–12) in expansion valve III (EV III) and enters the flash tank (FT) which outputs two streams. One is the flash gas (14) which is mixed with stream (9) and another one is the saturated liquid refrigerant stream (13) which enters the CASCOND for extraction of heat rejected by LTC refrigerant. Figure 1b, c shows the T-s and p–h diagrams of both LTC and HTC.

3 Mathematical Modelling

For carrying out the thermo-economic optimization using different natural refrigerant pairs, thermodynamic and economic modelling of the system is done on EES software [32] through mass, energy, exergy balance equations and cost functions of different system components. The thermophysical properties of refrigerants at different required state points are calculated by built-in mathematical and thermodynamic functions provided with the software.

3.1 Energy Analysis

In energy analysis, the first law of thermodynamics is applied to the process equipment which contributes to the overall thermodynamic performance of the system using different refrigerant pairs. However, the following assumptions have been made to make the overall analysis simplified and feasible [15, 24]:

- Pressure losses, heat losses and difference in kinetic and potential energies across pipes are negligible.
- All process equipment of system run at steady-state conditions.
- Refrigerants are at their respective saturated states at all inlet and exit of heat exchangers.

Based on the above assumptions, thermodynamic balance equations (mass, energy and exergy) have been formulated for each process equipment of the system and summarized in Table 1.

Compressor work can be expressed as:

$$\dot{W}_{comp} = \frac{\dot{m}\left(h_s - h_i\right)}{\eta_s \eta_{el} \eta_m} = \frac{\dot{m}\left(h_s - h_i\right)}{\eta_{net}} \tag{1}$$

η_{net} is the net compressor efficiency which depends on pressure ratio (δ) across the compressor and can be expressed for HTC and LTC compressors as follows [10]:

$$\eta_{net,\,HTC} = \begin{cases} 0.0071\delta^5 - 0.1264\delta^4 + 0.9023\delta^3 - 3.2277\delta^2 + 5.7871\delta - 3.3429, & \delta < 4.3 \\ -0.0261\delta + 0.9069, & \delta \geq 4.3 \end{cases} \tag{2}$$

$$\eta_{net,\,LTC} = \begin{cases} -0.1234\delta^4 + 1.1251\delta^3 - 3.8902\delta^2 + 6.0433\delta - 2.8860, & \delta < 2.7 \\ -0.0237\delta^4 + 0.3051\delta^3 - 1.4740\delta^2 + 3.1348\delta - 1.7978\,, & \delta \geq 2.7 \end{cases} \tag{3}$$

Operating pressure of flash intercooler can be expressed as the geometric mean of LTC evaporator and condenser [31]:

$$p_{FIIS} = \sqrt{p_{evap} \times p_{cc}} \tag{4}$$

Torella et al. [33] introduced subcooling and de-superheating parameters for FIIS which regulate the amount of subcooling and intercooling, respectively. They are defined as:

$$\text{Subcooling parameter}, a = \frac{h_5 - h_8}{h_5 - h_{6'}} \tag{5}$$

where 'a' is the subcooling parameter, its value lies between 0 to 1 and it shows the extent to which subcooling has been done. If $a = 1$, that means that refrigerant exiting the

Table 1 Mass, energy and exergy balance equations for different system components

Process equipment	Mass balance equation	Energy balance equation	Exergy destruction
EVAP	$\dot{m}_8 = \dot{m}_1$ $\dot{m}_{ea,i} = \dot{m}_{ea,e}$	$\dot{Q}_{evap} = \dot{m}_{ea,i}(h_{ea,i} - h_{ea,e})$ $= \dot{m}_1(h_1 - h_8)$	$(\dot{E}_d)_{evap} = (\dot{E}_{ea,i} - \dot{E}_{ea,e})$ $+ (\dot{E}_8 - \dot{E}_1)$
COND	$\dot{m}_{10} = \dot{m}_{11}$ $\dot{m}_{ca,i} = \dot{m}_{ca,e}$	$\dot{Q}_{cond} = \dot{m}_{ca,i}(h_{ca,e} - h_{ca,i})$ $= \dot{m}_{10}(h_{10} - h_{11})$	$(\dot{E}_d)_{cond} = (\dot{E}_{ca,i} - \dot{E}_{ca,e})$ $+ (\dot{E}_{10} - \dot{E}_{11})$
CASCOND	$\dot{m}_4 = \dot{m}_5$ $\dot{m}_{13} = \dot{m}_9$	$\dot{m}_4(h_4 - h_5) = \dot{m}_{13}(h_9 - h_{13})$	$(\dot{E}_d)_{cas} = (\dot{E}_4 - \dot{E}_5) + (\dot{E}_{13} - \dot{E}_9)$
FIIS	$\dot{m}_2 + \dot{m}_6 = \dot{m}_3$ $\dot{m}_{5'} = \dot{m}_7$	$\dot{m}_6 h_6 + \dot{m}_{5'} h_{5'} + \dot{m}_2 h_2 = \dot{m}_3 h_3 + \dot{m}_7 h_7$	$(\dot{E}_d)_{FIIS} = (\dot{E}_2 + \dot{E}_{5'} + \dot{E}_6)$ $- (\dot{E}_3 + \dot{E}_7)$
FT	$\dot{m}_{12} = \dot{m}_{13} + \dot{m}_{14}$	$\dot{m}_{12} h_{12} = \dot{m}_{13} h_{13} + \dot{m}_{14} h_{14}$	$(\dot{E}_d)_{FT} = \dot{E}_{12} - \dot{E}_{13} - \dot{E}_{14}$
LTC Comp I	$\dot{m}_1 = \dot{m}_2$	$\dot{W}_{LTC\ comp\ I} = \dot{m}_1(h_2 - h_1)$	$(\dot{E}_d)_{LTC\ comp\ I} = (\dot{E}_1 - \dot{E}_2)$ $+ \dot{W}_{LTC\ comp\ I}$
LTC Comp II	$\dot{m}_3 = \dot{m}_4$	$\dot{W}_{LTC\ comp\ II} = \dot{m}_3(h_4 - h_3)$	$(\dot{E}_d)_{LTC\ comp\ II} = (\dot{E}_3 - \dot{E}_4)$ $+ \dot{W}_{LTC\ comp\ II}$
HTC Comp	$\dot{m}_9 = \dot{m}_{10}$	$\dot{W}_{HTC\ comp} = \dot{m}_{10}(h_{10} - h_{15})$	$(\dot{E}_d)_{HTC\ comp} = (\dot{E}_9 - \dot{E}_{10})$ $+ \dot{W}_{HTC\ comp}$
EF	$\dot{m}_{cs} = \dot{m}_{ea,i}$	$\dot{W}_{EF} = \dot{m}_{cs}(h_{ea,i} - h_{cs})$	$(\dot{E}_d)_{EF} = (\dot{E}_{cs} - \dot{E}_{ea,i})$ $+ \dot{W}_{EF}$
CF	$\dot{m}_{env} = \dot{m}_{ca,i}$	$\dot{W}_{CF} = \dot{m}_{env}(h_{ca,i} - h_{env})$	$(\dot{E}_d)_{CF} = (\dot{E}_{env} - \dot{E}_{ca,i})$ $+ \dot{W}_{CF}$
EV I	$\dot{m}_7 = \dot{m}_8$	$h_7 = h_8$	$(\dot{E}_d)_{EV\ I} = (\dot{E}x_7 - \dot{E}x_8)$
EV II	$\dot{m}_{5''} = \dot{m}_6$	$h_{5''} = h_6$	$(\dot{E}_d)_{EV\ II} = (\dot{E}_{5''} - \dot{E}_6)$
EV III	$\dot{m}_{11} = \dot{m}_{12}$	$h_{11} = h_{12}$	$(\dot{E}_d)_{EV\ III} = (\dot{E}_{11} - \dot{E}_{12})$
Separator	$\dot{m}_5 = \dot{m}_{5'} + \dot{m}_{5''}$	$\dot{m}_5 h_5 = \dot{m}_{5'} h_{5'} + \dot{m}_{5''} h_{5''}$	$(\dot{E}_d)_{Separator} = \dot{E}_5 - \dot{E}_{5'} - \dot{E}_{5''}$
Connector	$\dot{m}_{15} = \dot{m}_9 + \dot{m}_{14}$	$\dot{m}_{15} h_{15} = \dot{m}_9 h_9 + \dot{m}_{14} h_{14}$	$(\dot{E}_d)_{Connector} = \dot{E}_9 + \dot{E}_{14} - \dot{E}_{15}$

condenser has been subcooled down to the FIIS temperature (*i.e.* 100% subcooling), and if $a = 0$, it means that no subcooling has been done. 'h_5' and 'h_8' are the specific enthalpies of exit refrigerant streams of LTC condenser and EV I, respectively, whereas 'h_6' is the specific enthalpy of the refrigerant at saturated liquid state at FIIS pressure.

$$\text{De - superheating parameter}, b = \frac{h_2 - h_3}{h_2 - h_{3'}} \tag{6}$$

where 'b' is the de-superheating parameter, its value lies between 0 to 1 and it shows the extent to which intercooling has been done. If $b = 1$, that means that refrigerant exiting LTC compressor I has been intercooled down to the FIIS temperature (*i.e.* 100% intercooling), and if b = 0, it means that no intercooling has been done. 'h_2' and 'h_3' are the specific enthalpies of refrigerant streams at exit of LTC compressor I and suction of LTC compressor II, respectively, whereas 'h_3' is the specific enthalpy of the refrigerant at saturated vapour state at FIIS pressure.

The power consumption of evaporator and condenser fans are expressed as [34]:

$$\dot{W}_{EF} = 0.075(\dot{Q}_{evap}) \tag{7}$$

$$\dot{W}_{CF} = 0.027\left(\dot{Q}_{evap} + \sum_j \dot{W}_{comp}\right) \tag{8}$$

where

$$\sum_j W_{comp} = \dot{W}_{LTC\ comp\ I} + \dot{W}_{LTC\ comp\ II} + \dot{W}_{HTC\ comp} \tag{9}$$

Total work input is estimated as:

$$\dot{W}_{\text{Gross}} = \dot{W}_{\text{LTC comp I}} + \dot{W}_{\text{LTC comp II}} + \dot{W}_{\text{HTC comp}} + \dot{W}_{\text{EF}} + \dot{W}_{\text{CF}} \tag{10}$$

$$\text{COP} = \frac{\dot{Q}_{\text{evap}}}{\dot{W}_{\text{Gross}}} \tag{11}$$

3.2 Exergy Analysis

Exergy is the maximum amount of work that can be obtained from a system when it is brought from its original state to the dead state. The rate of physical exergy of a fluid stream can be written as [15]:

$$\dot{E}_j = \dot{m}_j\left[\left(h - h_0\right)_j - T_0\left(s - s_0\right)_j\right] \tag{12}$$

where the surface tension, electrical and electromagnetic effects and changes in kinetic and potential energies are assumed to be negligible. T_0 is the thermodynamic averaged temperature of the ambient environment which can be expressed as [35]:

$$T_0 = \frac{\left(T_{\text{ca,e}} - T_{\text{env}}\right)}{\ln\left(\frac{T_{\text{ca,e}}}{T_{\text{env}}}\right)} \tag{13}$$

Total input exergy to the system can be written as:

$$\dot{E}_{\text{input}} = \dot{W}_{\text{Gross}} \tag{14}$$

The exergy output of the system can be written as:

$$\dot{E}_{\text{output}} = \dot{E}_{\text{ca,e}} - \dot{E}_{\text{ca,i}} \tag{15}$$

Total exergy destroyed by the system can be calculated as:

$$\left(\dot{E}_{\text{d}}\right)_{\text{total}} = \dot{E}_{\text{input}} - \dot{E}_{\text{output}} = \sum_k \left(\dot{E}_{\text{d}}\right)_k \tag{16}$$

where $\sum_k \left(\dot{E}_d\right)_k$ is the sum of exergy destroyed by all components of the system. The system exergetic efficiency can be expressed as:

$$\psi = \frac{\dot{E}_{\text{output}}}{\dot{E}_{\text{input}}} = \frac{\dot{E}_{\text{ca,e}} - \dot{E}_{\text{ca,i}}}{\dot{W}_{\text{Gross}}} = 1 - \frac{\sum_k \left(\dot{E}_d\right)_k}{\dot{W}_{\text{Gross}}} \tag{17}$$

3.3 Economic Analysis

Economics analysis accounts for the total cost of the system which is calculated on per year basis and is termed as total cost rate ($/year). Total cost rate has two components, one is the annual running cost and another is annualized capital and maintenance cost of the system. The economic modelling of the system is done using the following cost balance equations:

Total system cost rate can be estimated as:

$$\dot{W}_{\text{total}} = \dot{Z}_{\text{op}} + \sum_k \dot{Z}_k \tag{18}$$

The system operating cost per year ($/year) may be calculated as:

$$\dot{Z}_{\text{op}} = N \times \dot{W}_{\text{Gross}} \times \alpha_{\text{el}} \tag{19}$$

The rate of capital and maintenance cost ($/year) of kth component is estimated by [36]:

$$\dot{Z}_k = Z_k \times \emptyset \times \text{CRF} \tag{20}$$

Cost functions used for the estimation of capital cost (Z_k) of different system components are written in Table 2.

The capital recovery factor (CRF) is defined as [37]:

$$\text{CRF} = \frac{i(1+i)^n}{(1+i)^n - 1} \tag{21}$$

where i represents annual interest rate and n is a system lifetime in years.

where $\dot{W}$ is the electric power consumption of compressors, $\dot{V}$ is the volume of air handled by EF and CF per unit time, $\dot{m}_i$ is the charging rate of FT and A is the heat exchanger surface area which can be expressed as:

$$A = \frac{\dot{Q}}{U_0 F \Delta T_{\text{lm}}} \tag{22}$$

The correction factor F for logarithmic mean temperature difference (LMTD) of heat exchangers is calculated by the relation given by Fettaka et al. [41].

3.4 Optimization Procedure

Twenty-two refrigerant pairs with four HTC refrigerant groups have been compared based on the results obtained from two single-objective optimizations including

Table 2 Cost functions used for the estimation of capital costs of the components [14, 38–40]

Component	Capital cost function (Z_k)
EVAP or COND	$1397 \times A_{\text{evap or cond}}^{0.89}$
CASCOND	$383.5 \times A_{\text{cas}}^{0.65}$
LTC comp I or II	$10167.5 \times \dot{W}_{\text{LTC comp I orII}}^{0.46}$
HTC comp	$9624.2 \times \dot{W}_{\text{HTC comp}}^{0.46}$
FT	$280.3 \times \dot{m}_{\text{i}}^{0.67}$
FIIS	$1438.1 \times (A_{\text{FIIS}})^{0.65}$
EV I or II or III	$114.5 \times \dot{m}$
EF and CF	$155 \times \left(\dot{V} + 1.43\right)$
Installation of refrigeration system	$150.2 \times \dot{Q}_{\text{evap}}$

thermodynamic optimization (maximization of exergetic efficiency) and economic optimization (minimization of total cost rate), respectively. Best LTC refrigerant has been selected for each of the four HTC refrigerants resulting in four most compatible refrigerant pairs from thermodynamic optimization. Similarly, four refrigerant pairs are selected from the results of economic optimization. Therefore a total of eight refrigerant pairs are obtained out of twenty-two pairs which are either most efficient or most economical. After taking the union of these eight members, seven most compatible refrigerant pairs are obtained. Multi-objective optimization is performed on these seven refrigerant pairs to carry out further comparative analysis and identify the best refrigerant pair.

3.4.1 Conjugate Direct Method

The thermodynamic and economic optimizations are done by the conjugate direct method using minmax tool provided with EES software. Exergetic efficiency and total cost rate are the objective functions to be maximized and minimized, respectively, and the operating conditions a, b, T_{evap}, T_{cond}, T_{cc} and ΔT_{cas} are the six design variables. In the conjugate direct method, a series of one-dimensional searches are used to find an optimal point which is a function of design variable X_1, X_2, … X_n. Thereafter, this process is repeated to confirm the optimum point. These one-dimensional searches do not essentially require numerical derivatives which is an advantage of this algorithm because it makes this algorithm capable of dealing with functions that are not easily differentiable. (Bakeem et al. [31]).

3.4.2 Multi-objective Genetic Algorithm

Multi-objective optimization is a realistic approach to optimize more than one objective functions (commonly conflicting in nature) simultaneously. Model [M] shows the formulation of a general multi-objective optimization problem mathematically [42].

$$\begin{aligned} &\text{minimize } F(\vec{x}) = \left[f_1(x), f_2(x), \ldots, f_p(x)\right]^T \\ &\text{Subject to:} \\ &\vec{g}(\vec{x}) \le 0 \qquad \text{[M]} \\ &\vec{h}(\vec{x}) = 0 \\ &\vec{x} \in R^n, \vec{f}(\vec{x}) \in R^p, \vec{g}(\vec{x}) \in R^m \text{ and } \vec{h}(\vec{x}) \in R^q \end{aligned}$$

where integers p, n, m and q are the number of objective functions, independent variables, inequality and equality constraints, respectively, and $\vec{x} \in R^n$ and $\vec{f}(\vec{x}) \in R^p$ are decision variable vector and objective function vector, respectively.

Multi-objective genetic algorithm (MOGA) tool provided with MATLAB software is used in the present work which is one of the most efficient and frequently used tools based on an artificial neural network to search the solutions of real-world multi-objective problems. Genetic algorithm is a method of optimization involving iterative search procedures in a binary search space based on an analogy with the process of natural selection (Darwinism) and evolutionary genetics. The search procedure does not result in a global solution which maximizes or minimizes all the objective functions concurrently; however, the attention is focussed on the use of Pareto optimal solutions that are not dominated by other solutions and cannot be enhanced for any objective without worsening at least one objective. The set of all non-dominated solutions is represented as Pareto optimal set, and the relevant objective function values are called the Pareto optimal front [43].

Exergetic efficiency and total cost rate are the two conflicting objective functions, whereas the operating parameters a, b, T_{evap}, T_{cond}, T_{cc} and ΔT_{cas} are the six design variables considered in the formulation of the multi-objective optimization problem. Expressions for the objective functions in terms of six design variables are developed through

Table 3 Lower and upper bounds of design variables for different refrigerant pairs

Parameter	Refrigerant pair	Lower bound	Upper bound
a	For all pairs	0	1
b	For all pairs	0	1
T_{evap} (°C)	For pairs having R744 as LTC refrigerant	− 56	− 40
	For pairs having R744a as LTC refrigerant	− 88.47	− 40
	For pairs having R290 as LTC refrigerant	− 42.39	− 40
	For pairs having R1270 as LTC refrigerant	− 48	− 40
	For pairs having R170 as LTC refrigerant	− 88.84	− 40
	For pairs having R1150 as LTC refrigerant	− 104	− 40
T_{cond} (°C)	For all pairs	35	55
T_{cc} (°C)	For all pairs	− 9	1
ΔT_{cas} (°C)	For all pairs	2	4

multiple linear regression analysis with the help of linear regression tool provided in EES software. Since all the possible constraints related to practical and theoretical issues have been taken care of while setting the variable limits and formulating the thermodynamic and economic equations, there are no equality or inequality constraints involved in formulation of multi-objective optimization problem. The upper and lower bounds of design variables along with the other system settings are mentioned in Table 3.

3.4.3 Non-dominated Shorting Genetic Algorithm-II (NSGA-II)

NSGA-II is implemented to validate and prove the accuracy of final results obtained via MOGA. It is an enhanced version of non-dominated shorting genetic algorithm (NSGA) and works on a principle that allows the elites of population to be carried to the next generation. It is a fast shorting algorithm which explicitly preserves the diversity of population through a mechanism based on crowding distance and emphasizes on non-dominance and optimality [44]. A MATLAB code for NSGA-II is run using the same objective functions obtained through multiple linear regression analysis and same boundary conditions as tabulated in Table 3. Pareto frontier obtained via NSGA-II is subjected to the TOPSIS method for getting the best optimal solution.

3.5 TOPSIS Decision making

MOGA provides a set of non-dominated solutions which is represented by Pareto optimal front. Although all the solutions are potentially optimum and desired but due to practical reasons, only one solution has to be opted for design considerations. That unique solution is selected by the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method which is based on the concept that the best solution is always farthest to the non-ideal solution (worst solution) from the single-objective optimization criterion of an objective function. In this procedure, A $u \times v$ weighted normalized decision matrix (WNDM) is formed where u is the number of solutions and v is the number of objective functions. Euclidian distances of all the solutions from ideal and non-ideal solutions are calculated which can be expressed as [16]:

Euclidian distance of a solution from an ideal solution,

$$S_i^+ = \sqrt{\sum_{j=1}^{v} \left(V_{ij} - V^{\text{ideal}}\right)^2} \tag{23}$$

Euclidian distance of a solution from non-ideal solution,

$$S_i^- = \sqrt{\sum_{j=1}^{v} \left(V_{ij} - V^{\text{non - ideal}}\right)^2} \tag{24}$$

where V_{ij} is the weighted normalized value (WNV) of jth objective function for ith solution and V^{ideal} and $V^{\text{non - ideal}}$ are the best and worst values of WNVs of objective functions, respectively. Based on S_i^+ and S_i^- a performance score of each solution is calculated which can be expressed for ith solution as:

$$C_{\text{i}} = \frac{S_{\text{i}}^-}{S_{\text{i}}^- + S_{\text{i}}^+} \tag{25}$$

In the TOPSIS method, a solution with the highest value of performance score is selected.

Table 4 Validation of thermodynamic and economic models

Parameters	Operational conditions		
	$Q_{\text{evap}} = 500$ kW, $T_{\text{evap}} = -35.2$ °C, $T_{\text{cond}} = 35.01$ °C, $T_5 = -1.98$ °C, $\Delta T_{\text{cas}} = 2.27$ °C, $\Delta T_{\text{sup}} = 0.45$ °C, $T_{\text{cs}} = -20$ °C, $T_{\text{env}} = 25$ °C, $N = 4266$ h, $\alpha = 0.09\$\ \text{kWh}^{-1}$		
	Results		
	Present model	Reference [10]	% deviation
A_{evp} (m^2)	1671	1671	0
A_{cond} (m^2)	625.3	627.6	− 0.37
A_{cas} (m^2)	61.43	59.86	+2.62
$\dot{W}_{\text{comp}}$ (kW)	264.6	267.37	− 1.04
$\dot{E}x_d$ (kW)	220.6	223.5	− 1.3
COP	1.549	1.536	+0.85
ψ (%)	31.66	31.3	+1.15
$\dot{C}_{\text{total}}$($/year)	554,353	675,530	− 17.94

Table 5 Natural refrigerant pairs (HTC-LTC)

R717 refrigerant group	R290 refrigerant group	R1270 refrigerant group	R600a refrigerant group
R717-R744	R290-R744	R1270-R744	R600a-R744
R717-R744a	R290-R744a	R1270-R744a	R600a-R744a
R717-R290	R290-R1270	R1270-R290	R600a-R290
R717-R1270	R290-R170	R1270-R170	R600a-R1270
R717-R170	R290-R1150	R1270-R1150	R600a-R170
R717-R1150			R600a-R1150

Table 6 Properties of natural refrigerants [22, 45]

Cycle	Chemical Names of Refrigerants	Refrigerant codes	Critical Pressure (bar)	Critical Temperature (°C)	Specific heat ratio	Triple Point (°C)	Latent heat (kJ/kg)	NBP (°C)
HTC	Ammonia	R717	113	132.3	1.346	− 77.7	1370	− 33.35
	Isobutane	R600a	36.4	134.7	1.127	− 159.6	366.2	− 12
HTC/ LTC	Propane	R290	42.47	96.98	1.184	− 187.1	426.1	− 42.39
	Propylene	R1270	46.67	92.42	1.211	− 185	439.5	− 48
LTC	Carbon dioxide	R744	73.77	30.98	1.3	− 56.6	349.5	− 78.4
	Nitrous oxide	R744a	72.45	36.37	1.7949	− 90.82	379.43	− 88.47
	Ethane	R170	48.72	32.12	1.296	− 182.8	489.7	− 88.84
	Ethylene	R1150	50.4	9.195	1.358	− 169.16	482.5	− 104

Table 7 Values of operational parameters used in thermodynamic modelling of the system

Parameter	Value
Cooling capacity $(\dot{Q}_{evap})$	500 kW
Environment temperature (T_{env})	25 °C
Temperature of cold air at the evaporator inlet (T_{cs})	−30 °C
Ambient pressure (p_0)	101.3 kPa
Overall heat transfer coefficient of the evaporator (U_{evap})	0.03 kW/m^{2}K^1
Overall heat transfer coefficient of the condenser (U_{cond})	0.04 kW/m^{2}K^1
Overall heat transfer coefficient of the cascade heat exchanger (U_{cas})	1 kW/m^{2}K^1
Overall heat transfer coefficient of the flash intercooler (U_{FIIS})	1 kW/m^{2}K^1
Temperature difference of the air in evaporator and condenser	10 °C

$$i_{choice} = i \in \max\left(C_i\right) \tag{26}$$

4 Results and Discussion

4.1 Model Validation

The thermodynamic and economic models are validated from the results of Mosaffa et al. [10]. Table 4 displays the values of some important system parameters obtained through the present model and the reference data at the similar operating conditions. A deviation of − 1.3% to + 2.62% in corresponding values is observed to the published data whereas the total cost rate obtained from the present model comes out to be 17.94% lower than that of the reference value. This is because the cost of penalty due to emission of carbon dioxide has not been included in total cost estimation in the present work.

4.2 Natural Refrigerant Pair formation

The system requires two different refrigerants to be used in two different cycles (*i.e.* HTC and LTC). For HTC, a refrigerant with higher NBP is suitable whereas, for LTC, a refrigerant with lower NBP is suitable. Keeping this in view, out of eight natural refrigerants twenty-two possible refrigerant pairs are formed which are arranged in four groups and each

Table 8 Results of thermodynamic optimization

Refrigerant pair		Thermodynamic optimal operating condition						Maximum exergetic efficiency
HTC	LTC	a	b	T_1 (°C))	T_{11} (°C)	T_5 (°C)	ΔT_{cas} (°C)	ψ_{max} (%)
R717	R744	1	1	− 40	36.09	− 9	4	38.63
	R744a	1	1	− 40	37.84	− 7.809	4	37.92
	R290	0.9967	1	− 40	36.09	− 9	4	40.71
	R1270	0.9963	1	− 40	36.09	− 9	4	40.61
	R170	1	1	− 40	36.09	− 9	4	38.2
	R1150	1	1	− 40	36.09	− 9	4	35.82
R290	R744	1	1	− 40	35	− 6.455	2	35.63
	R744a	1	1	− 40	35	− 4.745	2	35.78
	R1270	0.9963	1	− 40	35	− 1.534	2	37.69
	R170	1	1	− 40	35	− 4.993	2	35.31
	R1150	1	1	− 40	35	− 8.421	2	33.07
R1270	R744	1	1	− 40	35	− 6.901	2	35.63
	R744a	1	1	− 40	35	− 5.228	2	35.77
	R290	0.9964	1	− 40	35	− 2.372	2	37.72
	R170	1	1	− 40	35	− 4.994	2	35.31
	R1150	1	1	− 40	35	− 9	2	33.12
R600a	R744	1	1	− 40	35	− 9	2	36.33
	R744a	0.9997	1	− 40	35	− 9	2	36.41
	R290	0.9965	1	− 40	35	− 1.562	2	38.34
	R1270	0.996	1	− 40	35	− 1.159	2	38.27
	R170	1	1	− 40	35	− 9	2	35.94
	R1150	1	1	− 40	35	− 9	2	33.77

group is named after the HTC refrigerant being used in the refrigerant pairs falling in the group. Table 5 shows the four refrigerant groups and Table 6 depicts the basic thermophysical properties of refrigerants.

Table 7 presents the values of operational parameters and settings used to model different system components. Other important parameters regarding economic model are the maintenance factor (ϕ), the annual interest rate (i), system lifetime (n), the average cost of electricity (α_{el}) and average annual operational hours (N) which are taken as 1.06, 14%, 15 years, 0.12 \$ kWh^{-1} and 4266 h, respectively [23].

4.3 Single-Objective optimization

Here twenty-two refrigerant pairs have been studied based on two single-objective optimizations subjected to maximization of exergetic efficiency (thermodynamic optimization) and minimization of total cost rate (economic optimization), respectively, out of which seven pairs have been chosen for multi-objective optimization based on their performance obtained via thermodynamic and economic optimizations.

Table 8 presents the results of thermodynamic optimization which aims to maximize the exergetic efficiency and find a set of thermodynamic optimal values of six operational parameters. The presented results build an understanding of tuning these parameters to achieve best thermodynamic performance of cascade refrigeration system working with these pairs. It can be noticed that the subcooling (a) and de-superheating parameters (b) approach towards their maximum values (*i.e.* 1) for all the refrigerant pairs which signify that the flash intercooler is configured to deliver maximum subcooling and de-superheating to achieve maximum lower temperature cycle performance. The thermodynamic optimal evaporator temperature (T_1) is its maximum value whereas the condenser temperature (T_{11}) approaches towards its lowest possible value of provided operational range for all the pairs to minimize the temperature lift of the overall cycle which further leads to maximization of overall cycle performance. The cascade condenser temperature (T_5) and cascade temperature gap (ΔT_{cas}) attain different optimal values depending upon properties of different refrigerant pairs to maximize the exergetic efficiency which is presented in Table 8. After a comparative analysis, it can be observed that, out of six refrigerant pairs of R717 group, R717-R290 pair shows the maximum optimum exergetic efficiency of 40.71% whereas R717-1150 gives the minimum value of 35.82%. Similarly, from R290, R1270 and R600a groups R290-R1270, R1270-R290 and R600a-R290

Table 9 Results of economic optimization

Refrigerant pair		Economic optimal operating condition						Minimum total cost rate
HTC	LTC	a	b	T_1 (°C)	T_{11} (°C)	T_5 (°C)	ΔT_{cas} (°C)	$\dot{C}_{Total,\ min}$(M$/year)
R717	R744	0	1	− 53.93	38.61	− 9	2.286	0.4568
	R744a	0	1	− 54.24	38.61	− 9	2.286	0.4534
	R290	0	1	− 42.39	38.61	− 9	2.286	0.5358
	R1270	0	1	− 48	38.61	− 9	2.286	0.4529
	R170	0	1	− 54.47	38.61	− 9	2.286	0.4586
	R1150	0	1	− 54.05	38.61	− 9	2.286	0.4849
R290	R744	0	1	− 52.46	53.94	− 8.132	2	0.546
	R744a	0	1	− 53.11	53.94	− 8.132	2	0.5424
	R1270	0	1	− 48	53.94	− 8.132	2	0.5361
	R170	0	1	− 53.31	53.94	− 8.132	2	0.5485
	R1150	0	1	− 52.89	53.94	− 8.129	2	0.5795
R1270	R744	0	1	− 52.02	54.32	− 9	2	0.5326
	R744a	0	1	− 53.31	54.32	− 9	2	0.5295
	R290	0	1	− 42.39	54.32	− 9	2	0.6048
	R170	0	1	− 53.53	54.32	− 9	2	0.5353
	R1150	0	1	− 53.12	54.32	− 9	2	0.5642
R600a	R744	0	1	− 52.11	54.32	− 1.098	2	0.5653
	R744a	0	1	− 51.37	54.32	− 1.099	2	0.5588
	R290	0	1	− 42.39	54.32	− 1.099	2	0.6229
	R1270	0	1	− 48	54.32	− 1.099	2	0.5457
	R170	0	1	− 52.59	54.32	− 1.098	2	0.5677
	R1150	0	1	− 51.88	54.32	− 1.098	2	0.6188

show maximum optimum exergetic efficiencies as 37.69%, 37.72% and 38.32%, respectively, whereas R290-1150, R1270-R1150 and R600a-R1150 result in minimum optimum exergetic efficiencies of 33.07%, 33.12% and 33.77%, respectively. As a result of comparative analysis based on thermodynamic optimization it is clear that, out of twenty-two refrigerant pairs categorized in four different groups, R717-R290, R290-R1270, R1270-R290 and R600a-R290 pairs are the best candidates (one pair from each group).

Thermodynamic optimization often leads to excessive cost; therefore, an economic optimization is required to look into the economic side of the behaviour of natural refrigerant pairs. In this regard, Table 9 depicts the results of economic optimization which is subjected to the minimization of the total cost rate of the system. The values of economic optimal operational parameters and resulting minimum total cost rate for twenty refrigerant pairs have been tabulated here. It can be noticed that the subcooling parameter (a) attains zero value for each refrigerant pair to configure the flash intercooler for zero indirect subcooling through minimization of subcooler coil area which further leads to minimization of the capital cost of the flash intercooler. De-superheating parameter (b) attains maximum value for all pairs which signifies that the superheated LTC refrigerant coming out of LTC compressor I is completely intercooled to the corresponding saturated vapour state which results in minimization of gross LTC compressor work causing minimization of the overall running cost of the system. Other parameters are also tuned to different economic optimal values depending upon the thermophysical properties of different refrigerant pairs being used which can be seen in Table 9. After comparing the optimum total cost rates for six refrigerant pairs of R717 group, it can be found that R717-R1270 yields the minimum optimum cost rate of 0.4529 M$/year whereas R717-R290 shows the maximum optimum cost rate of 0.5358 M$/year. Similarly in R290, R1270 and R600a groups, R290-R1270, R1270-R744a and R600a-R1270 give minimum optimum cost rates of 0.5361 M$ year^{-1}, 0.5295 M$ year^{-1} and 0.5457 M$ year^{-1}, respectively, whereas R290-R1150, R1270-R290 and R600a-R290 pairs result in maximum optimum cost rates of 0.5795 M$ year^{-1}, 0.6048 M$ year^{-1} and 0.6229 M$ year^{-1}, respectively. Based on comparative analysis based on economic optimization, it can be drawn that, out of twenty-two refrigerant pairs categorized in four different groups, R717-R1270, R290-R1270, R1270-R744a, and R600a-R1270 are the most economical refrigerant pairs (one pair from each group).

Twenty-two refrigerant pairs were studied through two single-objective optimizations and a group-wise comparative analysis revealed that R717-R290, R290-R1270, R1270-R290 and R600a-R-290 are the best pairs from a thermodynamic point of view whereas R717-R1270,

R290-R1270, R1270-R744a and R600a-R1270 are the best pairs from an economic point of view. In these eight pairs, R290-R1270 refrigerant pair is common in both the results; therefore, there is a total of seven unique pairs which are either thermodynamically most efficient or economically most reasonable. Hence, out of twenty-two refrigerant pairs, R717-R290, R290-R1270, R1270-R290, R600a-R290, R717-R1270, R1270-R744a and R600a-R1270 may be selected as seven best refrigerant pairs from either of the criteria. Now it is imperative to perform a thermo-economic multi-objective optimization on these refrigerant pairs to investigate their potential on both thermodynamic as well as economic criteria concurrently.

4.4 Multi-objective Optimization

Thermodynamic and economic trends are generally divergent in nature; therefore, a bi-criteria optimization is required to reveal new insights to thermo-economic behaviour of these refrigerant pairs and compare them on a mediated level platform. Here, multi-objective optimization is performed on seven best refrigerant pairs selected via single-objective optimization. The functions of exergetic efficiency and total cost rate along with lower and upper bounds of design variables are fed to multi-objective genetic algorithm (MOGA) tool of MATLAB software to perform the multi-objective optimization procedure which aims to minimize the total cost rate and maximize the exergetic efficiency. MOGA ends up providing a set of seventy non-dominated optimal solutions for each refrigerant pair which are represented in the form of Pareto optimal fronts.

4.4.1 Pareto Optimal Front

Figure 2a–g shows Pareto optimal fronts for the seven refrigerant pairs. All the solutions shown in Pareto optimal fronts are non-dominated solutions and any one of them can be used to design an optimal system configuration. But the designer has to opt for any one solution which depends on the importance or weightage given to the objective functions as per his constraints. When exergetic efficiency is important, it has to be given more weightage while the selection of the final solution and when the cost is the major constraint, it has to be given more weightage than exergetic efficiency while selecting the final solution. Here, for each refrigerant pair, TOPSIS decision-making technique has been used for selection of final solutions considering five different weightages of exergetic efficiencies taken as 1, 0.75, 0.50, 0.25 and 0 which are shown in each figure by five larger bullets 'A', 'B', 'C', 'D' and 'E', respectively. Solution 'A' is called an ideal solution for exergetic efficiency because the weightage of exergetic efficiency involved in this solution is 1 (weightage of the total cost as 0) which signifies that exergetic efficiency is the utmost priority and there are no concerns about the cost while selecting the solution and so the value of exergetic efficiency in this solution is the highest. Similarly, solution 'E' is called an ideal solution for total cost rate with a weightage of exergetic efficiency as 0 (weightage of the total cost as 1) resulting in a solution with a minimum total cost even though the exergetic efficiency is compromised. Table 10 shows the ideal solutions of exergetic efficiency and total cost rate for each refrigerant pair.

It can be observed from Fig. 2a–g that as the weightage of exergetic efficiency is increased, the TOPSIS selection shifts towards a solution of higher cost rate. This shows a clear conflict between the two objective functions for each of the refrigerant pair. Figure 2a is the Pareto optimal front of R717-R290 refrigerant pair which shows that as the exergetic efficiency increases from 37.68% to 40.41% which is an increase of 7.23%, the total cost rate also increases from 0.559401 M$ year^{-1} to 0.646909 M$ year^{-1} which is a hike of 15.64%. Similarly Fig. 2b–g shows the Pareto optimal fronts for R717-R1270, R290-R1270, R1270-R290, R1270-R744a, R600a-R290 and R600a-R1270 in which a hike in total cost rate of 46.02%, 50.05%, 18.01%, 68.12%, 36.65% and 73.31% are recorded with 18.14%, 28.85%, 12.89%, 42.14%, 12.09% and 26.40% increase in exergetic efficiency, respectively, which is tabulated in Table 10.

It is the sole choice of decision-maker to take the value of weightage for different objective functions depending upon their relative importance. However, following a common practice here, equal weightage (both 0.50) has been given to both the objective functions while selecting the final solution (which is point 'C' in this case) for each refrigerant pair. Table 11 presents the optimum values of operational parameters and objective functions for each refrigerant pair. It can be seen that R717-R290 gives maximum exergetic efficiency of 38.95% with a cost of 0.566361 M$ year^{-1} and R717-R1270 gives the minimum cost rate of 0.453638 M$ year^{-1} with an exergetic efficiency of 35.84%. These two pairs show the non-dominated results, *i.e.* R717-R290 gives more exergetic efficiency but demands more cost than R717-R1270 whereas R717-R1270 is economical but gives lesser exergetic efficiency. In such a case, no clear decision can be made about which refrigerant pair is superior. A similar conflict can be observed in other refrigerant pairs also. In this situation, it is imperative to apply a decision making procedure to assign an order of preference or rank to different refrigerant pairs based on thermo-economic performance data. Therefore, TOPSIS decision-making technique is applied to the seven alternative pairs with exergetic efficiency and total cost rate being the two selection criteria with equal weights. The ranks assigned to different refrigerant pairs by TOPSIS method are indicated in Table 11. It

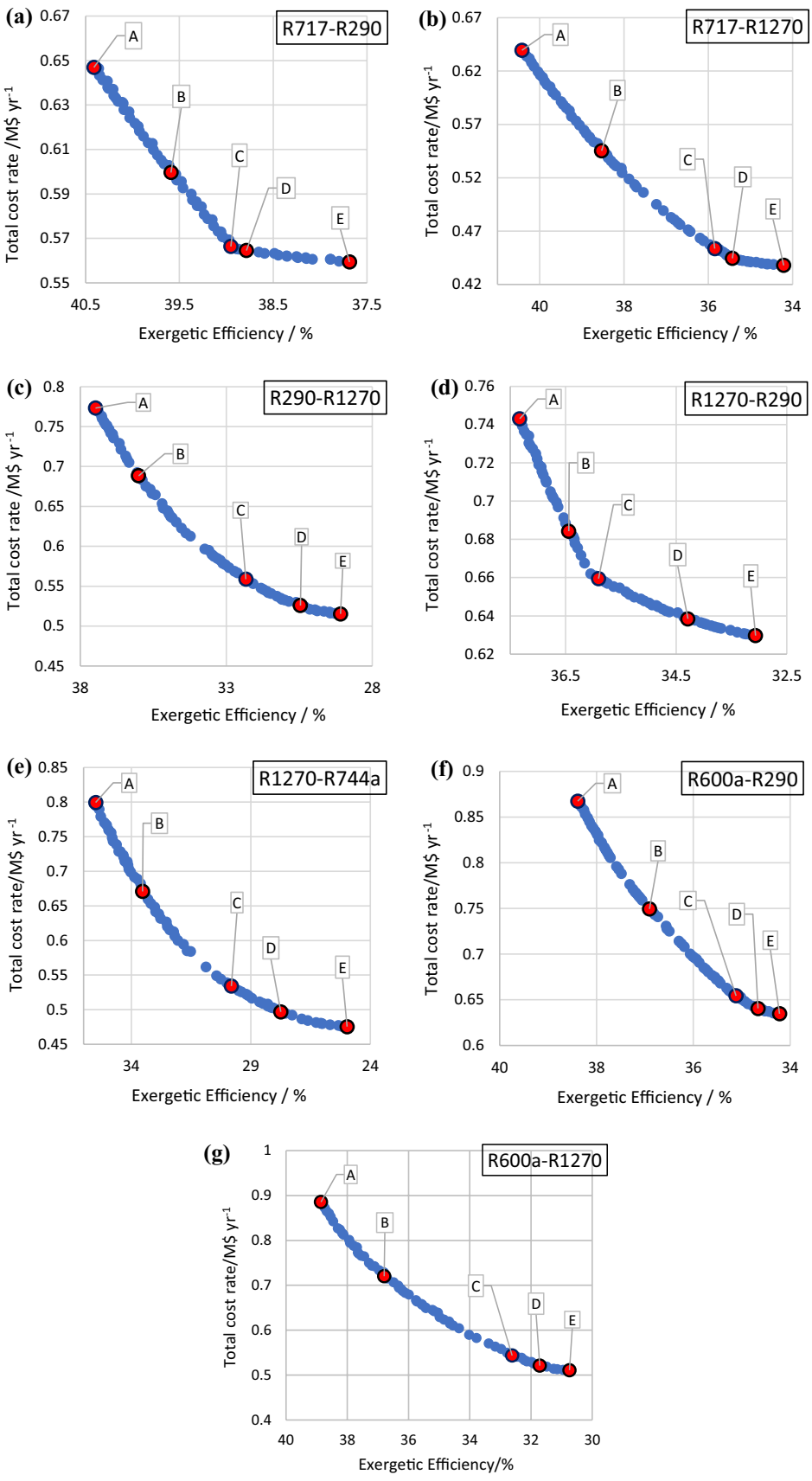


Fig. 2 Pareto optimal fronts for **a** R717-R290, **b** R717-R1270, **c** R290-R1270, **d** R1270-R290, **e** R1270-R744a, **f** R600a-R290 and **g** R600a-R1270

Table 10 Ideal solutions for exergetic efficiency and total cost rates

Refrigerant pair	Ideal solution for exergetic efficiency		Ideal solution for total cost rate		% increase in exergetic efficiency	% increase in total cost rate
	ψ (%)	$\dot{C}_{Total}$ (M$/yr)	ψ (%)	$\dot{C}_{Total}$ (M$/yr)		
R717-R290	40.4119	0.646909	37.6865	0.559401	7.23	15.64
R717-R1270	40.4208	0.63948	34.2118	0.437928	18.14	46.02
R290-R1270	37.4801	0.773281	29.0868	0.515331	28.85	50.05
R1270-R290	37.3287	0.743141	33.0643	0.629709	12.89	18.01
R1270-R744a	35.4964	0.799432	24.9715	0.47551	42.14	68.12
R600a-R290	38.3884	0.867401	34.2190	0.634739	12.09	36.65
R600a-R1270	38.8588	0.885587	30.7423	0.510982	26.40	73.31

Table 11 Order of preferences of refrigerant pairs based on TOPSIS results

Refrigerant pair (HTC-LTC)	Thermo-economic optimal operating conditions						Optimum performance parameters		Rank
	a	b	T_1 (°C)	T_{11} (°C)	T_5 (°C)	ΔT_{cas} (°C)	ψ (%)	$\dot{C}_{Total}$(M$/year)	
R717-R290	0.9999	0.9999	− 42.365	35	− 8.9993	2.0004	38.9512	0.566361	2
R290-R1270	0.9983	0.8677	− 47.7023	38.1676	− 8.8424	2.0542	32.3312	0.558809	5
R1270-R290	0.9996	0.8579	− 42.3803	35.6519	− 8.8834	2.1045	35.904	0.659494	6
R600a-R290	0.9982	0.613	− 42.3418	46.2399	− 8.6105	2.2516	35.1177	0.654438	7
R717-R1270	0.9995	0.8918	− 47.311	35.1369	− 8.9328	2.1262	35.8479	0.453638	1
R1270-R744a	0.9912	0.9833	− 50.5672	37.5787	− 8.9464	2.1571	29.8155	0.534131	4
R600a-R1270	0.9972	0.9817	− 46.4284	48.0651	− 8.8999	2.0568	32.6153	0.543655	3

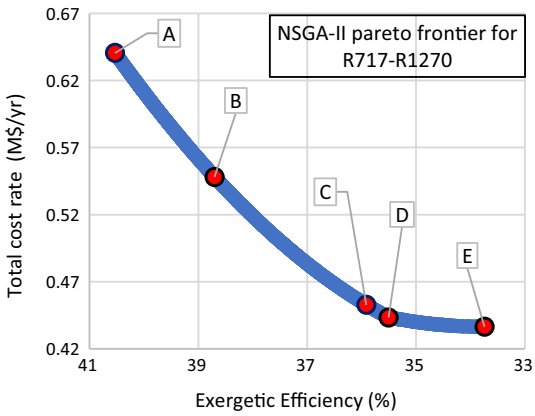


Fig. 3 Pareto frontier obtained via NSGA-II for R717-R1270 refrigerant pair

can be seen that R717-R1270 is the best refrigerant pair as it has been assigned rank 1. Since the top two ranked pairs involve R717 as HTC refrigerant, it can be drawn that R717 is the best HTC refrigerant. R600a-R290 pair should be the last choice and other refrigerant pairs may be chosen according to their ranks based on thermo-economic optimization results.

4.5 NSGA- II

Figure 3 shows the Pareto frontier for R717-R290 pair obtained via NSGA-II. The TOPSIS method is applied to identify solutions 'A', 'B', 'C', 'D' and 'F' corresponding to different weightages of exergetic efficiency as 1, 0.75, 0.50, 0.25 and 0, respectively. It can be seen that Pareto

Table 12 Comparison of NSGA-II and MOGA results for R717-R1270 refrigerant pair

TOPSIS Selections	NSGA-II		MOGA		% deviation in ψ	% deviation in $\dot{C}_{Total}$
	ψ (%)	$\dot{C}_{Total}$(M$/year)	ψ (%)	$\dot{C}_{Total}$(M$/year)		
A (Weightage of ψ = 1)	40.5323	0.640443	40.4208	0.639480	+ 0.2752	+ 0.1506
B (Weightage of ψ = 0.75)	38.7027	0.548077	38.5303	0.545328	+ 0.4452	+ 0.5041
C (Weightage of ψ = 0.5)	35.9147	0.452749	35.8478	0.453638	+ 0.1861	− 0.1959
D (Weightage of ψ = 0.25)	35.5048	0.443390	35.4290	0.444517	+ 0.2134	− 0.2536
E (Weightage of ψ = 0)	33.7419	0.436388	34.2118	0.437928	− 1.3926	− 0.3515

Table 13 Comparison of performance parameters using R717-R744 and R717-R1270 refrigerant pairs

Parameter	Reference [10] (R717-R744)	Present work (R717-R1270)	% deviation
COP	1.530	1.649	+7.77
ψ (%)	31.300	33.719	+7.72
$\dot{W}_{comp}$ (kW)	267.37	245.70	− 8.10
$\dot{C}_{total}$ ($/year)	554,523	525,022	− 5.32

frontier obtained from NSGA-II technique is broader with 1327 optimal solutions compared to multi-objective genetic Algorithm (MOGA) technique with 70 optimal solutions.

Table 12 presents comparison of the multi-objective optimization results of R717-R1270 pair obtained via MOGA and NSGA-II techniques for different exergetic efficiency weightages (*i.e.* solutions A, B, C, D and E). It can be seen that in solution 'C' optimal exergetic efficiency obtained by NSGA-II is 0.1861% greater and optimum total cost rate is 0.1959% lower than that of MOGA. So NSGA-II gives slightly better results as compared to MOGA. Similar is the case with solution 'D' where NSGA-II gives 0.2134% higher exergetic efficiency and 0.2536% lower cost. But the other solutions (A, B and E) given by MOGA are non-dominated by those of NSGA-II. Overall no significant difference in the results from both the technique is observed which validates and proves the accuracy of MOGA results.

4.6 Comparison of the Present Work with the Existing Literature

Mosaffa et al. [10] worked on the aforementioned system using NH_3-CO_2 refrigerant pair. In Table 13 some important parameters obtained by using R717-R1270 refrigerant pair have been compared with the results reported in the reference study. It is evident that COP of the system improves by 7.77% whereas the exergetic efficiency increases by 7.72%, while the compressor work and total system cost rate reduces by 8.10% and 5.32%, respectively, when R717-R744 is replaced with R717-R1270 and the system operates under similar operating conditions ($\dot{Q}_{evap}$ = 500 kW, T_{evap} = − 35.20 °C, T_{cond} = 35.01 °C, T_5 = − 1.98 °C, ΔT_{cas} = 2.27 °C, N = 4266 h, α_{el} = 0.09 $ kWh^{-1}). This implies that R717-R1270 can be an effective replacement of NH_3-CO_2 pair for the system under study.

5 Conclusion

A comparative analysis based on single- and multi-objective optimization of on a 500 kW cascade refrigeration system with flash gas removal in HTC and intercooling with indirect subcooling in its LTC using twenty-two natural refrigerant pairs has been done in this study. The study gives new insights to the behaviour of natural refrigerants in a thermodynamic and economic optimization framework which can be useful in future endeavours of selection and implementations of natural refrigerants to the cascade refrigeration systems. Following conclusions are drawn from the work presented:

- For maximization of the exergetic efficiency the subcooling parameter (a), de-superheating parameter (b) and evaporator temperature (T_{evap}) approach towards their highest possible values whereas the condenser temperature (T_{cond}) comes closer to its lowest possible value for each refrigerant pair.
- For minimization of cost, the flash intercooler is configured for maximum de-superheating and zero subcooling.
- R717-R290 is thermodynamically the best refrigerant pair with an optimum exergetic efficiency of 40.71% whereas R290-R1150 is the worst refrigerant pair with optimum exergetic efficiency of 33.07%.
- R717-R1270 is the most economical refrigerant pair with an optimum total cost rate of 0.4529 M$ year^{-1} and R600a-R290 is the worst pair with optimum cost rate of 0.6229 M$ year^{-1}.
- Comparative analysis of seven potential refrigerant pairs based on their performances obtained via multi-objective optimization reveals that R717-R1270 is the most suitable refrigerant pair from thermo-economic point of view and R717-R290 is the second best pair whereas R600a-R290 is the worst refrigerant pair.
- As both the top-performing refrigerant pairs (*i.e.*R717-R290 and R717-R1270) involve R717 as the HTC refrigerant, it can be concluded that R717 is the best HTC refrigerant among the four HTC refrigerants from multi-objective optimization criteria.
- As the weightage of exergetic efficiency is increased in TOPSIS method, it chooses a solution with higher cost rate for each refrigerant pair.
- The COP and exergetic efficiency of system improve by 7.77% and 7.72%, respectively, whereas the compressor work and total cost rate reduces by 8.10% and 5.32%, respectively, when R717-R744 is replaced by R717-R1270 refrigerant pair.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by KKS, RK and AG. The first draft of the manuscript was written by KKS and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding No funds, grants or other support was received.

Availability of Data and Material All the data generated or analysed in current study is available from corresponding author on reasonable request.

Code Availability Codes used in the current study are available from corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interest to disclose.

References

1. Park, H.; Kim, D.H.; Kim, M.S.: Thermodynamic analysis of optimal intermediate temperatures in R134a–R410A cascade refrigeration systems and its experimental verification. Appl. Therm. Eng. **54**, 319–327 (2013). https://doi.org/10.1016/j.applthermaleng.2013.01.005
2. Dubey, A.M.; Kumar, S.; Agrawal, G.D.: Thermodynamic analysis of a transcritical CO2/propylene (R744–R1270) cascade system for cooling and heating applications. Energy Convers Manage. **86**, 774–783 (2014). https://doi.org/10.1016/j.enconman.2014.05.105
3. Emani, M.S.; Mandal, B.K.: The use of natural refrigerants in refrigeration and air conditioning systems: a review. IOP Conf. Ser. Mater. Sci. Eng. **377**, 012064 (2018). https://doi.org/10.1088/1757-899X/377/1/012064
4. Abas, N.; Kalair, A.R.; Khan, N.; Haider, A.; Saleem, Z.; Saleem, M.S.: Natural and synthetic refrigerants, global warming: a review. Renew. Sustain. Energy Rev. **90**, 557–569 (2018). https://doi.org/10.1016/j.rser.2018.03.099
5. Bingming, W.; Huagen, W.; Jianfeng, L.; Ziwen, X.: Experimental investigation on the performance of NH_3/CO_2 cascade refrigeration system with twin-screw compressor. Int. J. Refrig. **32**, 1358–1365 (2009). https://doi.org/10.1016/j.ijrefrig.2009.03.008
6. Bhattacharyya, S.; Garai, A.; Sarkar, J.: Thermodynamic analysis and optimization of a novel N_2O-CO_2 cascade system for refrigeration and heating. Int. J. Refrig. **32**, 77–84 (2009). https://doi.org/10.1016/j.ijrefrig.2008.09.008
7. Lee, T.S.; Liu, C.H.; Chen, T.W.: Thermodynamic analysis of optimal condensing temperature of cascade-condenser in CO_2/NH_3 cascade refrigeration systems. Int. J. Refrig. **29**, 1100–1108 (2006). https://doi.org/10.1016/j.ijrefrig.2006.03.003
8. Bhattacharyya, S.; Bose, S.; Sarkar, J.: Exergy maximization of cascade refrigeration cycles and its numerical verification for transcritical CO_2-C_3H_8 system. Int. J. Refrig. **30**, 624–632 (2007). https://doi.org/10.1016/j.ijrefrig.2006.11.008
9. Getu, H.M.; Bansal, P.K.: Thermodynamic analysis of an R744–R717 cascade refrigeration system. Int. J. Refrig. **31**, 45–54 (2008). https://doi.org/10.1016/j.ijrefrig.2007.06.014
10. Mosaffa, A.H.; Farshi, L.G.; Infante Ferreira, C.A.; Rosen, M.A.: Exergoeconomic and environmental analyses of CO2/NH3 cascade refrigeration systems equipped with different types of flash tank intercoolers. Energy Convers. Manage. **117**, 442–453 (2016). https://doi.org/10.1016/j.enconman.2016.03.053
11. Dopazo, J.; Fernández-Seara, J.; Sieres, J.; Uhía, F.J.: Theoretical analysis of a CO_2–NH_3 cascade refrigeration system for cooling applications at low temperatures. Appl. Therm. Eng. **29**, 1577–1583 (2009). https://doi.org/10.1016/j.applthermaleng.2008.07.006
12. Dopazo, J.; Fernández-Seara, J.: Experimental evaluation of a cascade refrigeration system prototype with CO_2 and NH_3 for freezing process applications. Int. J. Refrig. **34**, 257–267 (2011). https://doi.org/10.1016/j.ijrefrig.2010.07.010
13. Yarmohammadi, S.; Mohammadzadeh, K.; Farhadi, M.; Hajimiri, H.; Modir, A.: Multi-objective optimization of thermal and flow characteristics of R-404A evaporation through corrugated tubes. J. Energy Storage. **27**, 101137 (2020). https://doi.org/10.1016/j.est.2019.101137
14. Rezayan, O.; Behbahaninia, A.: Thermoeconomic optimization and exergy analysis of CO2/NH3 cascade refrigeration systems. Energy **36**, 888–895 (2011). https://doi.org/10.1016/j.energy.2010.12.022
15. Aminyavari, M.; Naja, B.; Shirazi, A.; Rinaldi, F.: Exergetic, economic and environmental (3E) analyses, and multi-objective optimization of a CO_2/NH_3 cascade refrigeration system. Appl. Therm. Eng. **65**, 42–50 (2014). https://doi.org/10.1016/j.applthermaleng.2013.12.075
16. Eini, S.; Shahhosseini, H.; Delgarm, N.; Lee, M.; Bahadori, A.: Multi-objective optimization of a cascade refrigeration system: exergetic, economic, environmental, and inherent safety analysis. Appl. Therm. Eng. **107**, 804–817 (2016). https://doi.org/10.1016/j.applthermaleng.2016.07.013
17. Soltani, R.; Dincer, I.; Rosen, M.A.: Optimization of a cascade refrigeration system using C_3H_8 in high temperature circuit (HTC) and a mixture of C_2H_6/CO_2 in low temperature circuit (LTC). Appl. Therm. Eng. **104**, 96–103 (2016). https://doi.org/10.1016/j.applthermaleng.2016.05.059
18. Kilicarslan, A.; Hosoz, M.: Energy and irreversibility analysis of a cascade refrigeration system for various refrigerant couples. Energy Convers. Manage. **51**, 2947–2954 (2010). https://doi.org/10.1016/j.enconman.2010.06.037
19. Sun, Z.; Liang, Y.; Liu, S.; Ji, W.; Zang, R.; Liang, R.: Comparative analysis of thermodynamic performance of a cascade refrigeration system for refrigerant couples R41/404A and R23/R404A. Appl. Energy. **184**, 19–25 (2016). https://doi.org/10.1016/j.apenergy.2016.10.014
20. Sun, Z.; Wang, Q.; Xie, Z.; Liu, S.; Su, D.; Cui, Q.: Energy and exergy analysis of low GWP refrigerants in cascade refrigeration system. Energy **170**, 1170–1180 (2019). https://doi.org/10.1016/j.energy.2018.12.055
21. Sun, Z.; Wang, Q.; Dai, B.; Wang, M.; Xie, Z.: Options of low global warming potential refrigerant group for a three-stage cascade refrigeration system. Int. J. Refrig. **100**, 471–483 (2019). https://doi.org/10.1016/j.ijrefrig.2018.12.019
22. Sarkar, J.; Bhattacharyya, S.; Lal, A.: Selection of suitable natural refrigerants pairs for cascade refrigeration system. Proc IMechE Part A: J Power Energy. **227**(5), 612–622 (2013). https://doi.org/10.1177/0957650913487730
23. Singh, K.K.; Kumar, R.; Gupta, A.: Comparative energy, exergy and economic analysis of a cascade refrigeration system incorporated with flash tank (HTC) and a flash intercooler with indirect subcooler (LTC) using natural refrigerant couples. Sustain. Energy Technol. Assess. **39**, 100716 (2020). https://doi.org/10.1016/j.seta.2020.100716
24. Roy, R.; Mandal, B.K.: Thermo-economic analysis and multi-objective optimization of vapour cascade refrigeration system using different refrigerant combination. J. Therm. Anal. calorim. **139**, 3247–3261 (2020). https://doi.org/10.1007/s10973-019-08710-x
25. Roy, R.; Mandal, B.K.: Energetic and exergetic performance comparison of cascade refrigeration system using R170–R161 and R41–R404A as refrigerant pairs. Heat Mass Transf. **55**, 723–731 (2019). https://doi.org/10.1007/s00231-018-2455-7
26. Paradeshi, L.; Srinivas, M.; Jayaraj, S.: Thermodynamic analysis of a direct expansion solar-assisted heat pump system

working with R290 as a drop-in substitute for R22. J. Therm. Anal. Calorim. **136**(1), 63–78 (2019). https://doi.org/10.1007/s10973-018-7928-x
27. Shovon, M.K.B.; Raman, S.K.; Suryam, A.; Kim, T.H.; Kim, H.D.: Performance of ejector refrigeration cycle based on solar energy working with various refrigerants. J. Therm. Anal. Calorim. **141**, 301–312 (2020). https://doi.org/10.1007/s10973-020-09319-1
28. Saleh, S.; Pirouzfar, V.; Alihosseini, A.: Performance analysis and development of a refrigeration cycle through various environmentally friendly refrigerants. J. Therm. Anal. Calorim. **136**, 1817–1830 (2019). https://doi.org/10.1007/s10973-018-7809-3
29. Turgut, M.S.; Turgut, O.E.: Comparative investigation and multi objective design optimization of a cascaded vapor compression absorption refrigeration system operating with different refrigerants in the vapor compression cycle. Heat Mass Transf. **55**, 467–488 (2019). https://doi.org/10.1007/s00231-018-2430-3
30. Sun, X.; Liu, L.; Dong, Y.; Zhuang, Y.; Li, J.; Du, J.; Yin, H.C.: Multi-objective optimization and thermo-economic analysis of an enhanced compression-absorption cascade refrigeration system and ORC integrated system for cooling and power cogeneration. Energy Convers. Manage. **236**, 114068 (2021). https://doi.org/10.1016/j.enconman.2021.114068
31. Baakeem, S.S.; Orfi, J.; Alabdulkarem, A.: Optimization of a multistage vapor-compression refrigeration system for various refrigerants. Appl. Therm. Eng. **136**, 84–96 (2018). https://doi.org/10.1016/j.applthermaleng.2018.02.071
32. Klein, S.A.; Alvarado, F.L.: EES: engineering equation solver for the Microsoft Windows operating system. F-Chart software (1992)
33. Sanaye, S.; Shirazi, A.: Four E analysis and multi-objective optimization of an ice thermal energy storage for air-conditioning applications. Int. J. Refrig. **36**, 828–841 (2013). https://doi.org/10.1016/j.ijrefrig.2012.10.014
34. Mosaffa, A.H.; Farshi, L.G.: Exergoeconomic and environmental analyses of an air conditioning system using thermal energy storage. Appl. Energy. **162**, 515–526 (2016). https://doi.org/10.1016/j.apenergy.2015.10.122
35. Xu, G.; Liang, F.; Yang, Y.; Hu, Y.; Zhang, K.; Liu, W.: An improved CO2 separation and purification system based on cryogenic separation and distillation theory. Energies **7**, 3484–3502 (2014). https://doi.org/10.3390/en7053484
36. Torrella, E.; Larumbe, J.A.; Cabello, R.; Llopis, R.; Sanchez, D.: A general methodology for energy comparison of intermediate configurations in two-stage vapour compression refrigeration systems. Energy **36**, 4119–4124 (2011). https://doi.org/10.1016/j.energy.2011.04.034
37. Wijbenga, H.; Van Der Hoff, M.; Janssen, M.; Infante Ferreira, C.A.: Life cycle performance of refrigeration systems in the Dutch food and beverages sector. In: 4th IIR conf thermophys prop transf process refrig, Delft, The Netherlands (2016)
38. Infante Ferreira, C.A.: Refrigeration and heat pumping in food processing plants. In: Stougie L, editor. Energy effic qual energy food process ind. Delft, The Netherlands: Interduct: Delft University of Technology. 57–79 (2002)
39. Navidbakhsh, M.; Shirazi, A.; Sanaye, S.: Four E analysis and multi-objective optimization of an ice storage system incorporating PCM as the partial cold storage for air-conditioning applications. Appl. Therm. Eng. **58**, 30–41 (2013). https://doi.org/10.1016/j.applthermaleng.2013.04.002
40. Bejan, A.;, Tsatsaronis, G.;, Moran, M.J.: Thermal design and optimization. Wiley (1996)
41. Fettaka, S.; Thibault, J.; Gupta, Y.: Design of shell-and-tube heat exchangers using multi-objective optimization. Int. J. Heat Mass Transfer. **60**, 343–354 (2013). https://doi.org/10.1016/j.ijheatmasstransfer.2012.12.047
42. Srinivas, N.; Deb, K.: Multiobjective optimization using nondominated sorting in genetic algorithms. Evol. Comput. **2**, 221–248 (1994). https://doi.org/10.1162/evco.1994.2.3.221
43. Knowles, J.; Corne, D.: The Pareto archived evolution strategy: a new baseline algorithm for Multiobjective optimization, in Proc. of the 1999 Cong. on Evolutionary Computation., Washington: IEEE Service Center. (1999)
44. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. **6**, 182–197 (2002). https://doi.org/10.1109/4235.996017
45. Tooru, A.; Hideaki, C.: A new condensed gas calorimeter: thermodynamic properties of solid and liquid dinitrogen oxide. Bull. Chem. Soc. Jpn. **47**(9), 2126–2136 (1974). https://doi.org/10.1246/bcsj.47.2126

Proceedings of the Fifth International Conference on Trends in Electronics and Informatics (ICOEI).
IEEE Xplore Part Number:CFP21J32-ART; ISBN:978-1-6654-1571-2

Multiplex Regulation System With Personalised Recommendation Using ML

Anshul Shroff
Department of CSE
Delhi Technological University
Delhi. India
anshul.shroff223@gmail.com

Bickey Kumar Shah
Department of CSE
Delhi Technological University
Delhi. India
Bickeyshsh721@gmail.com

Aayush Jha
Department of Civil Engineering
Delhi Technological University
Delhi. India
aayushjha452@gmail.com

Amar Kumar Jaiswal
Department of CSE
Delhi Technological University
Delhi. India
jaiswalamar22@gmail.com

Pooja Sapra
Department of CSE
Delhi Technological University
Delhi. India
dspooja17@gmail.com

Manoj Kumar
Department of CSE
Delhi Technological University
Delhi. India
mkumarg@dce.ac.in

2021 5th International Conference on Trends in Electronics and Informatics (ICOEI) | 978-1-6654-1571-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICOEI51242.2021.9453005

***Abstract*—The main aim of this research work is to conduct a feasibility study and requirement analysis for a Multiplex Management system (MMS). This research work includes the design diagrams that facilitate the production of the software and its use in making a frontend and backend for the system according to the elicited requirements, by integrating them. Its alpha version, which strictly adhering to relevant software engineering practices that test the software with suitable methods has also developed. It automates various functions of Multiplex Theater with particular emphasis on ticket reservation, show screening, personnel management, statistical report generation, and increasing ticket sales by giving personalized movie suggestions by learning from the previous booking history or characteristics of users. MMS is a Web application that provides different interfaces for different stakeholders like users, admins, and employees. The admins can add movies and schedule screenings and evaluate the booking statistics and the users and employees can book and cancel tickets. The system also analyzes various factors using Machine Learning to suggest movies to the User to maximize their chances of booking them. The data and model generated from this system are also a precious commodity that can be leveraged successfully to bring in large gains for the Service.**

Keywords—requirement analysis, Machine learning, web application, central Database, Graphical User Interface.

I. INTRODUCTION

The majority of tickets sold are through an online platform compared to the counter at the Multiplex and since a large number of them are not part of big Multiplex Chains developing and maintaining their system is very costly [3]. Also, the data generated through this system is highly valuable from which the Service can derive important analytics and predictions. Multiplexes also employ many people and having some tool to manage them simplifies a lot of work. Manual management of the functions of Multiplex is a cumbersome task because there are multiple screens and simultaneous screenings and there is a scope of various errors. MMS is capable of managing the daily activities of the Multiplex, handle reservations of tickets and also 4 manage employees. MMS tries to automate much of this work and also maximize the revenue in different ways. It increases the efficiency of the employee and also automates the show screening. It helps in giving the movie suggestions and provides the statistical data to the admins and makes informed decisions. This research work includes the analysis of different requirements and all of them are deeply classified into functional and non-functional requirements [1].

A. Functional Requirements

1) Multiplex

- Automate their current multiplex management and ticket reservation system
- Account for employees and staff
- Model the multiplex structure and manage movie and screen timings
- Collecting user data on movie preferences for connecting with their local audience
- Connecting with their users through notifications and updates as well as promotions and branding

2) Customers

- Visually pleasing and intuitive UI design
- Recommend movies and shows similar to their preferences
- Updation of new shows and nearest multiplexes
- Advertising and Promotions

B. NON-functional Requirements

The software needs to be:

- Scalable - As the software needs to be accountable and bring in more multiplexes and more users into their

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:33:53 UTC from IEEE Xplore. Restrictions apply.

fold. The software needs to ensure that it can bear heavier usage and still run as needed.

- Maintainable - The Software needs to be developed such that it is possible for regular maintenance. Since the requirements wouldn't change much once established, regular maintenance would still be required to keep the system performing [2].
- Performance - the system should be able to handle huge loads of requests and the servers shouldn't fail to service the customers during peak times. The database and ticket reservation system should be error-proof and robust.
- Security - As the application collects user data as well as enables online payment systems there must be no security faults and holes in the software that can be exploited by a third party. Security should be of utmost priority and software needs to be designed to ensure the protection of data as well as the safety of online transactions[9].

II. FEASIBILITY STUDY

A. Financial Feasibility

The multiplex management system, being a web application, will subsequently have a cost for web hosting. Since the system does not consist of any multimedia transfer, the bandwidth required for the application is very low. Following freeware software standards, there will be no charges to potential customers [5]. Bug fixes and maintaining/updating the codebase would bear an additional cost for the company [18]. The major source of revenue for the company will be the Multiplexes that opt to use the service to automate their ticket reservation system. The Multiplex Management System can be an extremely attractive alternative to current ticketing systems in place. It can greatly boost their reach as well as retain potential customers by adding a host of secondary services, notifying new shows and offers, and personalized recommendations. The Multiplex is sure to benefit from making use of our system [16]. By offering this service to multiple multiplexes and expanding further users can later choose from multiplexes to view their preferred show. From there it's clear that the Multiplex Management System is financially feasible.

B. Technical Feasibility

The MMS is a complete web application. The Technologies and tools to be used for project development will be open source and freely available. The time limit of developing the product and the ease of using these technologies is synchronized and ensures that our progress matches with projected deadlines. There will be the use of already existing open-source software and APIs such as for the Payment portal.

- Git - Code Management
- Spring Boot
- SQL - Database Management

As the MMS is a web application, hosting the web application would be required. The use of open-source software for parallel development will help the developer teams stick to time constraints as well as keep up with the schedule. From there it's clear that the MMS is technically feasible [4].

C. Social/Legal Feasibility

The MMS uses freely available development tools and provides the system as an open-source system. The maintenance cost will be borne by the company. Since this system is a viable alternate to traditional and outdated ticket reservation systems and also encourages customer interaction with the multiplex, it will have a great impact as a Multiplex Management System[8-10].

D. Risk Feasibility

The feasibility of risk can be determined into different categories of size requirements and requirements changes. size of the codebase will be large as there are multiple stakeholders and as well as a customer portal and a portal for multiplex owners. The size of the Database would be moderately large as multiple multiplexes and users as well as their booking information would build a large temporary database as new bookings will be added and already existing data will be used for the recommender system. The requirements are well defined and understood before the implementation phase and since the requirements are general requirements would be changed only if additional functionality is to be added. Effects of the system on the company revenue matters. The system aims to bring together multiplexes under one umbrella.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:33:53 UTC from IEEE Xplore. Restrictions apply.

Moreover, it helps malls and banks to partner and brings coupons and promos, along with the movie recommender system, which will encourage more customers to be frequently in touch with the cinema. This all points to creating continuous and growing revenue with this system in place. The sophistication of end-users matters a lot. MMS is designed while maintaining the complexity at a very low level. The system is designed to be a highly usable system emphasizing Human-Computer Interaction. Usability is highly improved by providing help documents and making the GUI highly interactive and easy to use [7].

III. LITERATURE SURVEY

MMS is an age-old tradition which has been the primary source of entertainment and hobby for many individuals. It has been more than a century since, the theatre has blessed people with not only entertainment or time pass but has also given them idols to follow and in many cases, worship [13].

As this system has existed for so long in our society, it is only fair to believe that it has been through a tremendous amount of change to still be intact as a practice and be a formidable option for entertainment to this date [8]. This experience of visiting various Multiplex to watch a movie has become more and more of a luxury with the progressing time. For instance, booking of your tickets is a click away and that too with the seat of your choice to give you more of an immersive and a personalized experience. The handling of database with the records of customers and movies has been digitalized which has made the working of these theatres or multiplexes smooth and has definitely enhanced the accessibility of these records and made the current digitalized system more reliable than the traditional handwritten method. It has also made the tracking of employee and brand records much simpler and definitely the modern system of working has given these multiplexes a chance to tie up and generate a good amount of revenue from sponsorships and advertising [6-12].

Thus, the influence of technology has not only just been profitable but also advantageous and less tedious for the functioning of these multiplexes.

IV. METHODOLOGY

The MMS will be made using Incremental Process SDLC Model. An incremental build model is a method of software development where the model is designed, implemented, and tested incrementally (a little more is added each time) until the product is finished. The product is defined as finished when it satisfies all of its requirements. This model combines the elements of the waterfall model with the iterative philosophy of prototyping. The requirements are gathered, and specifications are defined as well as possible at the beginning of the project and an SRS document is produced. After Requirements gathering and specification, requirements are then spitted into several different versions starting with version1, in each successive increment, the next version is constructed and then integrated with the current version. Each incremental version is usually developed using an iterative waterfall model of development. As each successive version of the software is constructed and delivered, now the requirements for the next iteration are decided and incorporated in the next version. We plan to use the Parallel Development Model in which, different subsystems are developed at the same time. It can decrease the calendar time needed for the development, i.e TTM (Time to Market) if enough resources are available.

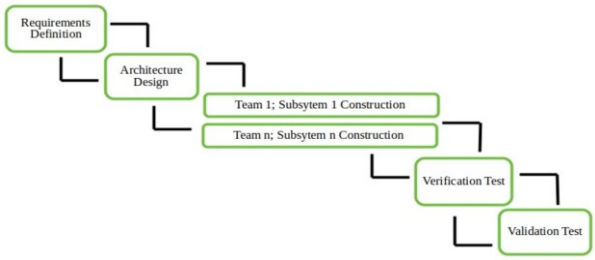


Fig.1 Model Diagram

There are some problems with this model. One is that each new build must be integrated with previous builds and any existing systems. The task of decomposing products into builds not trivial either. If there are too few builds and each build degenerates this turns into the Build And-Fix model. However, if there are too many builds then there is little added utility from each build. The software is then evaluated using some software quality metrics and tested to ensure it is reliable and efficient. In the next step, various documentation is created which are necessary for the maintenance and operation of the software.

A. FACILITIES REQUIRED

As the MMS is a complete web application, emphasis should be given to developing an intuitive and interactive front-end design and an error-free and robust back-end structure. The front end is developed using the popular JS framework while the backend is built along with SQL being used as the primary database. For building the movie recommender system, Machine Learning algorithms will be developed on the Python ML Framework TensorFlow. For code version control and project management Git

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:33:53 UTC from IEEE Xplore. Restrictions apply.

will be the primary tool. Automated Software Testing will be carried out using JUnit.

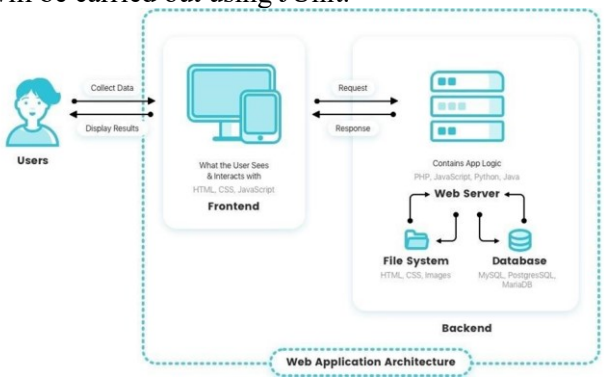


Fig.2 Web application Architecture

In the mentioned fig.2, it is the web application architecture that's shows the relationship and intra connection between the user, front end connection and connection of the frontend with the file system and database. The way and the request and respond is shown in the architecture clearly which is the intermediate od the interface in the frontend and backend steps.

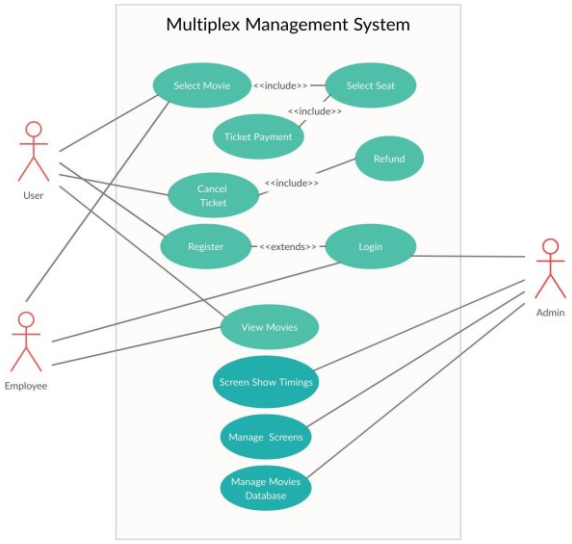


Fig.3 Use case Diagram

The use case diagram is shown in the figure between the user, employee and Admin. It is clearly shown that the user can select movie, cancel tickets and view movies [15]. It can also make the registration process and work. Employee can select movie, login and can view movie. And admin can do 4 tasks of login, screen show timings, manage screens and manage movies database. Intra connection of the tasks are also shown in the pictures.

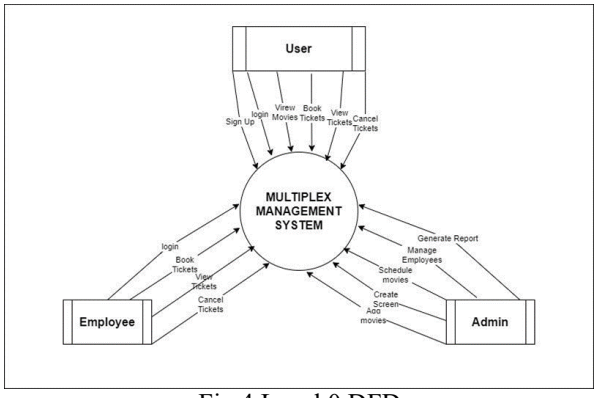


Fig.4 Level 0 DFD

Mentioned in the fig.4 is the data flow diagram of Level 0, It is well shown that how the system is singly related with all the entities. There is a single relationship of all entities with incoming and outgoing arrows.

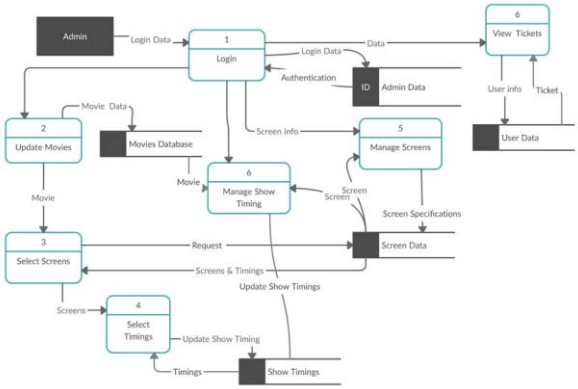


Fig.5 Level 1 DFD Admin

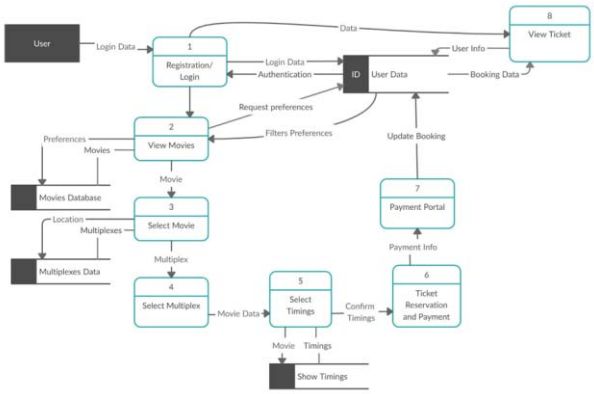


Fig.6 Level 1 DFD User

In the fig.5 and fig.6 the level 1 DFD is mentioned for the user and admin. In both the figures it is clearly noted the depth of the system for both the user and admin. It brings the efficiency to carry out the project

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:33:53 UTC from IEEE Xplore. Restrictions apply.

task. It makes the system understanding efficient and effective [14].

B. RECOMMENDER SYSTEM

The MMS will have a recommender system that trains a model on the user's previous ticket booking and cancellation history to predict the likelihood of the user purchasing tickets for a movie that is currently shown at the multiplex. Psychology of consumer behavior dictates that consumer are likely to avail/purchase products/services that are available to them right away. Therefore, it is intuitive that having movies that the user is more likely to book I the first screen he logs on to increases the chance of tickets being booked. This is especially true when users are quickly glancing at the homepage of the website prompting them to get tickets, whereas customers who use the service with a plan to book specific tickets to movies are not influenced by this as much [11]. We are proposing to use a logistic regression model as the base for the recommender and then build a more advanced model with greater accuracy. Logistic regression is a supervised classification algorithm. In a classification problem, the target variable (or output), y, can take only discrete values for a given set of features (or inputs), X. Contrary to popular belief, logistic regression is a regression model. The model builds a regression model to predict the probability that a given data entry belongs to the category numbered as "1". Just like Linear regression assumes that the data follows a linear function, Logistic regression models the data using the sigmoid function. Logistic regression becomes a classification technique only when a decision threshold is brought into the picture. The setting of the threshold value is a very important aspect of Logistic regression and is dependent on the classification problem itself. The features used in training the recommender are:

- Director
- Actors
- Genre
- Amount of ticket
- Language
- Subtitle-Language
- Time of the show
- Duration

C. APIs USED

- *Online Payment System*

As our system requires the facility for users to pay for the tickets, the book we plan to integrate a feature to pay online. To facilitate this, the system would require the use of already developed and tested Payment Gateways. Using Online Payment API like PayU, RazorPay, or the GooglePay API we can easily integrate it into our system. Configuring the API by using simple post requests the user will be transferred to the safe and secure payment gateway and on completion of the payment the user will be returned to our application.

As our system requires the facility for users to pay for the tickets, the book we plan to integrate a feature to pay online. To facilitate this, the system would require the use of already developed and tested Payment Gateways. Using Online Payment API like PayU, RazorPay, or the GooglePay API we can easily integrate it into our system. Configuring the APIs by using simple post requests the user will be transferred to the safe and secure payment gateway and on completion of the payment the user will be returned to our application.

1) Technologies used

The technologies that are being currently used to make MMS are:

- Desktop application – Java, C#, XAML
- Web application – Javascript, PhP, HTML, CSS
- Frameworks – Spring, Spring boot, React, Node.js
- Database – MySQL, MongoDB
- Diagram and Design tools - app.diagrams.net, Microsoft Project and Visio, NClass, Maven

There are various TMS and MMS available from different vendors available on different platforms and providing different sets of functionalities.

- Vista - Legacy software that has now started adapting to the latest technologies.
- CiniCloud - which provides everything from online ticketing to inventory and personnel management as well as cloud hosting (ticket commission).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:33:53 UTC from IEEE Xplore. Restrictions apply.

- Sony TMS – Most featured to manage multiplexes and internet Ticketing.

  Barco TMS – Mostly the screen/auditorium management.

The technology for the system can be based off of pre-existing technologies and aims to be more user-experience focused with the recommender system aiming to boost customer retention. The system will have many algorithms to suggest movie recommendations, to filter choices based on chosen preferences such as seating, location and other criteria.

V. RESULT ANALYSIS

The project implemented is a web-based user interface that has the options for both the user and admin. It is very much friendly to both of the user and admin. The frontend of the system is completely web based and interface for the user too. Recommender system is applied for the backend part with web partially used. Central databases are functionally used. Desktop applications like Java and C# with HTML, CSS and java script was used for the frontend option with node.js and spring boot framework.

VI. CONCLUSION AND FUTURE WORK

After completion of this project, Admins will be able to structure Screens/Auditoriums, add movies, schedule screenings, add or remove employees as well generate statistical reports [19]. Employees will be able to login and view, book, or cancel tickets. Users can sign up, sign in and view, book, or cancel tickets. The system can generate reports, manage to screen, and make personalized movie suggestions to users. And the actors will be able to edit some of their credentials according to permissions granted. The system cannot be linked to payment gateways and therefore all booking will be regarded as paid for and cancellations reimbursed immediately [17]. The system is limited to booking services for the users and does not offer any other services to them. The system cannot also integrate with movie distributors to project the movies on the screens. The system also does not handle tax-related tasks in accounting. Different newly tools and technologies are being developed day by day, the system can be made more efficient and can be structured again in good methods and design with high computation methods and techniques. We can also use deep learning technologies to enhance the model to the greater extend and bring the effectiveness into the system.

ACKNOWLEDGMENT *(Heading 5)*

We take this opportunity to express our profound gratitude and deep regards to firstly to Professor DR. Manoj Kumar for his exemplary guidance, monitoring, and encouragement throughout this project. His constant guidance and supervision helped a long way in this project's completion. Our thanks and appreciation also goes to our Department of Computer Science and Engineering, and Delhi Technological University for Helping Us to Develop this project.

REFERENCES

[1] O. F. O. R. E. Co-operation and I. E. Agency, "D Iffusion : T He C Ase of," *Energy*, vol. 87, no. October, pp. 645–659, 2006.

[2] Z. Wang *et al.*, "Failure prediction using machine learning and time series in optical network," *Opt. Express*, vol. 25, no. 16, p. 18553, 2017, doi: 10.1364/oe.25.018553.

[3] M. L. Popa, R. Albulescu, M. Neagu, M. E. Hinescu, and C. Tanase, "Multiplex assay for multiomics advances in personalized-precision medicine," *J. Immunoass. Immunochem.*, vol. 40, no. 1, pp. 3–25, 2019, doi: 10.1080/15321819.2018.1562940.

[4] P. Courlet *et al.*, "Development and validation of a multiplex UHPLC-MS/MS assay with stable isotopic internal standards for the monitoring of the plasma concentrations of the antiretroviral drugs bictegravir, cabotegravir, doravirine, and rilpivirine in people living with HIV," *J. Mass Spectrom.*, vol. 55, no. 6, pp. 1–13, 2020, doi: 10.1002/jms.4506.

[5] F. Shen, B. Sun, J. E. Kreutz, E. K. Davydov, W. Du, and R. F. Ismagilov, "Quantification of HIV and HCV viral load with large dynamic range using multivolume digital reverse transcription PCR on a rotational slipChip," *15th Int. Conf. Miniaturized Syst. Chem. Life Sci. 2011, MicroTAS 2011*, vol. 1, pp. 290–292, 2011.

[6] P. Szodoray, P. Alex, J. G. Brun, M. Centola, and R. Jonsson, "Circulating cytokines in primary Sjögren's syndrome determined by a multiplex cytokine array system," *Scand. J. Immunol.*, vol. 59, no. 6, pp. 592–599, 2004, doi: 10.1111/j.0300-9475.2004.01432.x.

[7] M. Cailleaux *et al.*, "Impact of a multiplex PCR assay (FilmArray®) on the management of patients with suspected central nervous system infections," *Eur. J. Clin. Microbiol. Infect. Dis.*, vol. 39, no. 2, pp. 293–297, 2020, doi: 10.1007/s10096-019-03724-7.

[8] J. P. Borde *et al.*, "Diagnosis of prosthetic joint infections using UMD-Universal Kit and the automated multiplex-PCR Unyvero i60 ITI® cartridge system: a pilot study," *Infection*, vol. 43, no. 5, pp. 551–560, 2015, doi: 10.1007/s15010-015-0796-4.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:33:53 UTC from IEEE Xplore. Restrictions apply.

[9] L. Bertazzi, G. Paletta, and M. G. Speranza, "Minimizing the total cost in an integrated vendor-managed inventory system," *J. Heuristics*, vol. 11, no. 5-6 SPEC. ISS., pp. 393–419, 2005, doi: 10.1007/s10732-005-0616-6.

[10] P. M. Valcour, "Managerial behavior in a multiplex role system," *Hum. Relations*, vol. 55, no. 10, pp. 1163–1188, 2002, doi: 10.1177/0018726702055010079.

[11] A. Alonso, *Chapter in DNA electrophoresis protocols for forensic genetics*, no. 2012. 2012.

[12] B. K. Shah, V. Kedia, and R. K. Jha, "Integrated Vendor - Managed Time Efficient Application to Production of Inventory Systems," pp. 275–280, 2021.

[13] B. K. Shah, A. Kumar, and A. Kumar, "Dog breed classifier for facial recognition using convolutional neural networks," *Proc. 3rd Int. Conf. Intell. Sustain. Syst. ICISS 2020*, pp. 508–513, 2020, doi: 10.1109/ICISS49785.2020.9315871.

[14] S. Adhikari, S. Thapa, and B. K. Shah, "Oversampling based Classifiers for Categorization of Radar Returns from the Ionosphere," *Proc. Int. Conf. Electron. Sustain. Commun. Syst. ICESC 2020*, no. Icesc, pp. 975–978, 2020, doi: 10.1109/ICESC48915.2020.9155833.

[15] M. H. Tania, K. T. Lwin, A. M. Shabut, K. J. Abu-Hassan, M. S. Kaiser, and M. A. Hossain, "Assay type detection using advanced machine learning algorithms," *2019 13th Int. Conf. Software, Knowledge, Inf. Manag. Appl. Ski. 2019*, 2019, doi: 10.1109/SKIMA47702.2019.8982449.

[16] H. Ye, L. Liang, G. Y. Li, J. B. Kim, L. Lu, and M. Wu, "Machine learning for vehicular networks," *arXiv*, no. April 2018, 2017.

[17] C. From, "O Vercoming I Nterference in S Patial," *Ieee Wirel. Commun.*, no. December, pp. 95–104, 2007.

[18] D. Ma, W. H. Ki, C. Y. Tsui, and P. K. T. Mok, "Single-inductor multiple-output switching converters with time-multiplexing control in discontinuous conduction mode," *IEEE J. Solid-State Circuits*, vol. 38, no. 1, pp. 89–100, 2003, doi: 10.1109/JSSC.2002.806279.

[19] L. Buitinck *et al.*, "API design for machine learning software: experiences from the scikit-learn project," pp. 1–15, 2013, [Online]. Available: http://arxiv.org/abs/1309.0238.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:33:53 UTC from IEEE Xplore. Restrictions apply.

Trans Indian Inst Met
https://doi.org/10.1007/s12666-021-02348-w

ORIGINAL ARTICLE

Non-Lubricated Sliding Wear Performance of LM13 Alloy Foam and its Composite Foams Reinforced with $ZrSiO_4$

Suresh Kumar[1] · Pradeep Kumar Teotia[2] · O. P. Pandey[3] · Ranvir Singh Panwar[4]

Received: 22 April 2021 / Accepted: 23 June 2021
© The Indian Institute of Metals - IIM 2021

Abstract In this present work, we have studied the effects of the amount and size of reinforced zircon sand particles ($ZrSiO_4$) on the cell geometry and wear performance of aluminium alloy (LM13) composite foam. An Al–Si alloy (LM13) as a matrix, zircon sand particles of different sizes as reinforcement, and $CaCO_3$ as a blowing agent were used to develop the alloy foam and hybrid composite foam. A stir casting process was used to develop alloy foam and its hybrid composite foams. The tribological study of LM13 alloy foam and its composite foam was carried out by using a pin on disc machine under dry sliding conditions at different loads in the range of 9.8–49 N. The results show that a higher amount of zircon sand particles above 5 wt.% with decreasing size leads to a decrement in the size of the cell having thicker cell walls. A comparative tribological study of alloy and its composite foam based on density, cell size, ligament and node size, and foam stability has been presented. An increment in the wear resistance was observed with increasing the amount and also with decreasing the size of reinforced particles.

Keywords Metallic composite foam · Dry sliding wear · Microhardness testing · Optical microscopy · SEM–EDS

✉ Ranvir Singh Panwar
ranvirsinghpanwar@pec.edu.in

[1] Terminal Ballistics Research Laboratory, DRDO, Chandigarh 160030, India

[2] Delhi Technological University, Delhi 110042, India

[3] Thapar Institute of Engineering and Technology, Patiala, Punjab 147004, India

[4] Department of Metallurgical and Materials Engineering, Punjab Engineering College, Chandigarh 160012, India

1 Introduction

Metallic foams are a new range of materials with outstanding properties having excellent strength-density ratio, lightweight structures, energy absorption, and improved high-temperature strength. These properties make them suitable for many industries, such as transportation, aerospace, and sports industries [1–4]. However, these properties of lightweight foam depend upon the processing conditions and have paved the ways for researchers to modify their structures by controlling the cell size of the structure because the cell morphology influences many properties including the tribological properties [2, 5]. Harders et al. [6] reported the mechanical properties of the foams with the variation in morphology of cell structures. It has been observed that with increasing the amount of ceramic particles (fine size) as a reinforcement, wear resistance of the composites improve significantly [7, 8]. Wang et al. [9] worked on that there is a good interface combination between copper foam and resin carbon matrix in the composites after the improvement of interface wettability by inserting a layer of pyrolytic carbon. Mondal et al. [7] studied the tribological properties of cenosphere reinforced aluminium syntactic foam using a pin-on-rotating disc test at different sliding velocities of 2–6 ms^{-1}. Jha et al. [8] carried out a comparative study of the tribological properties of Al-10 wt.% silicon carbide composite and its derived foam. The mechanical properties of foam materials depend on the cell geometry and also the distribution of ceramic particles inside the cell structure as ceramic particles provide strength to cell walls [8]. Metallic foam being a lighter substance is highly useful for transport industries.

The reduction in the weight of vehicles improves fuel efficiency, which is a major factor in the transport industry

Published online: 24 July 2021

[5]. However, the metallic foam used in transportation industries should have a comparable stiffness and strength to materials already being used for safety measures. Also, the impact of sliding load on foam during transportation may cause a lot of damage to the surface, causing wear and tear of materials. However, the packaging of sophisticated materials is an emerging area for the application of metal foam or metal matrix composite foam. Loading and unloading of packed materials by sliding from the cargo transport system is also an important application. So, keeping in mind this novel application of the ultimate loading and unloading system (dock to the trailer) for packaged/palletized products, we focus on the dry sliding wear tests of base LM13 alloy foam and its composite foam.

Many researchers have studied the effect of cell size and cell shape in the presence of different ceramic particles on the mechanical behaviour of metallic foam [6–10]. However, as per our best knowledge, no study exists on the wear performance of metal composite foam reinforced with mineral particles. The low cost and abundance of mineral particles makes them a smart choice for the production of metal composite foam at the industrial level.

Considering these facts, the present work has been taken up where we have reported the wear performance of lightweight LM13 alloy foam and its composite foam reinforced with mineral particles ($ZrSiO_4$) at different loads. The wear behaviour of the developed composite foams has been correlated with the cell size, relative density, and hardness of the cell wall as well as size and amount of reinforced particle which provide strength to cell walls. This wear study will be helpful for its application as a packaging material for the transport industries and many other applications like automobile and defence.

2 Materials and Experimental Procedures

2.1 Used Materials

LM13 alloy (an Al–Si alloy), a profitable grade alloy used as the matrix material was obtained from Emmes Metals Private Limited, Mumbai, India. Blowing agent ($CaCO_3$) with fine grade (1–25 μm) was purchased from SDFCL Maharashtra, India. Zircon mineral particles (also known as Zircon sand particles) were procured from IREL Odisha, India. Zircon sand ($ZrSiO_4$) of the particle size range (50–100 μm as a coarse grade) and (1–25 μm as a fine grade) was used as reinforcement. The particle size of zircon sand and calcium carbonate was analysed by using an optical particle analyser (model CAMSIZER XT by RETSCH–GMBH) which is shown in Fig. 1a–c). A wet chemical method was used to determine the chemical composition of LM13 alloy and mineral $ZrSiO_4$, which is shown in Tables 1 and 2, respectively.

2.2 Processing

For the development of metallic foam, small pieces of LM13 alloy in the desired quantity (1–1.5 kg) were taken in a clay-graphite crucible. The crucible was heated in an electric furnace up to 750 °C at a rate of 5 °C per minute and held at this for 20 min to ensure complete melting. After melting, the slag from the surface of the melt was removed. The melt was rotated mechanically with the help of a three-blade stirrer for 2 min. The rotational speed of the stirrer was increased slowly up to around 630 RPM to create a vortex inside the dynamics of fluid metal [5]. Blowing agent ($CaCO_3$) with 2 wt. % was charged inside of the vortex and mixed properly with the help of a stirrer for three minutes. After uniform mixing of blowing agent ($CaCO_3$) and LM13 alloy melt, the furnace temperature was increased up to 900 °C and held for five minutes to decompose of $CaCO_3$. The released gas (CO_2) forms bubbles and generates the LM13 alloy foam. Finally, the clay-crucible was taken down from the electric kiln and cooled in the open air to get foam. For the development of composite foam, required amount of zircon sand was added first in the vortex during stirring. After rotating the melt for 5 min, 2 wt.% blowing agent ($CaCO_3$) was added to the vortex again. The entire melt was further rotated for 3 min and, the temperature was raised to 900 °C to facilitate the decomposition of $CaCO_3$ and held for 5 min and taken out to cool in the open air. The manufacturing parameter's list is shown in Table 3. Sample pieces with the dimension $12 \times 12 \times 50$ mm^3 of the alloy foam and its composite foam were cut by the hot wire cutter. Foam sample pieces were ground on emery papers of 400–800 grit. A diamond paste (0.5 μm) was used for polishing and Keller's reagent was used for the etching to the bright surface of the foam samples. The density of the developed alloy foam and its composite foam materials were measured using the Archimedes principle. There was co-existence of $ZrSiO_4$ particles, Al matrix, and their interface in LM13 alloy composite foam. Microhardness on these regions (Al-matrix, $ZrSiO_4$ particle, Al-matrix/$ZrSiO_4$ particle interface) was estimated by using Vickers hardness testing machine which was purchased by Mitutoyo, Japan. The Vickers microhardness test was performed on the foam sample by considering at least five indentations at 100 gf load. Wear tests were conducted on the pin-on-disc apparatus at 25 °C. Wear and friction monitor with Model TR-20CH-400 was purchased from Ducom Instruments, Bangalore, India. Wear tests were conducted on a rectangular shaped sample (30 mm × 10 mm) of the LM13 alloy foam and its composites against a rotating EN32 steel disc with a hardness

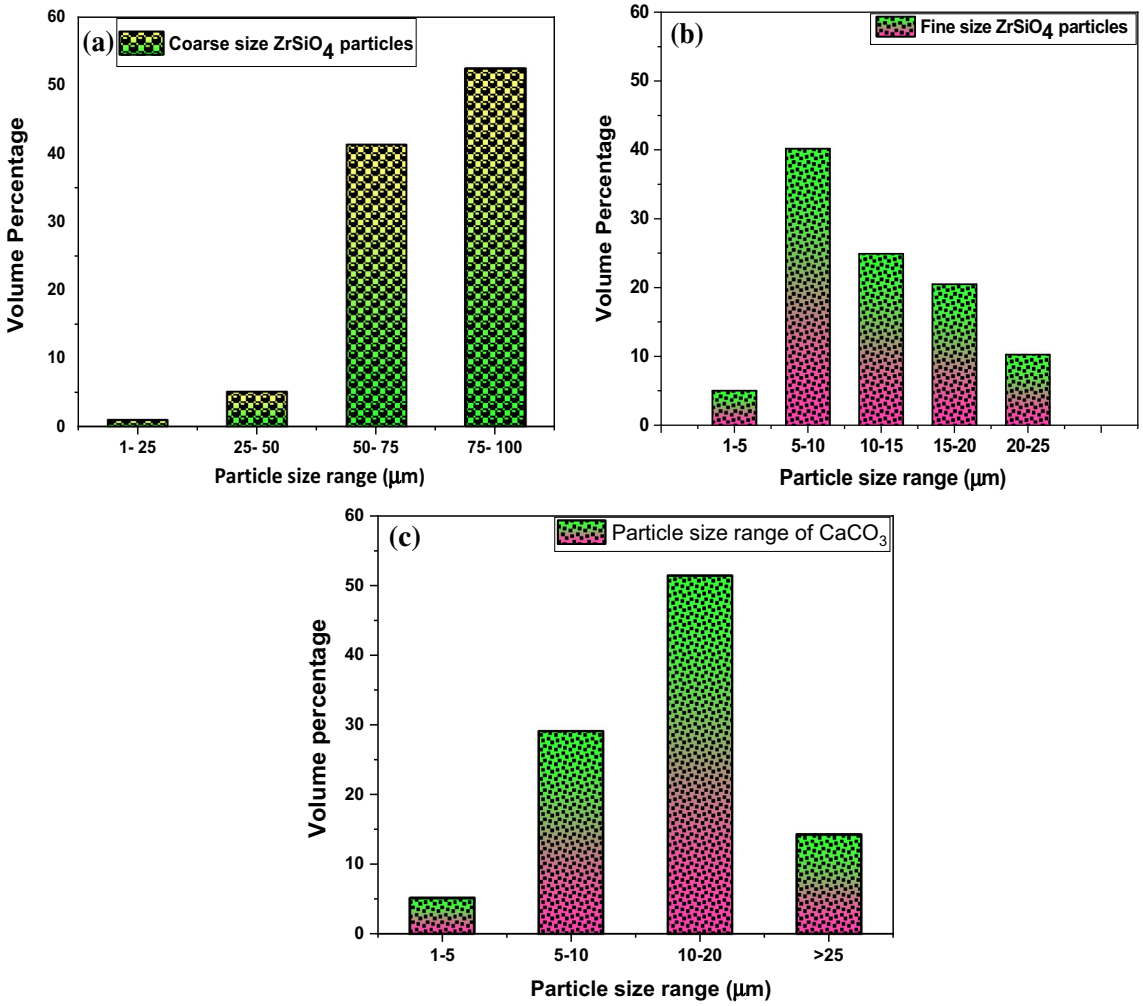


Fig. 1 Volume percentage of particle size range: **a** coarse size, **b** fine size $ZrSiO_4$ particles and **c** blowing agent ($CaCO_3$)

Table 1 Chemical composition of the LM13 alloy in wt.%

LM13 alloy	Si	Fe	Cu	Mn	Mg	Zn	Ti	Ni	Pb	Sn	Al
Wt.%	11.8	0.3	1.2	0.4	0.9	0.2	0.02	0.9	0.02	0.005	Balance

Table 2 Chemical composition of the zircon sand ($ZrSiO_4$)

Elements	ZrO_2	SiO_2	TiO_2	Fe_2O_3
% in bulk	65.30	32.80	0.27	0.12

of 65 HRC. For each type of developed LM13 alloy foam and its composites, wear tests were carried out with a constant sliding velocity (1.6 m/s) and up to a sliding distance of 3000 m at varying applied load from 9.8 to 49 N. Each pin sample was cleaned with acetone and dried before wear testing. The surface morphology of the worn surface was analysed with the help of SEM.

Table 3 List of parameters of the developed metal foams

List of processing parameters for foaming	
Temperature at the time of addition of zircon sand	750 °C
Temperature at the time of addition of blowing agent	750 °C
Foaming temperature	900 °C
Particle size of zircon sand	1–25 μm and 50–100 μm
Particle size of blowing agent	1–25 μm
No. of blades in stirrer	3
Blade angles of stirrer	45°
RPM of stirrer	630
Height of the stirrer in the melt	2/3 inside the melt

3 Results and Discussion

3.1 Microstructure Analysis

The macrographs of LM13 alloy foam and composite foams containing 2.5 wt.% and 15 wt.% of coarse and fine size $ZrSiO_4$ are shown in Fig. 2a–e. Apparently, the cell structure seems to be of the cellular type with a slight variation in their size. The influence of $ZrSiO_4$ particles on the macrostructure of composite foams has been analysed based on cell size and distribution. The composite foam containing 2.5 wt.% of coarse size $ZrSiO_4$ particles is observed with a smaller rounded cell structure as compared with alloy foam which is depicted in Fig. 2a–c. However, the addition of 2.5 wt.% of fine size $ZrSiO_4$ in the molten metal supports the smaller cell with thicker cell faces due to the increased viscosity of the molten alloy (Fig. 2c). Since the number of particles in the same fraction of fine size reinforced composite is more which supports higher viscosity of melt alloy. CO_2 gas pressure is not sufficient to produce a higher volume fraction of uniform cell in the melted alloy, hence cells obtained are smaller. In 15 wt. % of coarse size $ZrSiO_4$ reinforced composite foams (Fig. 2d), the cell geometry and their distribution are controlled by reinforcing ceramic particles which stabilizes the cell walls and prevents them from collapse [11]. Though the viscosity decreases with increasing the size of reinforced particles due to the lower surface area of coarse size particles as compared to fine size particles in the matrix.

The high viscosity and low surface tension are required to make stable composite foam. So, the addition of 15 wt. % of fine size $ZrSiO_4$ in molten metal is a better choice to make stable composite foam (Fig. 2e). However, the wettability of the ceramic phases with liquid metal is affected by bubbles stability. The distribution and position of the zircon particles in the centre of the cell wall and the node are dependent on the interface energy equilibrium at the boundary of zircon particles and the molten matrix. Gergely et al. [12] examined the effect of gas reaction on the cell wall surface and the amount of ceramic particles which is responsible for foamability. According to Ferreira et al. [13], the viscosity of composite continuously increases with an increase in the amount of ceramic particles in the melt. The higher viscosity of molten metal is responsible to delay the cell wall drainage before it solidifies through the adding of $ZrSiO_4$ in the fluid metal [14]. The addition of 15 wt.% fine size $ZrSiO_4$ particles increases viscosity which supports the thick cell walls (Fig. 2e). However, the average cell size, foaming efficiency, and density of the composite foam also change when the amount of the zircon sand is increased. The changes in the structural parameter generally depend on the formation rate of gas bubbles, size, and shape of the bubble.

Figure 3a–e show SEM images of LM13 alloy foam and composite foam containing 2.5 wt.% and 15 wt.% of coarse and fine size $ZrSiO_4$ composite foam, respectively. The slow rate of liquid drainage is due to higher viscosity which is maintained by the addition of finer size zircon sand. It also affects the critical cell wall thickness. It shows that 2.5 wt.% and 15 wt.% of coarse and fine size $ZrSiO_4$ are uniformly distributed within the node in the metal matrix which is depicted in Fig. 3b–e. Uniform distribution of zircon sand particles in cell walls and node is expected for better wear behaviour and mechanical properties. The high particle concentration at the node and ligament is responsible to increase the strength with decreased ductility. The same structural morphology is found in fine size zircon sand reinforced composite also. However, the strength at the node and ligament due to the higher particle concentration of the fine particles as compared to the coarse particle will be more. $ZrSiO_4$ particles perform as a barricade for the dendritic growth and this behaviour rapidly grows when the solidification rate of molten metal is fast. The presence of $ZrSiO_4$ in the Al matrix of node and ligament causes a modification in dendritic structure (Fig. 3f), due to thermal conductivity mismatch at the interface of $ZrSiO_4$ and melt [5]. The morphology of Si also gets modified from acicular to fine structure as shown in Fig. 3f. The effect of the amount and size of zircon sand on the cell geometry parameters and density of LM13 alloy foam and its composite foams is shown in the form of graphs in Fig. 4a–d.

3.2 Microhardness Measurement

The mechanical properties of alloy foam are affected by cell geometry and its uniform distribution. The smaller and uniform distribution of cells in the foam materials can be achieved with high strength. The microhardness measurement of LM13 alloy foam and composite foams in different regions has been carried out to know the effect of the amount and size of zircon particles in the alloy matrix. The effect of the amount and size of reinforced particles on the microhardness of LM13 alloy foam is given in Table 4. The high microhardness at the interface of $ZrSiO_4$/Al reveals the good interfacial bonding of $ZrSiO_4$ with Al-matrix. It is observed that composite foam reinforced with fine size particles shows better microhardness in comparison to composite foam reinforced with coarse size particles. Composite foam with 15 wt.% fine size $ZrSiO_4$ shows higher Vickers microhardness in comparison to 15 wt.% coarse size $ZrSiO_4$ reinforced composite foam and other composite foam at the interface of the $ZrSiO_4$–Al-matrix and Al-matrix phase. Vickers microhardness of the

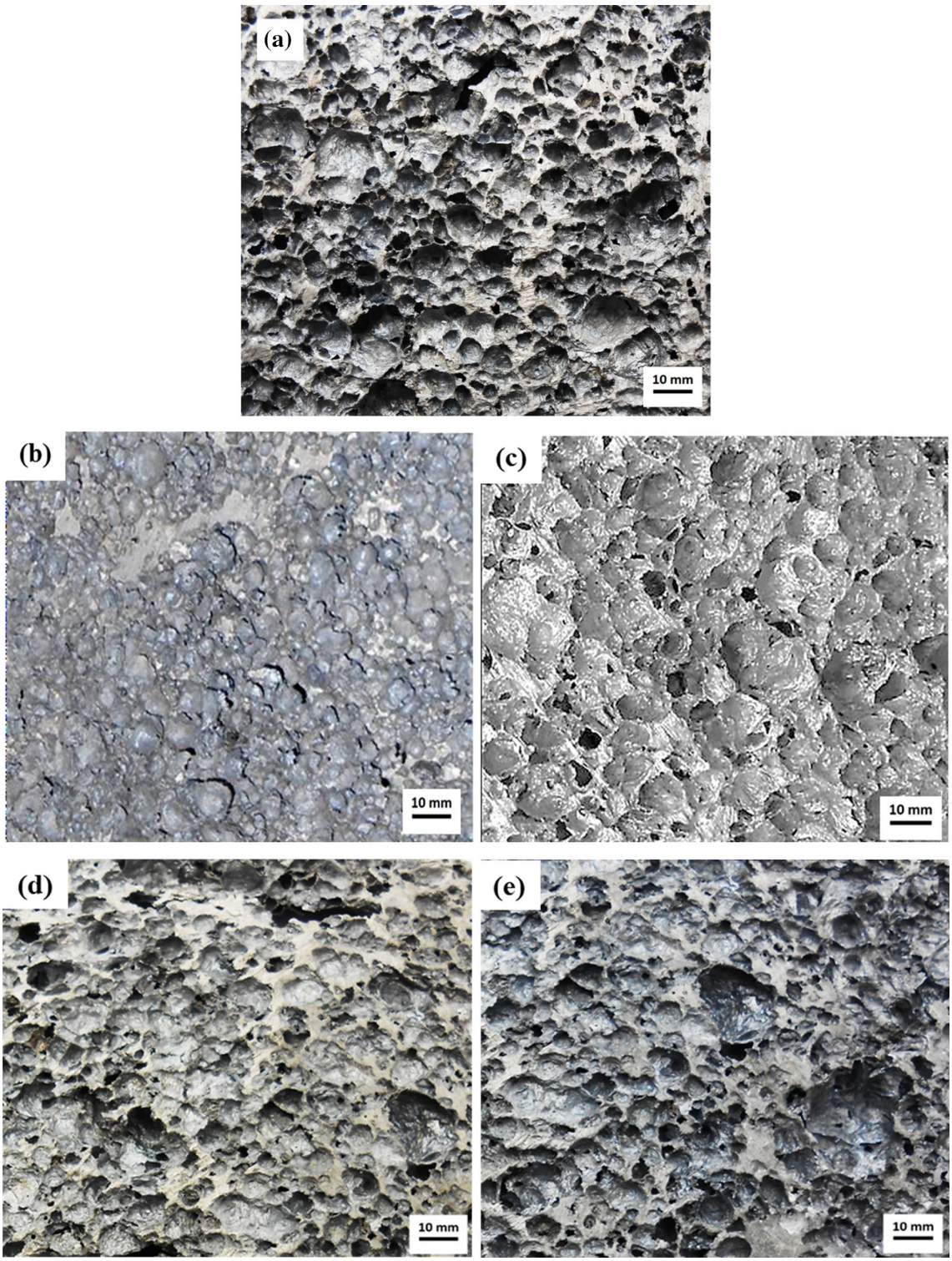


Fig. 2 Macrographs of LM13 alloy foam **a** LM13 alloy foam, **b** 2.5 wt.% coarse, **c** 2.5 wt.% fine, **d**15 wt.% coarse, and **e** 15 wt.% fine size reinforced composite foam

composite reinforced with 15 wt.% of fine size particles at the interface and matrix phase is due to the lower inter-particle distance which is responsible for higher interface region (brittle phase) per unit area in the matrix that increases the hardness of the composites at different phases.

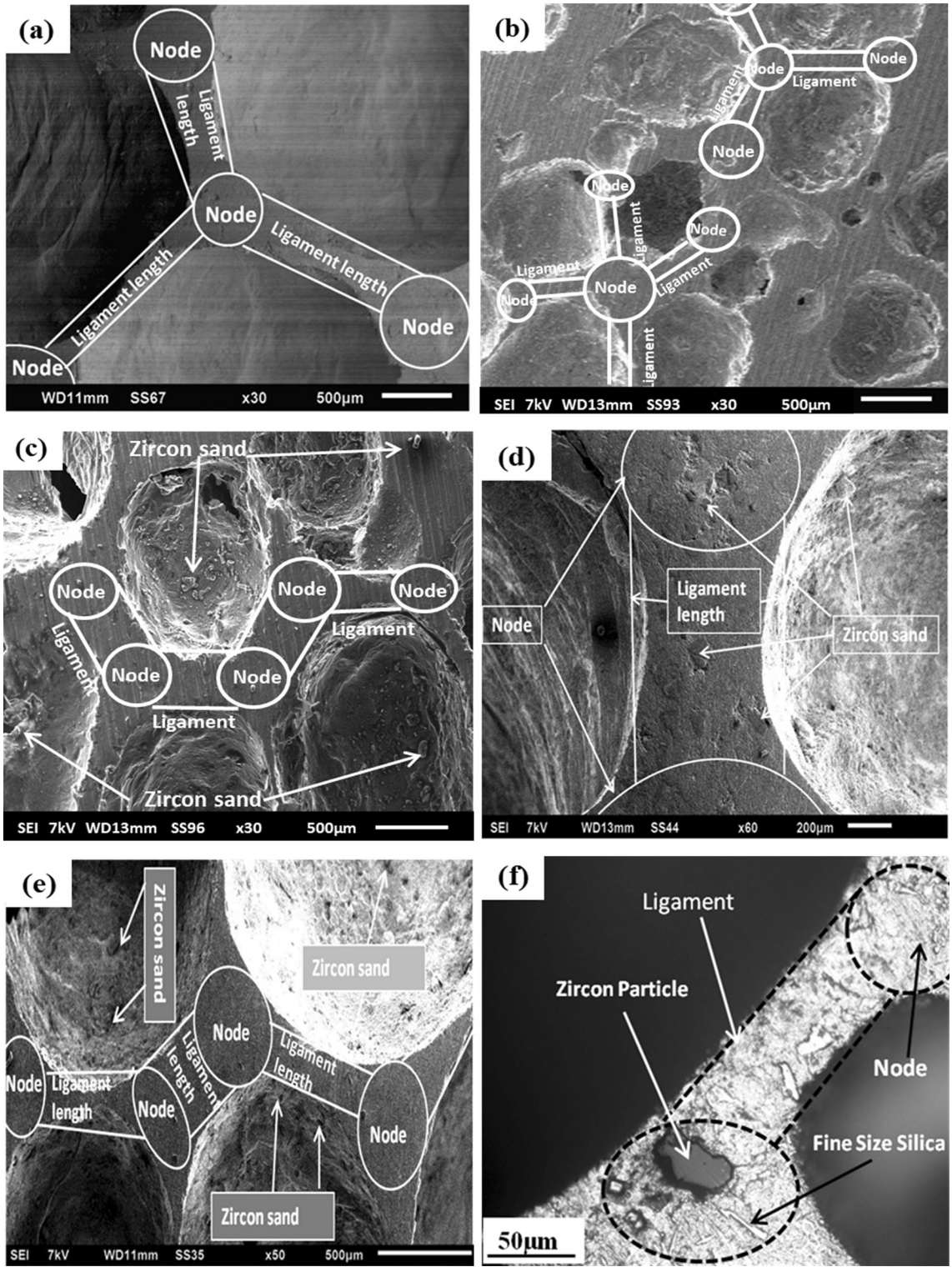


Fig. 3 SEM images of **a** LM13 alloy foam **b** 2.5 wt.% coarse, **c** 2.5 wt.% fine, **d** 15 wt.% coarse, **e** 15 wt.% fine and Optical image of **f** 15 wt.% coarse composite foam

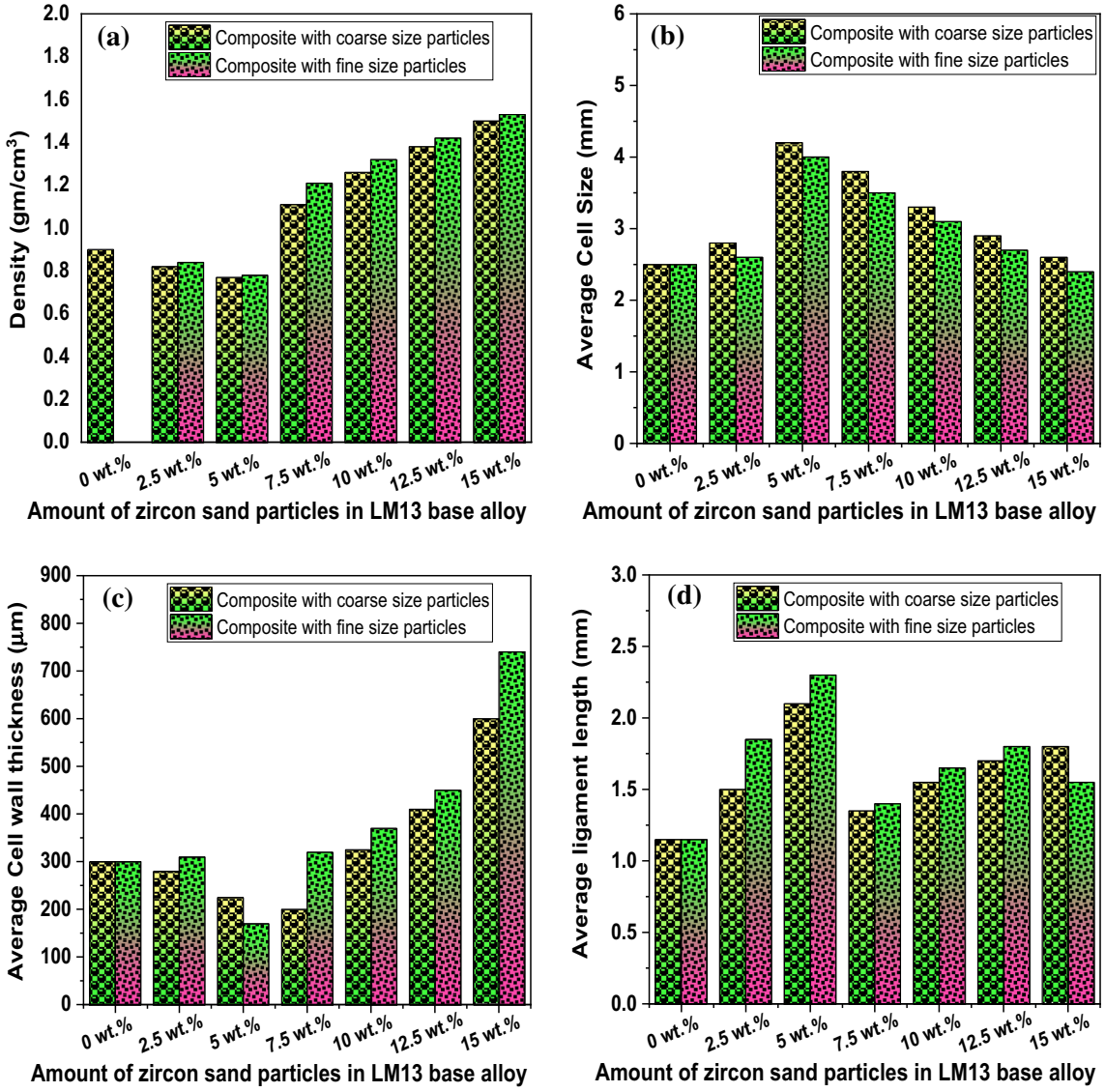


Fig. 4 Effect of the amount and size of zircon sand on the cell parameters and density of LM13 alloy foam and its composite foams

3.3 Analysis of Wear Behaviour

3.3.1 Impact of Applied Load on Wear Behaviour

The wear behaviour of LM13 alloy foam and its composite foams reinforced with 2.5 wt.% and 15 wt.% of $ZrSiO_4$ particles (coarse and fine size) performed at room temperature (at around 25 °C) is shown in Figs. 5 and 6. Wear tests of all developed composite were carried out and all composites showed the similar type of wear trend for respective operating parameters. However, to avoid the confusion due to large data, we provide the wear graph only for two wt.% reinforced composite foams having higher microhardness and average cell wall thickness.

It can be understood from the graph that the wear rate increases with an increase in the applied loads and sliding distance of foam materials. Wear graphs in Figs. 5 and 6 show that the wear resistance of LM13 alloy composite foam samples is much better in comparison to the base alloy foam samples. However, the wear rate of both alloy and composite foam samples increases with the increase in applied loads. This could be ascribed to the increase in the real contact area between the foam sample and counter disc [6]. The trend of wear behaviour could be alienated into two different parts for all foam samples with sliding distance as (a) run-in wear state: where wear rate changes abruptly concerning the sliding distance and (b) steady-state: where a change in wear rate is not significant with respect to sliding distance. Initially, in the run-in wear state, the wear resistance is low as related to the steady-state-wear with sliding distance at all loads. The wear rate in the run-in wear state increases abruptly due to the

Table 4 Variation of microhardness of LM13 alloy foam and its composite foam and different phases of composite foam (superscript shows amount of reinforcement and subscript for coarse and fine particles)

Microhardness (Hv)				
LM13 alloy/composite foam	At particle	At matrix	At interface	At node
LM13 alloy foam	–	69.49	–	69.60
LM13/2.5$_c$ZrSiO$_4$	650.55	68.28	117.34	70.54
LM13/2.5$_f$ZrSiO$_4$	652.45	78.54	118.72	78.91
LM13/5$_c$ZrSiO$_4$	655.93	80.88	126.19	78.10
LM13/5$_f$ZrSiO$_4$	667.22	85.60	130.92	86.40
LM13/7.5$_c$ZrSiO$_4$	666.45	85.89	130.78	86.82
LM13/7.5$_f$ZrSiO$_4$	670.51	86.56	132.12	88.79
LM13/10$_c$ZrSiO$_4$	665.55	87.78	132.33	88.46
LM13/10$_f$ZrSiO$_4$	666.74	88.12	133.67	90.78
LM13/12.5$_c$ZrSiO$_4$	664.98	88.55	134.54	90.45
LM13/12.5$_f$ZrSiO$_4$	671.34	89.34	136.67	91.55
LM13/15$_c$ZrSiO$_4$	669.56	88.34	137.34	92.12
LM13/15$_f$ZrSiO$_4$	672.22	90.56	139.12	93.54

abrasive wear of materials during the sliding [9]. The abrasive wear rate is related to the number of asperities which contacts with the counter surface, as the number of contact points is directly related to the real contact area and applied load. Wear rate is linearly related to contact area and plastic deformation rate of asperities. In all the studies, the steady-state is attained after the sliding distance of nearly 1300 m. As the applied load is increased from 9.8 to 49 N, the number of asperities engaged in ploughing is changed, and the wear rate is increased. During sliding, asperities get deformed plastically due to frictional heating increment with increasing the load to increase the contact area. The surface area of the fine $ZrSiO_4$ particles in the Al-matrix leads to an increase in the load transfer rate from the softening phase (matrix) to the harder phase (mineral particles). This higher load transfer rate has provided better wear resistance of fine size particle reinforced composite foam in comparison to coarse size particle reinforced composite foam. Generally, with the increased applied load on the composite sample, wear rate increases in a similar pattern as has been observed by Vishwakarma et al. [15]. During dry sliding, the thin layer of oxide shields the surface of the metal and brings two solid surfaces in contact [16, 17]. During sliding, the oxide film get fractured, which leads to exposing the substrate material under heavy load. This leads to the plastic deformation of the surface beneath. This close contact also leads to the welding of the materials at the edges of the cells. However, some will be trapped within the cells also. Comparison of wear rate of LM13 alloy foam and its composite foam with a constant sliding velocity at 1.6 m/s up to a sliding distance of 3000 m at variable loads from 9.8 to 49 N are shown in Fig. 7a–b. The decrement in the wear rate of the LM13 alloy foam and its composite foam is observed (Fig. 7a) with increasing the amount of $ZrSiO_4$ particles at 9.8 N load. Similarly, wear behaviour is noticed (Fig. 7b) with the increase in the amount of $ZrSiO_4$ particles at higher applied load 49 N. In the same way, the minimum wear rate is observed for fine size $ZrSiO_4$ particle composite foams as compared to coarse size $ZrSiO_4$ particles composite foams.

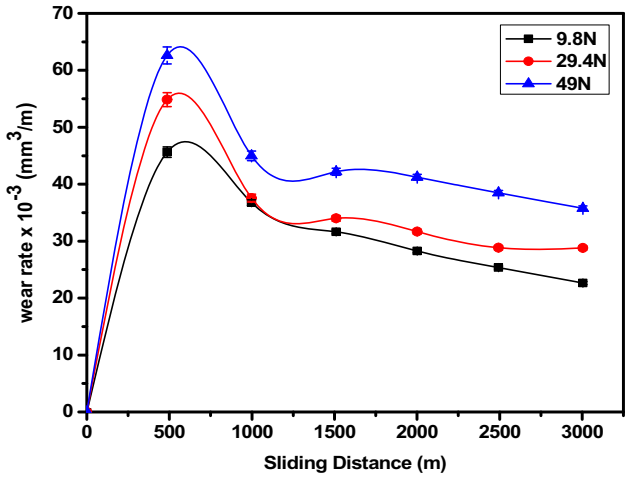


Fig. 5 Wear rate of LM13 alloy foam against sliding distance at different loads

3.3.2 *Impact of $ZrSiO_4$ Size and its Amount on Wear Behaviour*

Wear performance depends on the cell geometry of the foam. The size and amounts of reinforcing particles are important factors in determining the tribological behaviour

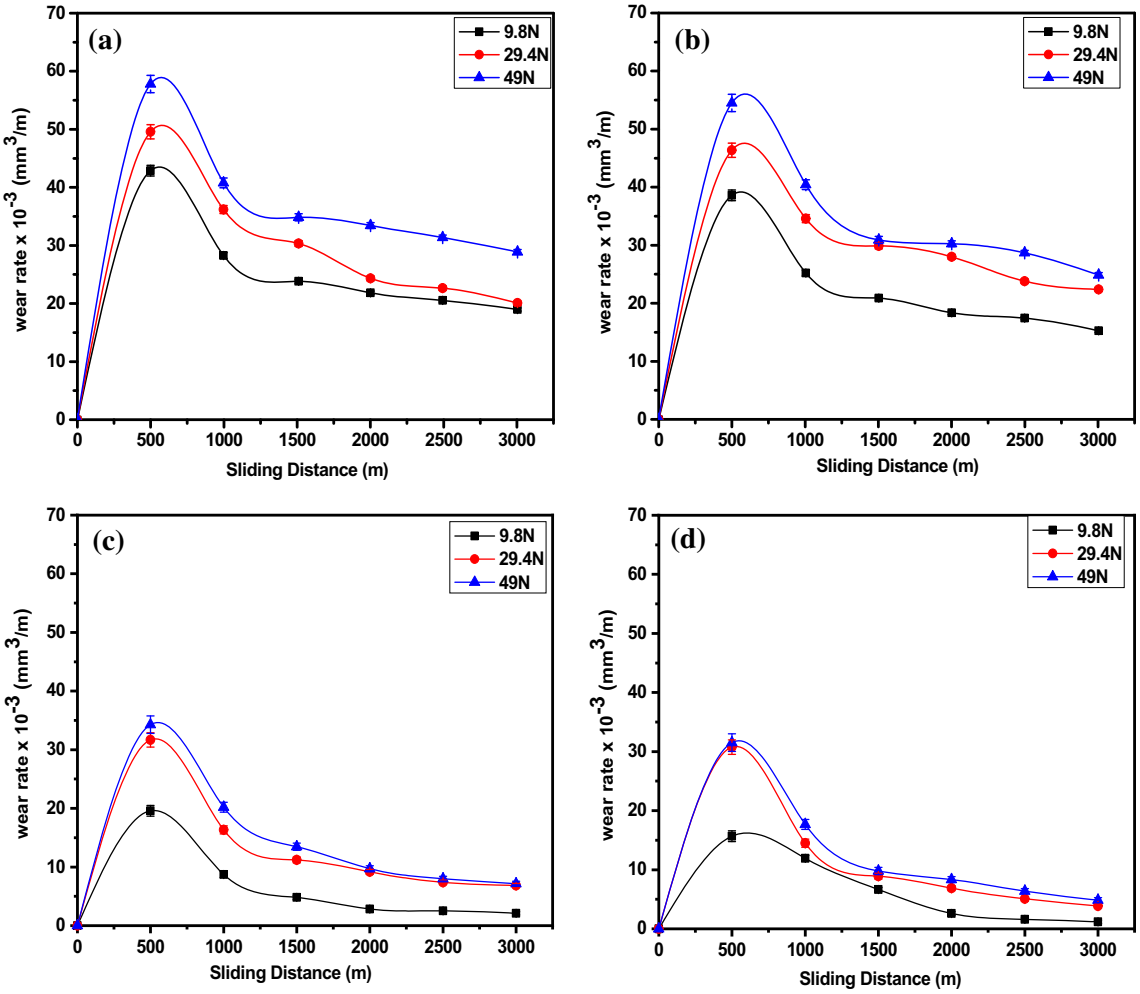


Fig. 6 Wear rate against sliding distance at different loads for **a** LM13/2.5$_c$ZrSiO$_4$, **b** LM13/2.5$_f$ZrSiO$_4$, **c** LM13/15$_c$ZrSiO$_4$, and **d** LM13/15$_f$ZrSiO$_4$ composite foam

of composite foam. It is observed from the wear graph that the composite foams show a lesser wear rate as compared to alloy foam, though the foaming temperature is the same. The wear behaviour of the alloy foam and composite foams (with fine and coarse particle reinforcement) at all applied load and sliding distance is represented in Fig. 8. It is observed from the wear graphs from Figs. 5 and 6 that the wear rate in the run-in state remains almost the same up to the applied load of 9.8 N and after this load wear rate increases sharply when the load increases from 29.4 to 49 N. The transition mode in wear performance is changed due to increased applied load [8]. This could be due to several reasons. However, the domination factor is cell structure where cell acts as craters which reduces the real contact area between pin sample and counter disc [8].

The abrasive nature of wear is controlled by the asperity-to-asperity real contact between the pin sample and the counter disc. The real contact area between the pin sample and counter disc depends on the roughness of these surfaces. The foam sample and counter surface contain large asperities of different geometry (height, shape, and sharpness) [5]. The size, height, and shape of asperity and inter-asperity distance of the ligaments and nodes of foam are controlled by the amount, size, distribution nature of reinforced ZrSiO$_4$ in the matrix and applied loads.

Kapoor and Johnson [18] theoretically estimated the steady-state wear of two real contact asperities. They have shown that the geometrical asperities, such as heights and curvatures, will be modified during the wear test, in a manner that the dependable real contact pressure drops to the shakedown limit. Also, sliding will involve the least amount of plastic deformation of asperities on the node of foam. As the applied pressure is increased, wear rate increases due to the decrease in separation between real contact surface areas, and frictional heat affects the geometry of asperities with the passage of the highest tough asperities of the ligament and node. The hard asperities (particles) will grow higher with applied contact

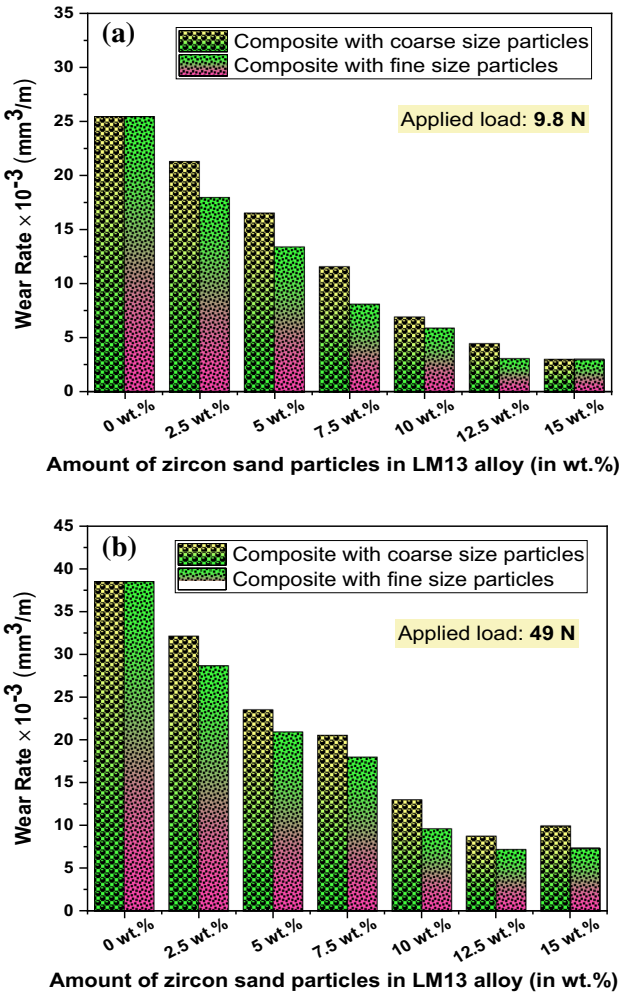


Fig. 7 Comparison of wear rate of LM13 alloy foam and its composite foam at **a** 9.8 N and **b** 49 N loads

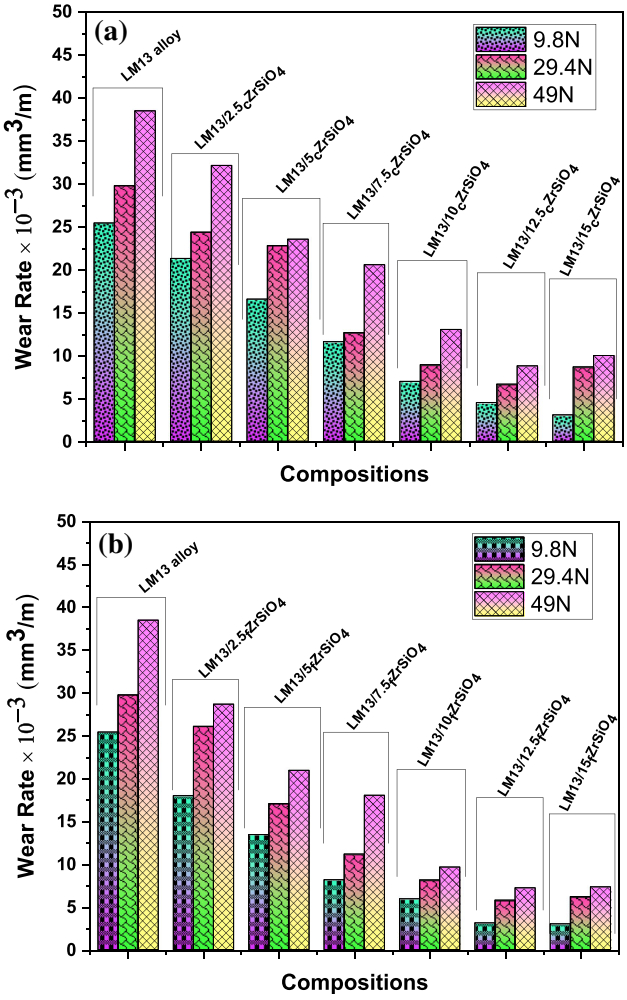


Fig. 8 Comparison of wear rate at different loads for LM13 alloy foam and its composite foam reinforced with **a** coarse size particles and **b** fine size particles

pressures in comparison with lower neighbours. The number of asperities with fine size $ZrSiO_4$ composite foams is more than that of the coarse size $ZrSiO_4$ composite foams. But the asperity size and inter-asperity distance of the ligament and the node of composite foams with fine size $ZrSiO_4$ depends on the particle-to-particle distance. However, fine size $ZrSiO_4$ composite foams have a smaller particle to particle distance in comparison to the composite foams with coarse size $ZrSiO_4$. Though, particle to particle distance decreases with the increased amount of reinforcement in the developed composite foams with increased applied load up to 49 N. So, the distance between real contact surfaces, reduced and frictional heat-affected asperities change the plastic deformation zone by the passage of the highest tough asperities.

Figure 5 exposes that LM13 alloy foam possesses thinner cell wall breadth with the minor actual contact area, which is the key reason for lower wear resistance as compared to the composite foam material. This is because of the lower contact area, as the transmittance rate of load per unit area of the foam sample is lower. At this condition, the contact stresses build up on the surface, which leads to a higher material removal rate of alloy foam. Composite foam specimens having a large cell wall thickness with a larger actual contact area leads to a fast load transmission rate per unit area and increased wear resistance. Figure 8 depicts the comparative study about the effect of the amount and size of zircon sand particles on the wear behaviour of LM13 alloy foam and its composites foam at different applied loads from 9.8 to 49 N. From the comparative study (Fig. 8), it can easily be observed that the wear rate of the LM13 alloy decreases with increasing the amount of reinforced zircon sand particles extensively. However, the size of the reinforced particle also plays an important role in the wear behaviour of LM13 alloy foam. By decreasing the size of particles, the wear rate of composites foam decreases significantly. Hence, composite

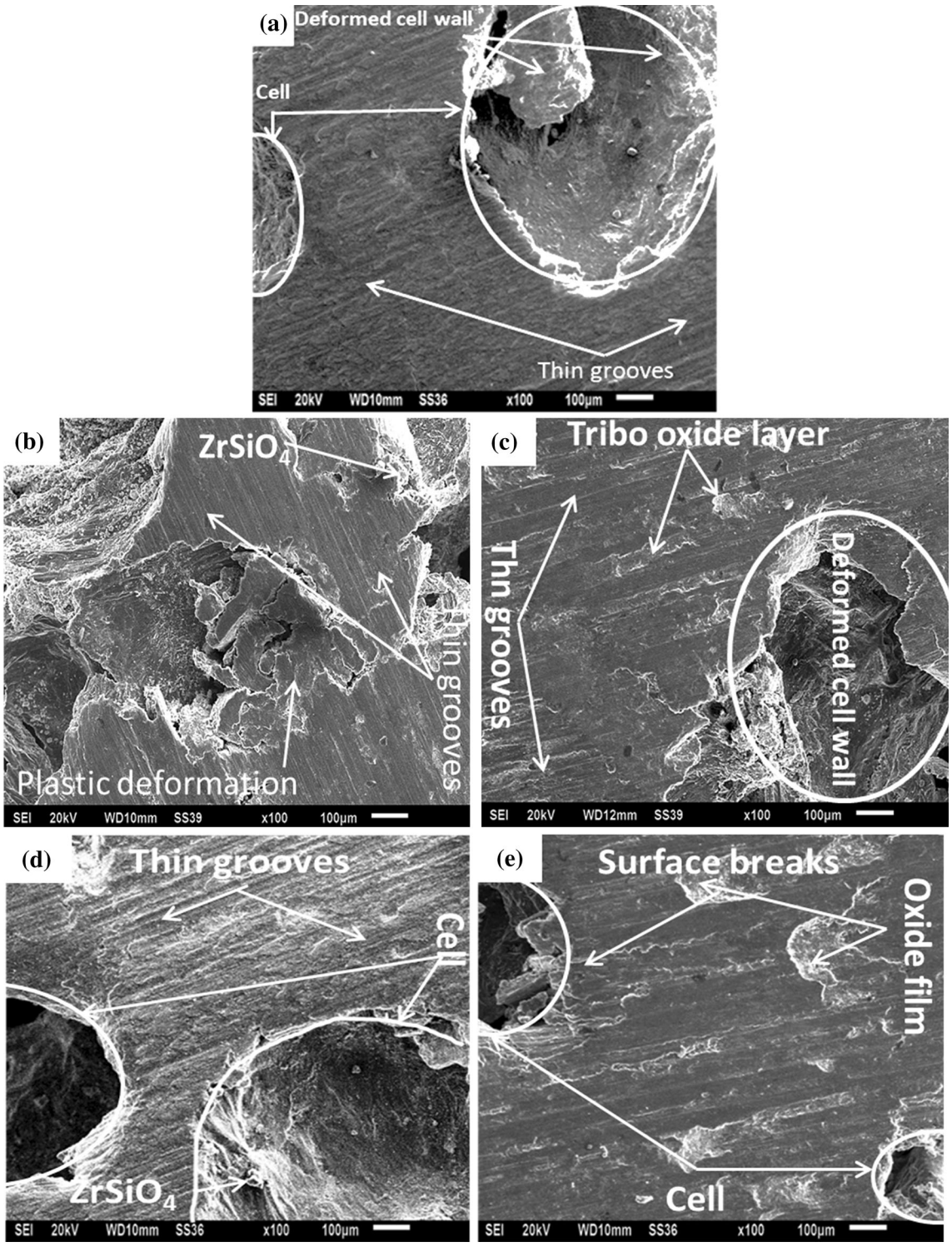


Fig. 9 SEM images of worn surfaces for **a** LM13 alloy foam, **b** LM13/2.5$_c$ZrSiO$_4$, **c** LM13/2.5$_f$ZrSiO$_4$, **d** LM13/15$_c$ZrSiO$_4$, and **e** LM13/15$_f$ZrSiO$_4$ composite foam

foam reinforced with a higher amount of fine size particles will have better wear resistance.

The composite foams with fine zircon particles have a large number of particles per unit area dispersed in the Al-matrix and the distance between particle-to-particles decreases. With the increase in the amount of zircon particles from 2.5 to 15 wt.%, nearly a similar type of trend is observed in Fig. 6a–d. At higher loads (49 N), there is a

significant change in the wear rate performance with 12.5 wt.% zircon sand composite foam compared to 15 wt.% zircon sand composite foam. The higher contents of zircon sand in composite foams lead to the late transition in the wear mode. The slope of wear rate for the fine zircon sand composite is less when compared to coarse size zircon sand. This is because of the material removal rate of 15 wt.% fine zircon sand composite foam is lower than 15 wt.% coarse size zircon sand composite foam. The above study clearly shows that composite foam with 15 wt.% fine zircon comes out to be improved wear-resistant foam.

3.3.3 Analysis of Wear Tracks

It is observed from the micrographs that the wear behaviour remains nearly the same up to the applied load of 9.8 N and the wear rate increases suddenly when the applied load is increased from 29.4 to 49 N. Examination of worn surfaces gives a better clue about the wear mechanism involved during the wear test. To avoid the confusion of similar types of micrographs, we have provided the SEM images of worn surfaces only for two wt.% and base alloy foam as shown in Fig. 9a–e. Smaller thin grooves and scratches at a few locations on the worn surface of the foam can be observed from the microstructure (Fig. 9a). Some craters help in depositing the wear debris due to applied contact pressure and get crushed in the craters [7]. Thus, the deformation rate of the material is reduced from the surface and reduces the degree of wear. In the same way, thin grooves, scratches and plastic deformation on the worn surface are seen in Fig. 9b. Some areas represent the tribo oxide layer and deformation of cell wall on the wear track in Fig. 9c. The depth, size, and the number of grooves present on the worn surface of composite foam (Fig. 9d–e) are lesser than that on alloy foam (Fig. 9a–c). The groove geometry (size, shape, height, and the number of grooves) designed on the worn surface of coarse size $ZrSiO_4$ reinforced composite foam (Fig. 9d) is large compared to fine size $ZrSiO_4$ reinforced composite foam (Fig. 9e). These phenomena can be enlightened as weak interface bonding between coarse zircon particles with Al-matrix in coarse size $ZrSiO_4$ reinforced composite foam supporting to higher removal of material. Many grooves due to plastics delamination, on the node and the wall, are observed in both foam samples, but with a decrease in wear rate performance in fine size zircon composite foam (Fig. 9e). This phenomenon is mostly due to the softening of the matrix phase by the frictional heat between the foam pin and steel disc. The shared effect of frictional heating and surface plastic deformation on the worn surface is represented on the mixed layer of fine $ZrSiO_4$ particle and Al-matrix; it makes a robust MML

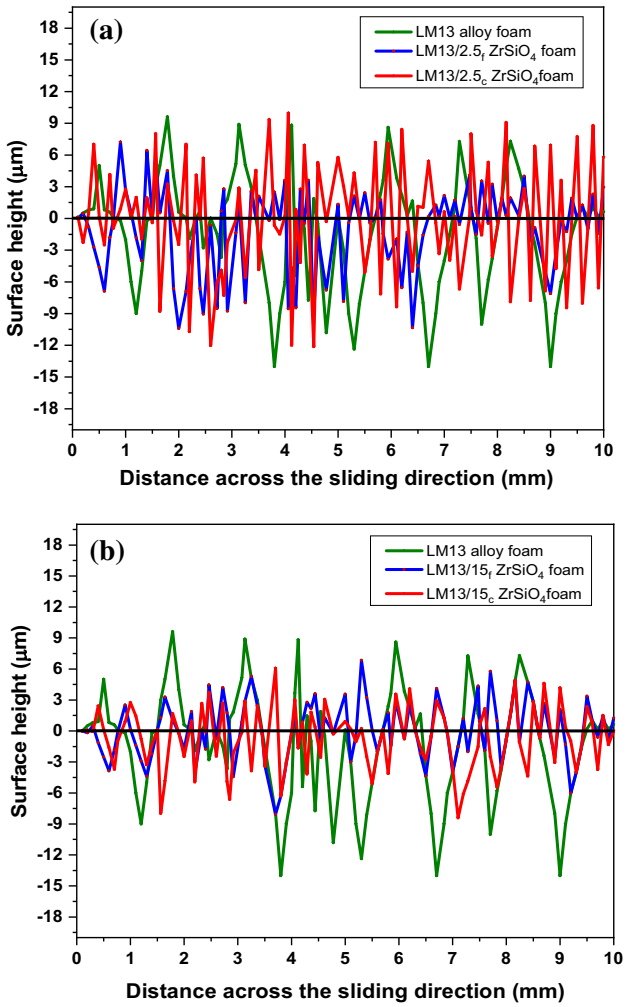


Fig. 10 Comparative surface texture of worn surfaces

(mechanically mixed layer) specifically in the presence of high applied contact pressure over the surface. At the higher applied contact pressure, during dry sliding, the oxide film (Al_2O_3/Fe_2O_3) working as a shield cover to the metallic surface breaks, thus bringing the two solid surfaces in contact (Fig. 9e). This fractured oxide film leads to the plastic deformation of the surface [5].

The temperature at the interface depends on the thermal boundary resistance of the matrix phase and the reinforced phase. Now, assuming a constant heat flux, due to this interfacial thermal resistance, a temperature difference exists between the matrix phase and reinforced phase at their interface. This thermal boundary resistance is due to carrier scattering at the interface and it is known as interfacial thermal resistance. The interfacial thermal resistance depends on the difference in the CTE (coefficient of thermal expansion) between the $ZrSiO_4$ particle and Al-matrix. The surface morphology of the wear track of LM13 alloy foam under different applied loads seems similar in the nature of asperities on the node and ligament of cells. The

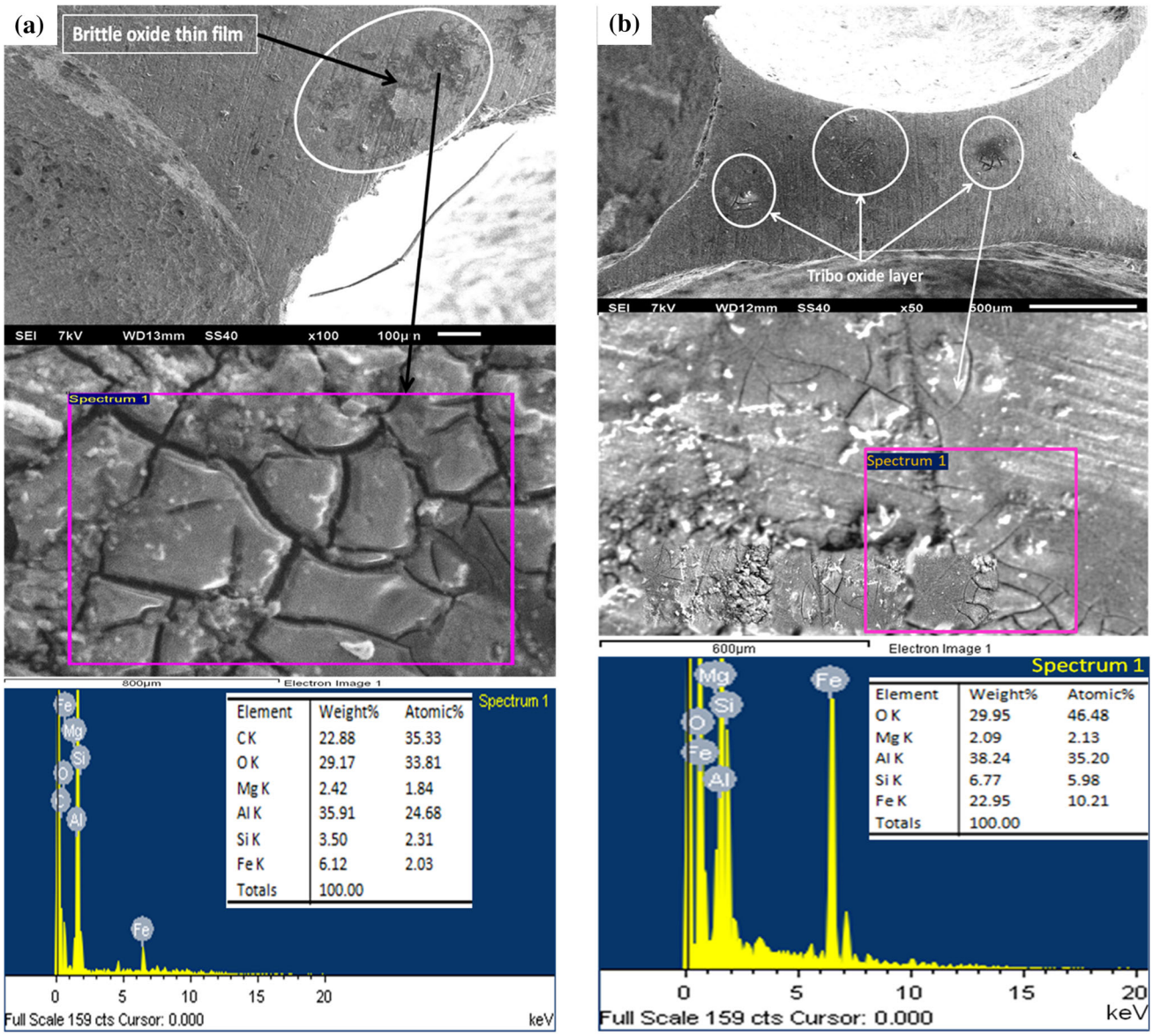


Fig. 11 SEM–EDS images of brittle oxide thin film on the wear track of **a** LM13/15$_f$ZrSiO$_4$ and **b** LM13/15$_c$ZrSiO$_4$ foam at 49 N load

plastic deformation of asperities is pressed and buried into the cells and welded with the parent material due to the softening of the matrix at high contact pressure. It is also found that higher friction can introduce more energy into the sliding interfaces. This energy can transfer to thermal energy, causing softening and removal of more materials from the worn surface. This leads to the increased wear rate of the alloy foam. However, in the composite foam materials, this energy is not easily transferred to thermal energy because ceramic particle acts as a barrier. This barrier lowers the softening rate of the LM13 alloy composite foam and decreases the wear rate performance. It is observed from the wear track morphology of both alloy foam (Fig. 9a) and composite foam samples (Fig. 9b–e), that alloy foam suffers more wear loss as compared to composite foam.

By using a Mitutoyo make surface roughness tester (surftest SJ-310), to avoid repetition and confusion with similar data, in this section we have planned to study the surface texture of the worn surfaces of LM13 alloy foam and its composite foams as shown in Fig. 10a–b. The surface texture of worn surfaces depicts that surface depth and width is higher in LM13 alloy in comparison to the composite foam, which supports the higher wear rate of LM13 alloy. The surface texture of LM13/15$_f$ZrSiO$_4$ foam shows more roughness along with lower surface depth and smaller width supporting a comparatively lower wear rate.

3.3.4 EDS Analysis of Wear Tracks

SEM–EDS pattern of brittle oxide thin film on the worn surface of composite foam with 15 wt.% coarse and fine size zircon sand particles at an applied load 49 N is shown Fig. 11a–b. It is observed that, during the sliding wear, some of the ceramic particles have been pulled out or have fractured to fine size as the pin sample come in contact with the rotating steel disc and acts as an abrasive mode. The crushed ceramic particles along with Fe_2O_3 from the counter surface and Al_2O_3 from the pin foam sample escape from the worn surface in the form of MML. The transfer of deformation materials between $ZrSiO_4$ particle and Al-matrix results in MML on the sample surface during sliding [6]. The mechanically mixed layer enhances the wear resistance by providing a brittle Fe_2O_3/Al_2O_3 oxide and Al_4C_3 carbide layer covering between the coupling surfaces [16, 19]. Hence, zircon sand particles enhance wear resistance of composite foams than alloy foam by the creation of MML at the interface of $ZrSiO_4$ particle and Al-matrix, which is confirmed by EDS analysis [18] and can be depicted in Fig. 10a. At higher contact pressure, the repetition of the formed brittle Fe_2O_3/Al_2O_3 oxide layer is shown on the surface with subsurface cracks/microcrack network propagation [14]. The intersection of these cracks/microcracks on the oxide layer results in the formation of shallow cracks in the worn surface of composite foam (Fig. 10a). At higher applied load 49 N, a similar SEM–EDS pattern of brittle oxide thin film on the worn surface of composite foam is noticed (Fig. 11b) with increase in the amount of Fe_2O_3/Al_2O_3 oxide layer.

4 Conclusions

1. LM13 alloy foam and LM13/$ZrSiO_4$ composite foam material with the help of blowing agents ($CaCO_3$) were successfully developed using a stir casting process.
2. The influence of the amount and size of $ZrSiO_4$ on the cell geometry and wear performance of alloy foams/composite foam was studied. The nature of the distribution of $ZrSiO_4$ was homogenous in the matrix. The microhardness of composite foam samples increased with the decreasing size of reinforced zircon particles.
3. The dendritic structure in the composite foam matrix changed with the presence of the reinforced particles in the matrix. As a result, the wear resistance performance of composite foam was better than the alloy foam.
4. The presence of $ZrSiO_4$ in the matrix enhanced the tribological properties as the composite foam was able to perform well with increased applied loads and sliding distance at constant velocity. However, wear rate performance was reduced by decreasing cell size and increasing the density of both alloy foam and composite foam.
5. The wear rate performance of fine size $ZrSiO_4$ composite foams got enhanced in comparison to both alloy foam and the coarse size $ZrSiO_4$ composite foams (if mixed in the same amounts of reinforcement). Formation of brittle Fe_2O_3/Al_2O_3 oxide and Al_4C_3 layer on the sliding components which had self-lubricating properties led to the improvement of the wear performance of the composites. The brittle oxide layer played a very vital task in reducing the wear rate performance.
6. The $ZrSiO_4$ particles acted as a barrier in the matrix phase of LM13 composite foam, which resisted the flow of thermal energy as compared to LM13 alloy foam. The presence of barrier adherently lowered the softening and wear rate of the alloy composite foam.

Data Availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

[1] Banhart J. Metal Foams: Production and stability. Adv Eng Mater 2006; 8:781-94.

[2] Vendra LJ, Brown JA, Rabiei A. Effect of processing parameters on the microstructure and mechanical properties of Al-steel composite foam. J Mater Sci 2011; 46:4574-81.

[3] Májlinger K, Kalácska G, Orbulov IN, Zsidai L, Bozóki B, Keresztes R. Global approach of tribomechanical development of hybrid aluminium matrix syntactic foams. Tribo Lett 2017; 65:16-29.

[4] Wang, S., Wang, Y., Li, C., Chi, Q., Fei, Z. The Dry Sliding Wear Behavior of Interpenetrating Titanium Trialuminide/Aluminium Composites. Appl. Compos. Mater., 2007; 14(2), 129–144.

[5] Kumar S, Sharma V, Panwar RS, Pandey OP. Wear behavior of dual particle size (DPS) zircon sand reinforced aluminum alloy. Tribol Lett 2012; 47:231-51.

[6] Harders H, Hupfer K, Rösler J. Influence of cell wall shape and density on the mechanical behaviour of 2D foam structures. Acta Mater 2005; 53:335-45.

[7] Mondal DP, Das S, Jha N. Dry sliding wear behaviour of aluminum syntactic foam. Mater. Des 2009; 30:2563-68.

[8] Jha N, Badkul A, Mondal DP, Das S, Singh M. Sliding wear behaviour of aluminium syntactic foam: A comparison with Al-10 wt% SiC composites. Tribol Int 2011; 44:220-31.

[9] Wang, P, Deng, G, Zhu, H, Yin, J, Xiong, X. Effect of Pyrolytic Carbon Interface Thickness on Conductivity and Tribological Properties of Copper Foam/Carbon Composite. Appl. Compos. Mater.. 2021; 28:219-233.

[10] Fricke S, Hager C, Solovyeva S, Wangenheim M, Wallaschek J. Influence of surface form deviations on friction in mixed lubrication. Tribol Int 2018;118:491-99.

[11] Akinci, A., Yilmaz, S., Sen, U. Wear Behavior of Basalt Filled Low Density Polyethylene Composites. Appl. Compos. Mater., 2011; 19(3-4), 499–511.
[12] Gergely V, Curran DC, Clyne TC. The FOAMCARP process: Foaming of aluminium MMCS by the chalk-aluminium reaction in precursors. Compos Sci Technol 2003; 63:2301-10.
[13] Ferreira SC, Velhinho A, Rocha LA, Braz Fernandes FM. Microstructure characterization of aluminium syntactic functionally graded composites containing hollow ceramic microspheres manufactured by radial centrifugal casting. Mater Sci Forum 2008; 207:587-88.
[14] Kumar S, Pandey OP., Role of fine size zircon sand ceramic particle on controlling the cell morphology of aluminum composite foams. J Manuf Process 2015; 20:172-80**.**
[15] Vishwakarmaa A, Mondal DP, Birla S, Dasa S, Prasanth N. Effect of cenosphere size on the dry sliding wear behaviour LM13-cenosphere syntactic foam. Tribol Int 2017;110:8-22.
[16] Májlinger K, Bozóki B, Kalácska G, Keresztes R, Zsidai L. Tribological properties of hybrid aluminum matrix syntactic foams. Tribol Int 2016;99:211-23.
[17] Daoud A. Synthesis and characterization of novel ZnAl22 syntactic foam composites via casting. Mater Sci Eng A 2008; 488:281-295.
[18] Kapoor A, Johnson KL. Steady State Topography of Surfaces in Repeated Boundary Lubricated Sliding (Elsevier, 1993).
[19] Chu, D., Ma, Y., Tang, P., Li, P. Effect of Thermal Exposure on the Interface Microstructure and Interfacial Shear Strength of the SiC Fiber Reinforced $AlFe_5Si_2$ Matrix Composite. Appl. Compos. Mater.. 2020; 27:95–105.

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

Analog Integrated Circuits and Signal Processing
https://doi.org/10.1007/s10470-021-01915-x

MIXED SIGNAL LETTER

Novel architecture of four quadrant analog multiplier/divider circuit employing single CFOA

Ajishek Raj[1] · Data Ram Bhaskar[1] · Pragati Kumar[1]

Received: 26 April 2021 / Revised: 4 July 2021 / Accepted: 7 July 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract
In this communication, a current feedback operational amplifier (CFOA) based novel architecture of voltage-mode analog multiplier/divider circuit has been proposed. The proposed circuit employs a single CFOA and four MOSFETs. The multiplier circuit can operate in four quadrant modes, whereas, the divider circuit can operate in two quadrant modes. The applications of the proposed multiplier circuit in amplitude modulator, squarer, and frequency doubler circuit have also been presented. To check the robustness of the proposed circuit, mismatch analysis, process corner-voltage-temperature analysis, and Monte-Carlo simulations have been performed. Noise analysis has also been carried out and the output noise for multiplier and divider circuits are found to be below 0.28 μV /sqrt(Hz). The presented circuit has been simulated with CMOS CFOA implemented using 0.18 μm TSMC technology parameters.

Keywords Current feedback operational amplifier · Voltage mode · Analog multiplier · Analog divider · Analog computational circuits

1 Introduction

In instrumentation and communication systems, analog multiplier and divider circuits may be used in various applications such as waveform generator, modulator, frequency doubler, phase-locked loop, gain control amplifier, amplitude locked loop, squarer, data converters, peak detectors, phase detectors, frequency synthesizers as well as in analog computational systems based on biological 'neural' paradigms [1–4]. Previously, many researchers have reported multipliers/dividers using different active building block(s) viz. operational amplifiers (Op-amps) [5–8] (wherein an interesting novel versatile all MOS VM analog reconfigurable multiplier/divider parameterised cell has been reported [7, 8], which employs a single active device (Op-amp) and eight MOSFETs to achieve different non-linear analog applications), operational transconductance amplifiers (OTAs) [9–13], current conveyors and their variants [14–19], current differencing transconductance amplifiers and their variants (CDTA) [20–22], current controlled CTTA [23], current follower transconductance amplifier (CFTA) [24], operational transresistance amplifiers (OTRAs) [25–27], current amplifier [28], current differencing buffered amplifiers (CDBAs) [29, 30], current controlled current differencing buffered amplifiers [31], dual-X second-generation current conveyor (DXCCII) [32], extra-X second generation current conveyor (EXCCII) [33], voltage differencing transconductance amplifiers (VDTAs) [34], etc. Some CMOS based multipliers and divider circuits have also been reported in open literature [35–39].

CFOA, because of its typical architecture, allows realization of constant bandwidth variable gain amplifiers, with very high slew rates, and thus, has been used for many linear analog circuit applications viz, inductance simulators, capacitance multipliers, filters and oscillators [40]. CFOAs have also been used for the realization of non-linear analog functions like multipliers and dividers [41, 42]. An analog divider circuit employing two CFOAs and four MOSFETs was reported in [41]. A differential analog multiplier cum

✉ Ajishek Raj
ajishek_phd2k16@dtu.ac.in

Data Ram Bhaskar
drbhaskar@dtu.ac.in

Pragati Kumar
pragatikumar@dtu.ac.in

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

Published online: 15 July 2021

Table 1 Features of the previously reported VM/CM analog multiplier and divider circuits with proposed circuit

Ref	Active device used	Quadrant		Technology	Output impedance	Supply voltage	Output noise	Power dissipation	BW	Input range	
		Multiplier	Divider							Multiplier	Divider
[5]	3 Op-amps, 6R, 2 M	Four	NA	0.35 µm	Low	1 V	14nV $\sqrt{Hz}$	0.75mW	9.6 MHz	$-0.4\ V \le V_1 \le 0.4\ V$ $-1\ V \le V_2 \le 1\ V$	NA
[6]	5 Op-amp, 11R	Four	NA	Bipolar	Low	8 V	NM	NM	0.84 MHz	$-1.5\ V \le V_x \le 1.5\ V$ $-1\ V \le V_y \le 1\ V$	NA
[7]	1 Op-amp, 8 M	Four	Two	Bipolar	Low	NM	NM	NM	NM	$-0.5\ V \le V_x \le 0.5\ V$ $-0.8\ V \le V_y \le 0.8\ V$	$0.25\ V \le V_x, V_y \le 0.25\ V$ $0\ V \le V_z \le 1\ V$
[8]	1 Op-amp, 8 M	Four	Two	CMOS	Low	5 V	NM	NM	NM	$-1.5\ V \le V_y \le 1.5\ V$ $-0.8\ V \le V_x \le 0.8\ V$	$0.8\ V \le V_y \le 0.8\ V$ $0.25\ V \le V_z \le 1.25\ V$
[9]	3 OTAs	Four	Two	Bipolar	High	10 V	NM	NM	186 MHz	$-1\ mA \le I_1 \le 1\ mA$ $-1\ mA \le I_2 \le 1\ mA$	NM
[10]	3 OTAs	Four	Two	Bipolar	High	10 V	NM	NM	155 MHz	$i_1 = \pm 1\ mA$ $i_2 = \pm 0.9\ mA$	$1\ mA \le I_B \le 2\ mA$
[11]	3 OTAs, 2B, 1C	NA	Two	3 um	High	5 V	NM	10mW	NM	$-0.75\ V \le V_1 \le 0.75\ V$ $-1\ V \le V_2 \le 1\ V$	$-0.75\ V \le V_1 \le 0.75\ V$ $V_2 \le 0$
[12]	4 OTAs,1R	Four	NA	0.13 µm	High	1.2 V	450nV $\sqrt{Hz}$	NM	6.56 MHz	NM	NM
	6 OTAs, 2R	NA	Two	0.13 µm	High	1.2 V	280nV $\sqrt{Hz}$	NM	6.56 MHz	NM	NM
[13]	1 OTA, 2 M, 1B	NA	Two	0.18 µm	High	0.9 V	NM	0.821mW	4.5 GHz	NA	$0.5\ V \le V_1 \le 1.5\ V$ $-50\ mV \le V_2 \le 50\ mV$
[14]	1 CCII + , 1 CCII- , 2 M	Four	NA	Bipolar	High	NM	NM	NM	NM	$-0.1\ V \le V_1 \le 0.1\ V$ $-0.5\ V \le V_2 \le 0.5\ V$	NA
	2 CCII + , 1 CCII- , 6 M	NA	Two	Bipolar	High	NM	NM	NM	NM	NA	$-0.1\ V \le V_1 \le 0.1\ V$ $-0.5\ V \le V_2 \le 0$
[15]	DVCC and Gilbert Cell	Four	NM	0.35 µm	High	1.65 V	NM	1.08 mW	90 MHz	$-0.2\ V \le V_1 \le 0.2\ V$ $-0.2\ V \le V_2 \le 0.2\ V$	NM
[16]	2 CCCII +	Two	NM	Bipolar	High	5 V	NM	NM	20 MHz	$i_1 = \pm 0.9\ mA$ $i_2 = \pm 0.1\ mA$	I = 0.1 mA
[17]	3 CCCII +	Four	Four	Bipolar	High	2 V	NM	NM	NM	$-0.2\ mA \le I_1 \le 0.2\ mA$ $20\mu A \le I_2 \le 300\mu A$	NM
[18]	1 CCIII + , 1 CCIII-, 6 M, 1B	Four	Two	0.35 µm	High	1.6 V	NM	NM	112.63 MHz	$-0.3\ V \le V_x \le 0.3\ V$ $-0.2\ V \le V_y \le 0.2\ V$	$1.05\ V \le V_z \le 1.35\ V$
[19]	2 CCII	Four	Four	Bipolar	High	1.5 V	NM	NM	3 MHz	$0.2\ mA \le I_1 \le 0.2\ mA$ $80\mu A \le I_2 \le 80\mu A$	$-40\mu A \le I_3 \le 40\mu A$
[20]	2 CDTA	Four	Two	Bipolar	High	3 V	NM	NM	30 MHz	NM	NM

Table 1 (continued)

Ref	Active device used	Quadrant		Technology	Output impedance	Supply voltage	Output noise	Power dissipation	BW	Input range	
		Multiplier	Divider							Multiplier	Divider
[21]	1 DO-CCCDTA	Four	Two	Bipolar	High	1.5 V	NM	1.39 mW	25.58 MHz	$150\mu A \leq I_1 \leq 150\mu A$ $120\mu A \leq I_2 \leq 120\mu A$	$20\mu A \leq I_b \leq 200\mu A$
[22]	1 CCCDTA	Two	Two	Bipolar	High	1.5 V	NM	NM	26.94 MHz	$20\mu A \leq I_c \leq 100\mu A$ $-10\mu A \leq I_a \leq 10\mu A$	$10\mu A \leq I_b \leq 100\mu A$
[23]	1 MO-CCCTTA	Four	Two	Bipolar	High	1.5 V	NM	1.83 mW	53.13 MHz	$-150\mu A \leq I_1 \leq 150\mu A$ $-120\mu A \leq I_2 \leq 120\mu A$	$20\mu A \leq I_a \leq 200\mu A$
[24]	2 CFTA	Four	Two	Bipolar	High	3 V	NM	NM	NM	$-0.2\ V \leq V_x \leq 0.2$ $-0.2\ V \leq V_y \leq 0.2$	$-0.15\ V \leq V_y \leq 0.15$ $-0.1\ V \leq V_z \leq 0.5$
[25]	1 OTRA, 6 M	Four	Two	0.5 µm	Low	1.5 V	NM	NM	NM	NM	NM
[26]	1 OTRA, 4 M	Four	NA	0.5 µm	Low	1.5 V	NM	0.83 mW	8 MHz	$-0.3\ V \leq V_1 \leq 0.3$ $-0.15\ V \leq V_2 \leq 0.15$	NA
[27]	1 OTRA, 4 M	NA	Two	0.25 µm	Low	1.5 V	NM	1.05 mW	NM	NA	$-0.1\ V \leq V_1 \leq 0.1\ V$ $-50\ mV \leq V_2 \leq 50\ mV$
[28]	1 CA	Two	Two	0.5 µm	High	2.5 V	NM	6.34 mW	NM	$10\mu A \leq I_1 \leq 100\mu A$ $5\mu A \leq I_2 \leq 25\mu A$	$0 \leq I_3 \leq 100\mu A$
[29]	1 CDBA, 6 M	Four	Two	0.35 µm	Low	5 V	NM	NM	82.31 MHz	$-0.23\ V \leq V_x \leq 0.23$ $-0.18\ V \leq V_y \leq 0.18$	$-2\ V \leq V_z \leq -0.6\ V$
[30]	1 CDBA, 4 M	Four	Two	0.35 µm	Low	2.5 V	NM	0.28mW	68.55 MHz	$-0.2\ V \leq V_x \leq 0.2$ $-0.25\ V \leq V_y \leq 0.25$	$-9\ V \leq V_z \leq -1\ V$
[31]	2 CCCDBA	Two	Two	Bipolar	High	1.5 V	NM	1.39 mW	26.42 MHz	$20\mu A \leq I_c \leq 80\mu A$ $-0.1\ mA \leq I_a \leq 0.1\ mA$	$10\mu A \leq I_b \leq 100\mu A$
[32]	1 DXCCII, 2 M	Four	NA	0.18 µm	High	0.9 V	NM	NM	19.3 GHz	$\pm$ 0.25 V	NA
[34]	1 EXCCII, 2 M	Four	NA	0.18 µm	High	1.25 V	NM	1.2 mW	NM	$\pm$ 0.4 V	NA
[41]	2 CFOAs, 4 M	NA	Two	Bipolar	Low	12 V	NM	NM	1.8 MHz	NM	NM
[42]	5CFOAs, 6 M, 4 C	NM	NM	Bipolar	Low	12 V	NM	NM	60 MHz	NM	NM
Proposed	1 CFOA, 4 M	Four	Two	0.18 µm	Low	1.25 V	$70nV/\sqrt{Hz}$ $280nV/\sqrt{Hz}$	1.23mW	200 MHz 182 MHz	$-0.2\ V \leq V_1 \leq 0.2$ $-0.1\ V \leq V_2 \leq 0.2$	$-0.2\ V \leq V_2 \leq 0.2$ $0.5\ V \leq V_3 \leq 2$

M MOSFET, *R* Resistor, *C* Capacitor, *B* Buffer, *BW* Bandwidth, *NA* Not available, *NM* Not Mentioned

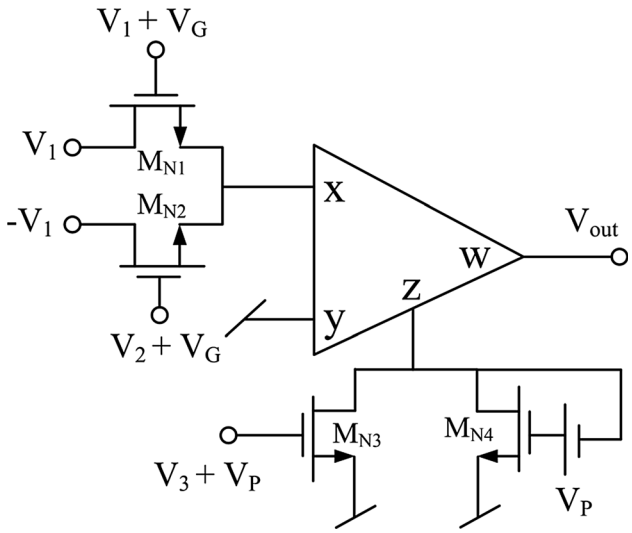


Fig. 1 VM analog multiplier/divider configuration

divider circuit architecture employing five CFOAs, six matched MOS transistors (operating in triode region) and four identical capacitors was presented in [42]. Table 1 below presents various features of previously reported multiplier and divider circuits.

A perusal of the available literature on the implementation of analog multipliers and dividers using different active building blocks has revealed that apart from the circuits presented in [30], which use a single CDBA and four MOSFETs, no other configuration using a single active building block and four MOSFETs has been proposed previously, in which, both multiplication and division functions are available within the same circuit topology in VM.

Table 2 Aspect ratios of MOSFETs as shown in Fig. 2

MOSFETs	Aspect ratio (W/L)
M_1–M_5, M_8–M_9, M_{17}–M_{20}, M_{23}–M_{24}	54/0.72
M_{11}, M_{26}	270/0.72
M_6, M_{21}	135/0.72
M_7, M_{10}, M_{12}–M_{16}, M_{22}, M_{25}, M_{27}–M_{30}	18/0.72

Therefore, the main aim of this communication is to present a new circuit configuration of VM multiplier and divider circuit that requires only single CFOA and four matched NMOS transistors operating in the triode region. Mismatch analysis, Monte-Carlo analysis, PVT analysis and noise analysis have been performed to validate the workability of the proposed multiplier and divider circuit using CMOS CFOA implemented in 0.18 μm TSMC CMOS technology parameters.

2 Proposed analog multiplier/divider circuit

The proposed multiplier/divider circuit is shown in Fig. 1.

CFOA is a four terminal active device which is characterized by the following terminal equations:

$i_y = 0,\ v_x = v_y,\ i_z = i_x,$ and $v_w = v_z.$

The drain current of MOSFET, in the triode region ($V_{DS} \leq (V_{GS} - V_{TH})$; $V_{GS} > V_{TH}$), neglecting channel length modulation is expressed as [43]:

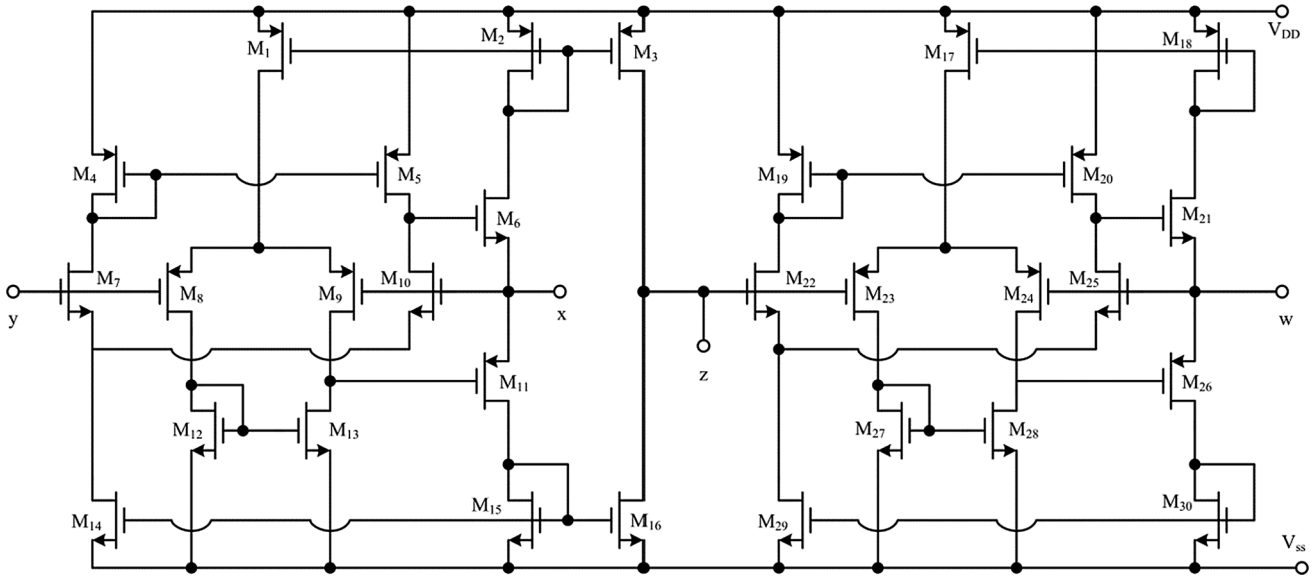
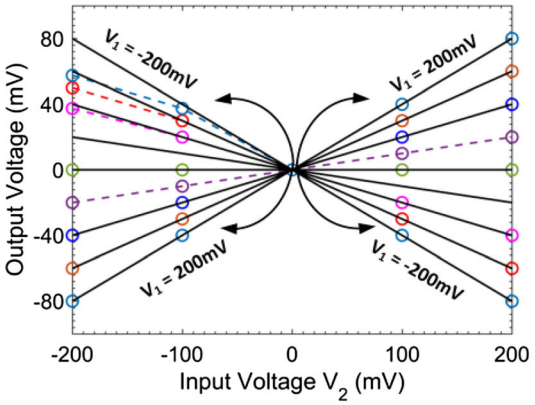
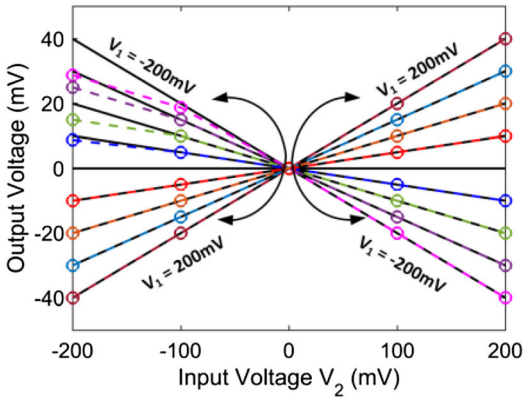


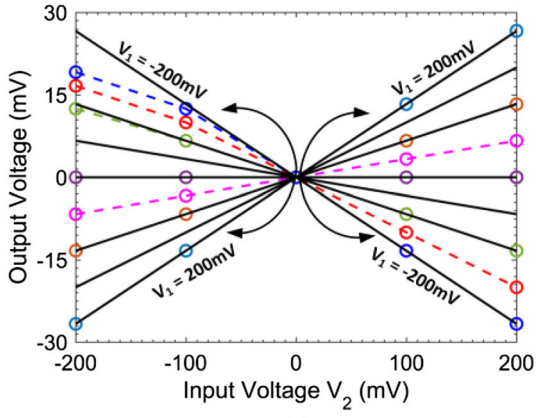
Fig. 2 CMOS implementation of CFOA [44]



(a)



(b)



(c)

Fig. 3 DC analysis curves of V_{out} and V_2 with V_1 as a variable parameter

$$I_D = K_n\left\{(V_{GS} - V_{TH})V_{DS} - \frac{1}{2}V_{DS}^2\right\} \tag{1}$$

where $K_n = \mu_n C_{ox}\left(\frac{W}{L}\right)$ is the MOSFET transconductance parameter. V_{TH} is the threshold voltage, V_{GS} and V_{DS} are the gate-to-source and drain-to-source voltages of the different MOSFETs.

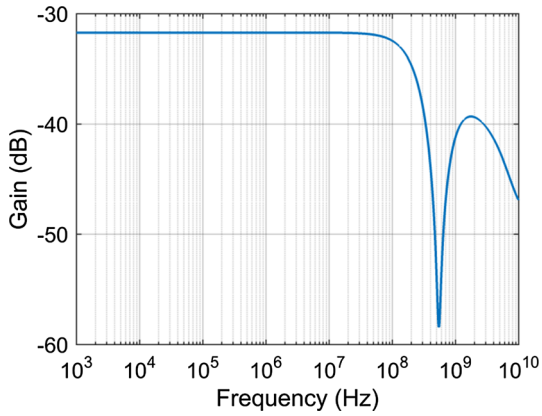


Fig. 4 Frequency response of proposed multiplier circuit

Assuming ideal CFOA, a routine analysis of the circuit shown in Fig. 1, yields the following expression relating the output voltage and the input voltage:

$$V_{out} = \frac{V_1 V_2}{V_3 + 2(V_P - V_{TH})} \tag{2}$$

If we select $V_P = V_{TH}$, Eq. (2) reduces to

$$V_{out} = \frac{V_1 V_2}{V_3} \tag{3}$$

where V_1, V_2, V_3 are input voltage signals.

From Eq. (3), it may be noted that the output voltage of the proposed circuit of Fig. 1 is directly proportional to the multiplication of input voltages V_1 and V_2, and is inversely proportional to V_3. The circuit, thus, can be used in (1) multiplier mode, if the voltage V_3 is kept constant and (2) divider mode, if either V_1 or V_2 is constant and the signals to be used in division being taken as V_2 or V_1 and V_3, without changing the structure of the circuit. It may also be noted from Eq. (3) that the output voltage of the proposed multiplier/divider topology is independent from transconductance parameters of MOSFETs.

3 Mismatch analysis between threshold voltages of MOSFETs

Because of variation in the process parameters, the threshold voltages of the four NMOS transistors may not be exactly same and the output voltage of the circuit may be different from its ideal value. We have evaluated the effects of the mismatch in the threshold voltages analytically. Assuming different threshold voltages V_{TH1}, V_{TH2}, V_{TH3} and V_{TH4} for MOSFETs M_{N1}, M_{N2}, M_{N3} and M_{N4} respectively, the output voltage of multiplier and divider circuits may be expressed as:

$$V_{out} = \frac{V_1 V_2 - V_1(V_{TH2} - V_{TH1})}{V_3 + 2V_P - (V_{TH3} + V_{TH4})} \tag{4}$$

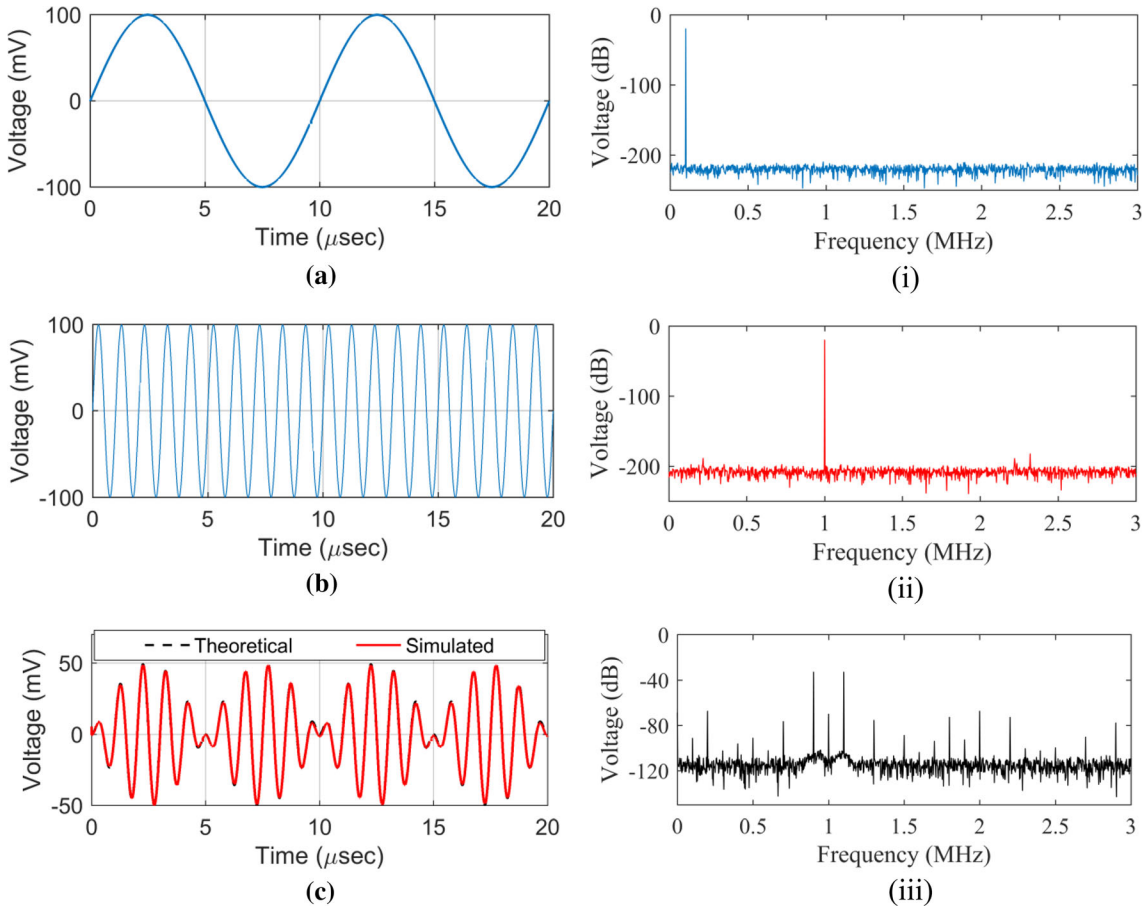


Fig. 5 Transient input–output voltages of modulator circuit and their respective frequency spectrums

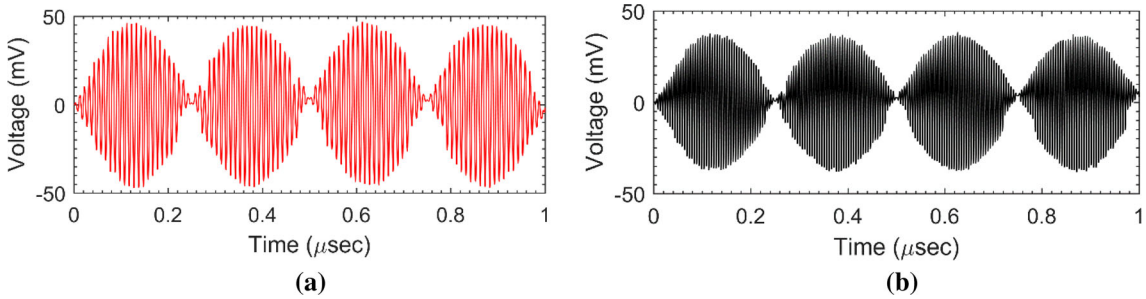


Fig. 6 Simulated transient output voltage of modulator circuit at **a** 100 MHz and **b** 200 MHz

From Eq. (4), it is observed that any mismatch between the threshold voltages of M_{N1} and M_{N2} will be reflected as DC offset in the output. However, the effect of any mismatch between the threshold voltages of M_{N3} and M_{N4} on the output may be eliminated altogether by selecting appropriate value of V_P.

4 Simulation results

To verify the theoretical propositions of the proposed multiplier/divider topology, PSPICE simulations have been carried out. An exemplary CFOA implemented using 0.18 μm TSMC technology parameters as shown in Fig. 2 has been used for the verification of proposed multiplier/divider topology. The aspect ratios used for various MOSFETs in Fig. 2 have been tabulated in Table 2. The power supply voltages $V_{DD} = -V_{SS}$ were taken as 1.25 V.

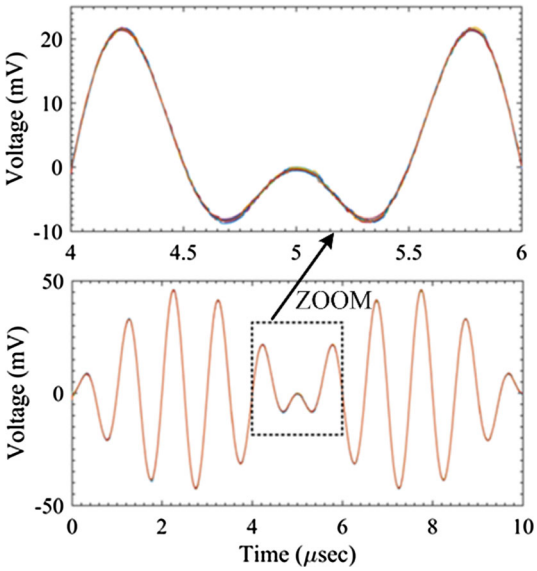
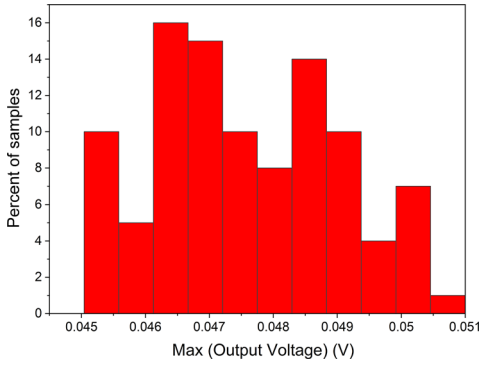


Fig. 7 Output voltage of amplitude modulator with variation in aspect ratios of MOSFETs



n samples	=	100	10th %ile	=	0.046238
n divisions	=	10	median	=	0.0477604
mean	=	0.0475368	90th %ile	=	0.086631
sigma	=	0.000938927	maximum	=	0.0489792
minimum	=	0.0453873	3*sigma	=	0.00281678

Fig. 8 Histogram of the amplitude modulator

The aspect ratios of all external MOSFETs used were taken as 3.6 µm/3.6 µm. The discrete MOSFETs used in the realization of multiplier/divider (M_{N1}, M_{N2}, M_{N3} and M_{N4}) were operating in triode region and the MOSFETs used for floating voltage source (M_{F1}, M_{F2}, M_{F3}, M_{F4}) were operating in saturation region.

4.1 Voltage multiplication mode

The DC characteristics of the circuit are illustrated in Fig. 3, wherein, the variation of the output voltage with V_2, when the input voltage V_1 was varied from −200 mV to + 200 mV in steps of 50 mV for different values of V_3 (a) 500 mV, (b) 1 V and (c) 1.5 V have been shown.

From the above simulated output voltages, it may be observed that the error between theoretical and simulated output voltage is more when the input voltage V_1 is less than -100 mV for different values of input voltage V_2. It may be pointed out that the MOSFETs M_{N3} and M_{N4} must operate in the triode region for the implementation of both the multiplier as well as divider circuit. The maximum output voltage V_{out} will thus be limited by the input voltage V_3 as given below.

$$V_{out} \leq (V_3 + V_G - V_{TH}) \tag{5}$$

Figure 4 shows the AC responses of the multiplier circuit and the −3 dB frequency was found to be 200 MHz when input signal were applied to V_1, V_2 and V_3 was kept constant at 100 mV.

4.1.1 Application examples of multiplier circuit

4.1.1.1 Amplitude modulator For the application of multiplier circuit as amplitude modulator, the input voltage (V_1) was taken as a sinusoidal wave of 100 kHz with peak to peak amplitude of 200 mV while the voltage (V_2) was taken as sinusoidal of 1 MHz with a peak to peak voltage of 200 mV as shown in Fig. 5(a) and (b) respectively. A DC input voltage applied as V_3 with amplitude of 200 mV. The simulated output voltage of modulator circuit has been shown in Fig. 5(c). The frequency spectrums (in dB) of such inputs and output are also displayed in Fig. 5.

In Fig. 6, we have shown the sample modulator output voltages when the frequency of V_2 signal was (a) 100 MHz and (b) 200 MHz while the frequency of V_1 was kept 2 MHz.

The performance of multiplier circuit has also been evaluated with variations in the aspect ratios of external MOSFETs (M_{N1}, M_{N2}, M_{N3} and M_{N4}). Figure 7 shows the output voltage of modulator when width of the MOSFETs have deviation of ± 10%.

To test the robustness of the multiplier circuit, we have done Monte-Carlo simulations by varying the input voltage V_2 with a nominal tolerance of 10% (considering Gaussian distribution) for 100 runs and the histogram of the Monte-Carlo simulations has been shown in Fig. 8. From Fig. 8, it may be noted that the maximum output voltage of the multiplier circuit is 48.97 mV after 100 iterations which is close to the theoretical value (50 mV).

To evaluate the prefabrication performances of multiplier circuit such as process-voltage-temperature (PVT) analyses have also been carried out. The effect of different process corners i.e., slow-slow (SS), slow fast (SF), fast-slow (FS), fast–fast (FF), and typical-typical (TT) on the output of modulator circuit have been shown in Fig. 9(a), whereas in Fig. 9(b), the output voltage of modulator has been displayed with the presence of ± 10% variations in bias

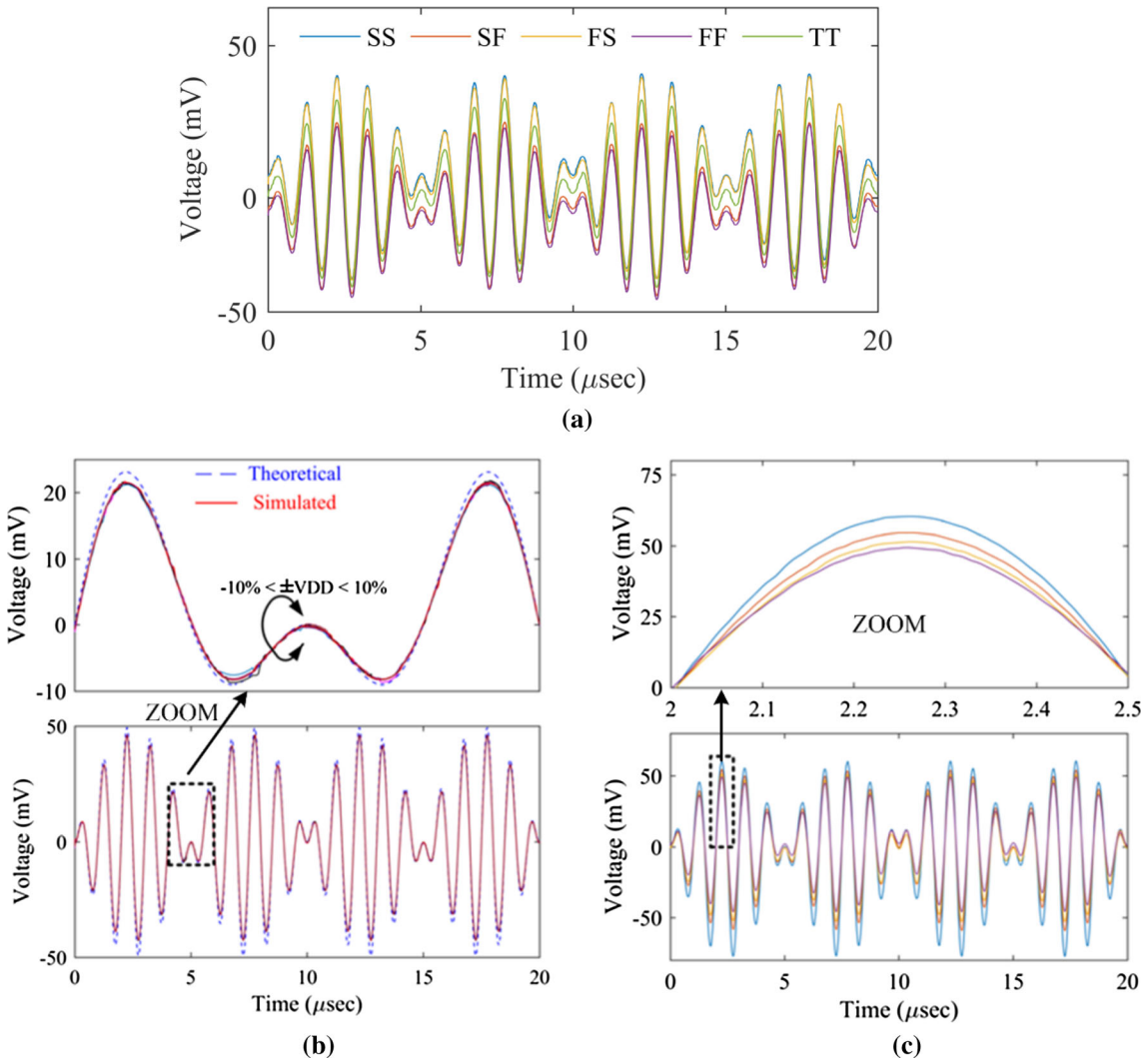


Fig. 9 Output voltage of amplitude modulator corresponding to **a** process variations **b** supply voltage variations and **c** temperature variations

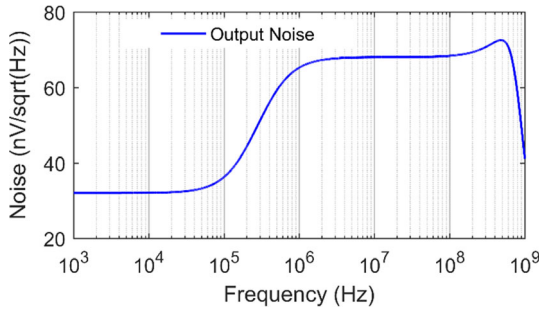


Fig. 10 Output noise of amplitude modulator

voltage. The effect of temperature on the output voltage of the amplitude modulator has also been evaluated at different temperatures (−10 °C, 10 °C, 20 °C, 30 °C) and the simulated results have been displayed in Fig. 9(c). The effect of process corners, variations in bias voltage and change in temperature on the output of proposed multiplier circuit has been found to be very small.

We have also evaluated the noise performance of the modulator circuit using PSPICE (V_1 -100 mV, 100 MHz, V_2—100 mV, 2 MHz) and the result is shown in Fig. 10. The maximum output noise for the selected frequency is found to be less than 70nV/sqrt(Hz).

4.1.1.2 Frequency doubler and squarer function The circuit of Fig. 1 can also be configured as frequency doubler if input voltages $V_1 = V_2 = V_{in}$ and V_3 is taken as a constant voltage. With same inputs applied to the Fig. 1, the output obtained from such circuit is square of the input signal. To verify the results of frequency doubler and squarer circuits, the input voltage V_{in} was taken as a sinusoidal voltage of 1 MHz with peak to peak amplitude of 200 mV and V_3 was also set to 200 mV DC. The transients of input and output voltages have been displayed in Fig. 11(a) and the

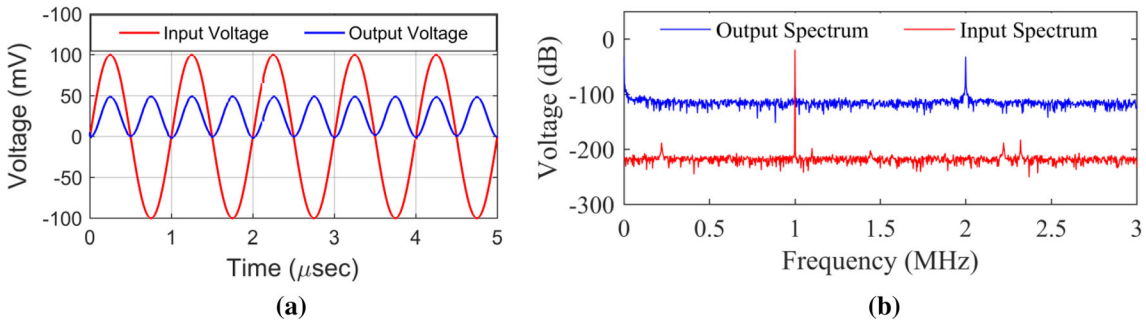


Fig. 11 Transients of input and output voltages of squarer circuit and its frequency spectrums

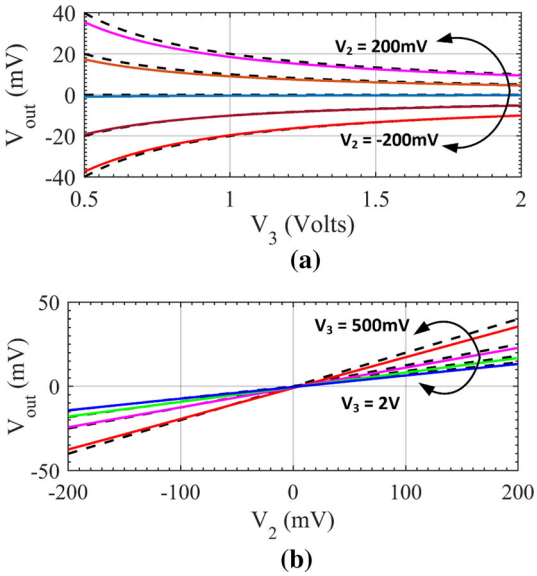


Fig. 12 DC responses of analog divider circuit **a** when V_2 is varied **b** when V_3 is varied

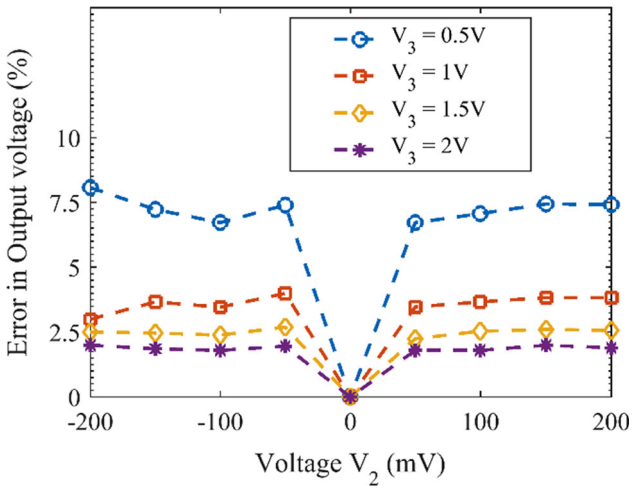


Fig. 13 Error in output voltage of divider circuit for different values of V_2 and V_3

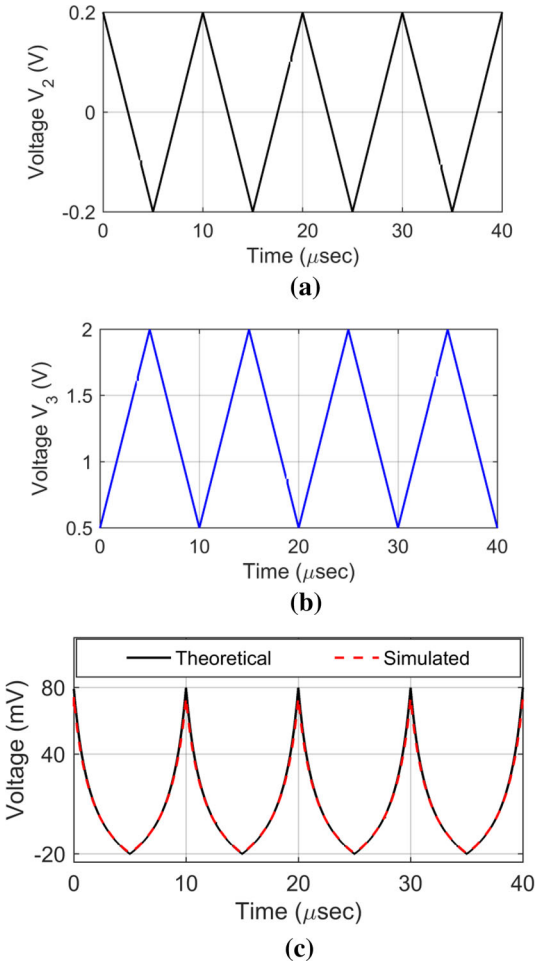


Fig. 14 Time responses of the circuit in division mode

frequency spectrums have been depicted in Fig. 11(b). As shown in the frequency spectrum, the frequency of output voltage is double that of the input frequency.

4.2 Voltage divider mode

The transfer characteristics for the voltage division mode of operation were plotted by keeping the input voltage V_1 as constant at 100 mV DC and determining the output voltage

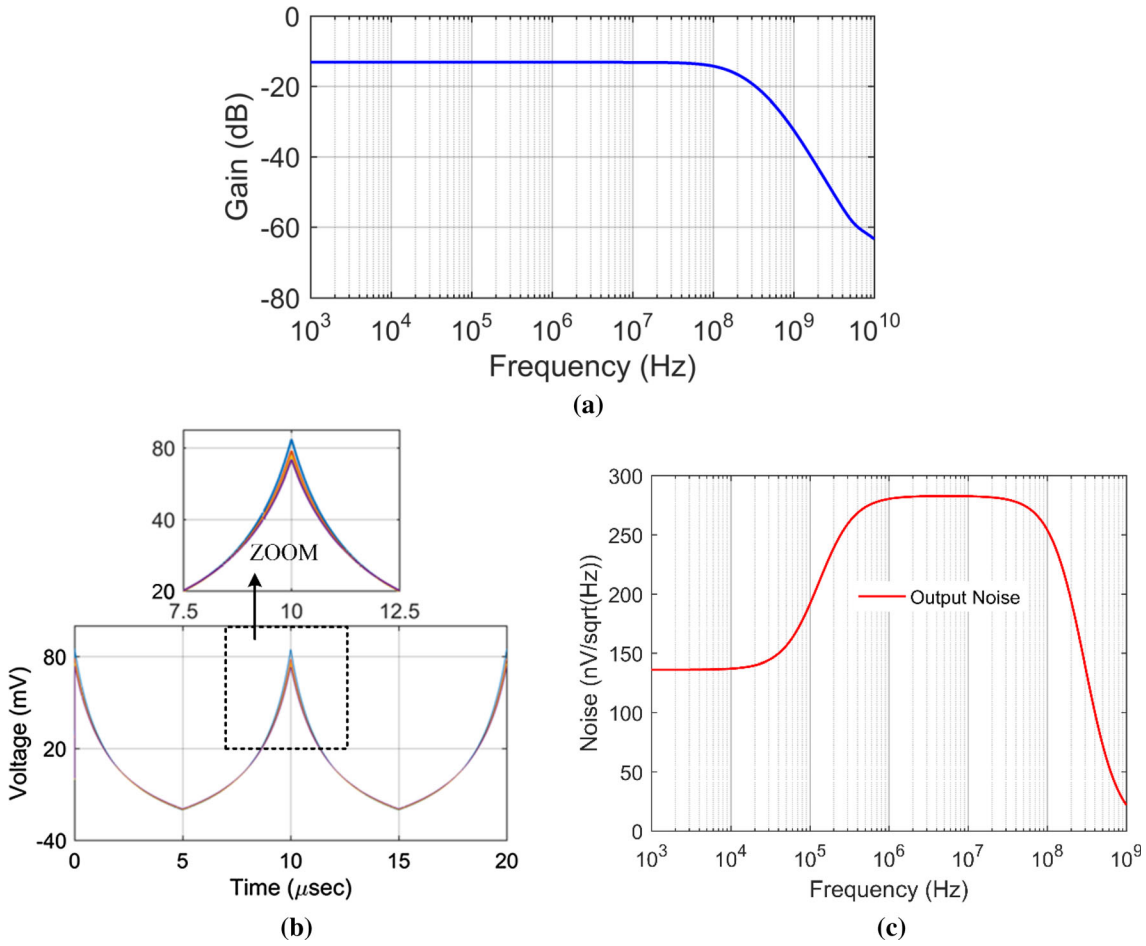


Fig. 15 **a** Frequency responses of proposed divider circuit **b** Variation in temperature on divider circuit **c** Output noise of analog divider circuit

variation with V_2 and V_3. Figure 12(a) shows the output when the input voltage V_2 was varied from −200 mV to + 200 mV in steps of 50 mV, whereas the variation of output voltage with V_2 is shown in Fig. 12(b) when the input voltage V_3 is varied from 0.5 V to 2 V (dashed line-Theoretical, coloured line- Simulated).

We have also computed the error in output voltage, when V_2 was varied from −200 mV to 200 mV and V_3 was varied between 0.5 V to 2 V. Figure 13 illustrates the error in output voltage of divider circuit. From the error graph, it may be observed that the percentage error in output voltage is quite low when V_3 increases gradually.

The time response of the divider circuit was tested by applying a triangular voltage of $400mV_{PP}$ (V_2) and $1.5V_{PP}$ (V_3) at a frequency of 100 kHz to both the inputs. The voltage V_1 was kept constant at 200 mV (DC). The inputs and output of divider circuit have been illustrated in Fig. 14.

The AC response of the proposed circuit for division mode of operation has been illustrated in Fig. 15(a). The −3 dB bandwidth of the divider circuit was found to be 182 MHz approximately. The effect of temperature on divider circuit has also been evaluated for different temperatures between −10 to 30 °C and depicted in Fig. 15(b). From Fig. 15(b), it may be noted that the effect of temperature on divider circuit is quite low. We have also carried out the noise analysis using PSPICE for analog divider circuit and the simulation results have been shown in Fig. 15(c). The maximum output noise measured for divider is 280nV/sqrt(Hz).

The above results, thus, verify the usability of the proposed circuit.

5 Conclusions

A simple architecture of VM analog multiplier/divider circuit employing a single CFOA has been presented. The circuit can function as four quadrant analog multiplier and two quadrant analog divider depending on the specified inputs, without any additional change. In contrast to the previously reported non-linear function generation circuits realized with CFOAs which employ more number of components, the circuit uses only one CFOA and four MOSFETs

(biased in triode region) for the realization of multiplier/divider function. Several applications of multiplier such as squarer, amplitude modulator and frequency doubler have also been included. PVT analysis, Monte-Carlo simulations and noise analysis have also been carried out to evaluate the robustness of the proposed configuration. The PSPICE simulations, using CMOS CFOA implemented with 0.18 μm TSMC technology parameters, have been used to verify the workability of the proposed circuits.

Acknowledgements The authors wish to thank the anonymous reviewers for their constructive comments and suggestions in improving the quality of the manuscript. This work has been performed at Analog Signal Processing Laboratory of Electrical Engineering Department, Delhi Technological University, New Delhi-110042, India.

Data availability All data in this study can be obtained by simulation results.

Declarations

Conflict of interest The authors declare that they have no conflict of interest regarding the publication of this article. 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. Doebelin, O. E. (2004). *Measurement systems: Application and design*. New York: McGraw Hill.
2. Saxena, N., & Clark, J. J. (1994). A four-quadrant CMOS analog multiplier for analog neural networks. *IEEE journal of solid-state circuits, 29*(6), 746–749. https://doi.org/10.1109/4.293124
3. Parsnejad, S., Hu, Y., Wan, H., Ashoori, E., & Mason, A. J. (2016). Wide dynamic range multi-channel electrochemical instrument for in-field measurements. *IEEE Sensors*. https://doi.org/10.1109/ICSENS.2016.7808670
4. Tadic, N., & Gobovic, D. (2007). Smart sensor interfacing circuit using square-rooting current-to-frequency conversion. *International Journal of Electronics, 94*(12), 1075–1098. https://doi.org/10.1080/00207210701799500
5. Roy, S., Paul, T. K., & Pal, R. R. (2017). A new method of realization of four-quadrant analog multiplier using operational amplifiers and MOSFETs. *Journal of Physical Sciences, 22*, 163–173.
6. Riewruja, V., & Rerkratn, A. (2011). Four-quadrant analogue multiplier using operational amplifier. *International Journal of Electronics, 98*(4), 459–474. https://doi.org/10.1080/00207217.2010.520155
7. Khachab, N. I., & Ismail, M. (1989). MOS multiplier/divider cell for analogue VLSI. *Electronics Letters, 25*(23), 1550–1552.
8. Khachab, N. I., & Ismail, M. (1991). A nonlinear CMOS analog cell for VLSI signal and information processing. *IEEE Journal of Solid-State Circuits, 26*(11), 1689–1699.
9. Surakampontorn, W., Khanittha, K., Chalermpan, F. (2002). A Simple Current-Mode Analog Multiplier-Divider Circuit Using OTAs, In Proceedings of the IEEK Conference, pp. 658–661.
10. Kaewdang, K., Fongsamut, C., Surakampontorn, W. (2003). A wide band current-mode OTA based analog multiplier-divider, In Proceedings of the 2003 international symposium on circuits and systems, 349–352, https://doi.org/10.1109/ISCAS.2003.1205572.
11. Sinencio, E. S., Angulo, J. R., Barranco, B. L., & Vazquez, A. R. (1989). Operational transconductance amplifier-based nonlinear function syntheses. *IEEE Journal of Solid-State Circuits, 24*(6), 1576–1586. https://doi.org/10.1109/4.44993
12. Bhanja, M., & Ray, B. N. (2016). OTA-based logarithmic circuit for arbitrary input signal and its application. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 24*(2), 638–649. https://doi.org/10.1109/TVLSI.2015.2406953
13. Raj, A., Bhaskar, D. R., & Kumar, P. (2020). Two quadrant analog voltage divider and square root circuits using OTA and MOSFETs. *Circuits Systems and Signal Processing, 39*, 6358–6385. https://doi.org/10.1007/s00034-020-01454-2
14. Liu, S.-I., Wu, D. S., Tsao, H. W., Wu, J., & Tsay, J. H. (1993). Nonlinear circuit applications with current conveyors. *IEE Proceedings of Circuits, Devices and Systems, 140*(1), 1–6. https://doi.org/10.1049/ip-g-2.1993.0001
15. Hwang, Y. S., Liu, W. H., Tu, S. H., & Chen, J. J. (2009). New building block: Multiplication-mode current conveyor. *IET Circuits, Devices & Systems, 3*(1), 41–48. https://doi.org/10.1049/iet-cds:20080156
16. Abuelma'Atti, M. T., & Al-Qahtani, M. L. (1998). A current-mode current-controlled current conveyor-based analogue multiplier/divider. *International Journal of Electronics, 85*(1), 71–77. https://doi.org/10.1080/002072198134364
17. Petchakit, W., Kiranon, W., Wardkien, P., Petchakit, S. (2010). A current-mode CCCII-based analog multiplier/divider. In ECTI-CON2010: The 2010 ECTI International Confernce on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 221–224, IEEE.
18. Arora, T. S., & Sharma, R. K. (2016). A novel cubic generator realised by CCIII-based four quadrant analog multiplier and divider. *Indian Journal of Science and Technology, 9*(38), 1–8. https://doi.org/10.17485/ijst/2016/v9i38/100026
19. Narksarp, W, Pawarangkoon, P., Kiranon, W., Wadkien, P. (2009). A four-quadrant current-mode multiplier/divider building block. In 2009 6th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, pp. 574–577, IEEE, https://doi.org/10.1109/ECTICON.2009.5137072.
20. Tangsrirat, W., Pukkalanun, T., Mongkolwai, P., & Surakampontorn, W. (2011). Simple current-mode analog multiplier, divider, square-rooter and squarer based on CDTAs. *International Journal of Electronics and Communications (AEU), 65*(3), 198–203. https://doi.org/10.1016/j.aeue.2010.02.017
21. Jaikla, W., Siripruchyanun, M. (2007). A novel current-mode multiplier/divider employing only single dual-output current controlled CDTA. In 2007 International Symposium on Communications and Information Technologies, 106–109, IEEE, https://doi.org/10.1109/ISCIT.2007.4391994.
22. Siripruchyanun, M., & Jakila, W. (2008). A current-mode analog multiplier/divider based on CCCDTA. *International Journal of Electronics and Communications (AEU), 62*(3), 223–227. https://doi.org/10.1016/j.aeue.2007.03.009
23. Pisutthipong, N., Siripruchyanun, M. (2009). A novel simple current-mode multiplier/divider employing only single multiple-output current controlled CTTA. In TENCON 2009 IEEE Region 10 Conference, 1–4, IEEE, https://doi.org/10.1109/TENCON.2009.5395877.
24. Mongkolwai, P., Tangsrirat, W. (2011). CFTA-based current multiplier/divider circuit. In 2011 International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS), 1–4, IEEE, https://doi.org/10.1109/ISPACS.2011.6146074.

25. Chadha, U. and Arora, T. (2016). Four quadrant analog multiplier/divider employing single OTRA. Proceedings of the International Conference on Communication and Computing Systems, pp. 635–639, https://doi.org/10.1201/9781315364094-115.
26. Pandey, R., Pandey, N., Sriram, B., & Paul, S. K. (2012). Single OTRA based analog multiplier and its applications. *International Scholarly Research Notices*. https://doi.org/10.5402/2012/890615
27. Nagar, B. C., & Paul, S. K. (2018). Single OTRA based two quadrant analog voltage divider. *Analog Integrated Circuits and Signal Processing, 94*(1), 161–169. https://doi.org/10.1007/s10470-017-1085-1
28. Kumngern, M., Torteanchai, U. (2013). A CMOS current-mode multiplier/divider using a current amplifier. In 2013 IEEE 7th International Power Engineering and Optimization Conference (PEOCO), 742–745, IEEE, https://doi.org/10.1109/PEOCO.2013.6564645.
29. Pathak, J. K., Singh, A. K., & Senani, R. (2014). New multiplier/divider using a single CDBA. *American Journal of Electrical and Electronic Engineering, 2*(3), 98–102.
30. Roy, S., Paul, T. K., Maiti, S., & Pal, R. R. (2018). Two new analog multipliers/dividers employing single current differencing buffer amplifier. *International Journal of Electronics and Communications (AEU), 88*, 11–19. https://doi.org/10.1016/j.aeue.2018.03.002
31. Siripruchyanun, M. (2007). A design of analog multiplier and divider using current controlled current differencing buffered amplifiers. In Proceedings of the international symposium on integrated circuits (ISIC'07), pp. 568–571, https://doi.org/10.1109/ISICIR.2007.4441925.
32. Rajpoot, J., & Maheshwari, S. (2020). High performance four-quadrant analog multiplier using DXCCII. *Circuits Syst Signal Process, 39*, 54–64. https://doi.org/10.1007/s00034-019-01179-x
33. Kumar, A. (2020). Single EXCCII-based analog multiplier structure. *Journal of Circuits, Systems and Computers*. https://doi.org/10.1142/S0218126621501073
34. Petrović, P. B. (2020). New current-mode RMS-to-DC converters and four-quadrant multiplier/divider based on VDTA. *IET Circuits, Devices & Systems, 14*(4), 490–497. https://doi.org/10.1049/iet-cds.2019.0373
35. Lopez-Martin, A. J., Carlos, A., Ramirez-Angulo, J., Carvajal, R.G. (2010). Compact low-voltage CMOS current-mode multiplier/divider. In Proceedings of 2010 IEEE International Symposium on Circuits and Systems, pp. 1583–1586, IEEE, https://doi.org/10.1109/ISCAS.2010.5537427.
36. Wasaki, H., Horio, Y., & Nakamura, S. (1991). Current multiplier/divider circuit. *Electronics Letters, 27*(6), 504–506. https://doi.org/10.1049/el.19910317
37. Liu, W., Liu, S.-I., & Wei, S. K. (2005). CMOS current-mode divider and its applications. *IEEE Transactions on Circuits and System-II: Express Briefs, 52*(3), 145–148. https://doi.org/10.1109/TCSII.2004.842041
38. Al-Absi, M. A., Hussein, A., & Abulma'atti, M. T. (2013). A low voltage and low power current-mode analog computational circuit. *Circuits, Systems, and Signal Processing, 32*(1), 321–331. https://doi.org/10.1007/s00034-012-9446-6
39. Wang, Z. (1991). A CMOS four-quadrant analog multiplier with single-ended voltage output and improved temperature performance. *IEEE Journal of Solid-State Circuits, 26*(9), 1293–1301. https://doi.org/10.1109/4.84947
40. Senani, R., Bhaskar, D. R., Singh, A. K., & Singh, V. K. (2013). *Current feedback operational amplifiers and their applications*. New York: Springer.
41. Liu, S.-I., & Chen, J. J. (1995). Realisation of analogue divider using current feedback amplifiers. *IEE Proceedings-Circuits, Devices and Systems, 142*(1), 45–48. https://doi.org/10.1049/ip-cds:19951609
42. Bajaj, N., Govil, J. (2006). Realisation of a differential multiplier-divider based on current feedback amplifiers. In 2006 IEEE International Conference on Semiconductor Electronics, 708–712, IEEE, https://doi.org/10.1109/SMELEC.2006.380728.
43. Allen, P. E., & Holberg, D. R. (1987). *CMOS analog circuit design*. New York: Oxford University Press.
44. Dogan, M., & Yuce, E. (2019). CFOA based a new grounded inductor simulator and its applications. *Microelectronics Journal, 90*, 297–305. https://doi.org/10.1016/j.mejo.2019.07.002

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

Ajishek Raj received B. Tech degree in Electrical and Electronics Engineering from Dr. M. G. R. Educational and Research Institute, Chennai in 2012, M. Tech degree from Delhi Technological University, Delhi in 2016. Currently, he is pursuing for Doctoral studies in the Department of Electrical Engineering at Delhi Technological University, Delhi. His research interests are in the area of analog signal processing, signal generation, active nonlinear applications, and inverse analog filters.

D. R. Bhaskar received B.Sc. degree from Agra University, B. Tech. degree from Indian Institute of Technology (IIT) Kanpur, M. Tech. from IIT Delhi and Ph.D. from University of Delhi. Prof. Bhaskar held the positions of Lecturer (1984–1990) and Senior Lecturer (1990–1995) at the Electrical Engineering Department of Delhi College of Engineering (now Delhi Technological University). He joined the Electronics and Communication Engineering (ECE) Department of Jamia Millia Islamia in July1995, as a Reader and became a Professor in January 2002. He served as the Head of the Department of ECE from 2002 to 2005. Presently he is working in the Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India. His teaching and research interests are in the areas of Bipolar and CMOS Analog Integrated Circuits and Systems, Current Mode Signal Processing, Communication Systems and Electronic Instrumentation. Prof. Bhaskar has authored or co-authored 103 research papers—all in international journals of repute. He has co-authored 4 monographs. He has acted/has been acting as a Reviewer for several journals of IEEE, IEE and other international journals of repute.

Pragati Kumar Professor in Department of Electrical Engineering, Delhi Technological University, New Delhi, India was born in Muzaffarpur, India. He received his Ph.D. from The University of Delhi in 2008. His teaching and research interests are in the area of Network Analysis and Synthesis, Microelectronics, CMOS Analog Integrated Circuits and Fractional Order Circuits/Systems. He has authored/co-authored 33 research papers in various international journal/conferences/edited book volumes.

NOVEL APPROACH TO GROUNDWATER CONTAMINANT TRANSPORT MODELLING

Manoj Kumar[1], Anunay Gour[2]

[1]Corresponding Author, Department of environmental engineering, Delhi Technological University, New Delhi, E-mail: m.kumar62.mk@gmail.com

[2]Department of environmental engineering, Delhi Technological University, New Delhi, E-mail: anunaygour@dtu.ac.in

Abstract: This paper comprehensively investigates literature about modelling techniques used in groundwater contaminant transport modelling. Modelling of groundwater is a useful way for the management of groundwater resources, also assessing the fate of contaminants and their remediation. Models very conveniently help to study complex real conditions and examine specific phenomena in addition to predicting the future behaviour of any problem. The use of groundwater simulation programming tools such as MODFLOW, MT3DMS, RT3D, FEFLOW, and MODPATH to model multi-directional contamination transport yields accurate results. Movement, storage, and change of solute concentration are largely regulated by groundwater flow gradient. As a result, a precise description of the flow mechanism is very important. If models are not properly constructed and interpreted, they can become complicated and may generate wide errors. Well-defined and clear modelling objectives produce suitable models for efficient error-free modelling processes. The study will assist modellers to clearly define their model objective and select appropriate modelling tools.

Keywords: *groundwater, modelling, contaminant transportation, MODFLOW, FEFLOW, MT3DMS, RT3D, MODPATH*

1. Introduction

Groundwater modelling is a technique to understand a system's response under certain conditions or to anticipate the system's future behaviour by representing it in another form. Groundwater modelling is an effective tool for groundwater resources management and protection. A logical relation between the info availability and model complexity should be considered for better presentation. When there's a shortage of data, selecting complex systems wouldn't end in desirable results and wouldn't show the real scenario. In such situations, we should review our expectations of the model's ability as per the data availability and accessibility. Representation of the real-world problem in simple form without compromising the accuracy is a challenge for the modellers. To get the best description of the real world situations modellers tend to get the maximum data possible and the selection of the right model tool plays a vital role in the best representation of reality also. Classification of groundwater models can be done as a physical model, analogue model, and mathematical model. Physical models are the laboratory portrayal of the groundwater system such as the Sand tank which poses a scaling problem. Analogue models are the electronics models developed by the analogy between water flow systems and electrical flow systems (Delleur, 2006). For example, “Electrical current flow through a circuit board with resistors to represent hydraulic conductivity and capacitors as storage coefficient”. Representation of the groundwater system mathematically by a set of equations is known as a mathematical model. The mathematical models can be solved either numerically or analytically. Through the ages, analytical methods have become standard tools for explaining groundwater quantity and quality dynamics. Also, these models are occasionally applied before a numerical model starts. These models have their disadvantages concerning mathematical assumptions that may not occur naturally. Another problem is that they do not allow for simulating spatial and temporal variations in various parameters simultaneously. Hence, these models are often inadequate to mimic a heterogeneous subsurface system closely. “Numerical solutions can handle more complicated problems than analytical solutions. With the rapid development of computer processors and

increasing speed, numerical modelling has become more effective and manageable"(Baalousha, 2005). A detailed explanation of numerical techniques is provided in the next chapter.

2. Numerical Models

The finite difference method and the finite element method are the most commonly used approaches to solve the governing groundwater flow or solute transport equations numerically" (Wang et al., 2019). The modelling approach to be selected depends upon the modelling objectives and concerning problems. "The finite difference method can produce different results than the finite element method if the concern is complicated"(Baalousha, 2005). The modelling approach isn't the solitary factor that influences the outcomes of the model. Different elements such as initial conditions, boundary conditions, space discretization, and nature of information impact the outcomes. With numerical modelling, hydrogeologists could approximate the solutions of complicated differential equations of the system by converting them to discrete equations and dissecting the domain into meshes or grids. In numerical models, the dissection of a region and the discretization of differential equations are generally accomplished using approximation approaches like finite difference, finite element, and finite volume boundary. Each of these approaches can be applied solely or combined with other methods to reduce numerical models' computational complexity. Considering the applicability, ease of use, performance, and computational speed, each method has its limitations and advantages.

2.1 Finite Difference Method

"The finite difference method (FDM) has been widely used in groundwater studies since the early 1960s. FDM was studied by Newton, Gauss, Bessel, and Laplace"(Pinder and Gray, 1978).
The finite-difference approximation is given by:

$$\frac{dy}{dx} = \lim_{\Delta x \to \infty} \left(\frac{\Delta y}{\Delta x} \approx \frac{\Delta y}{\Delta x} \right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textit{Eq.2.1}$$

The finite difference method relies upon the calculation of a "function derivative by a finite-difference". In this method, the area of concern is divided into small cells of regular shape or grid points called nodes, the derivatives of the partial differential equations are approximated in space, concerning the differences between values of variables at adjacent cells or nodes. The nodes represent the specific values of the system's computational properties (e.g., hydraulic head, solute concentration, effective porosity) as functions of time. All nodes are defined in a body-fitted coordinate system and labelled consecutively by a set of indices (i, j, and k along the direction x, y, and z respectively). Discretization of the area in the finite difference method within a model domain is presented in figure 2.1. FDM is well documented easy to implement and gives reasonably good results. The main limitation of the finite difference method is that we can not apply it appropriately to the irregular shape boundaries as the area of concern is divided into regular shape cells. Accuracy and the efforts required for computation are positively affected by the grid size and their distribution. "The output accuracy of the finite difference method is useful in the case of solute transport modelling. Mass balance is not guaranteed if conductivity or grid spacing varies"(Baalousha, 2005).

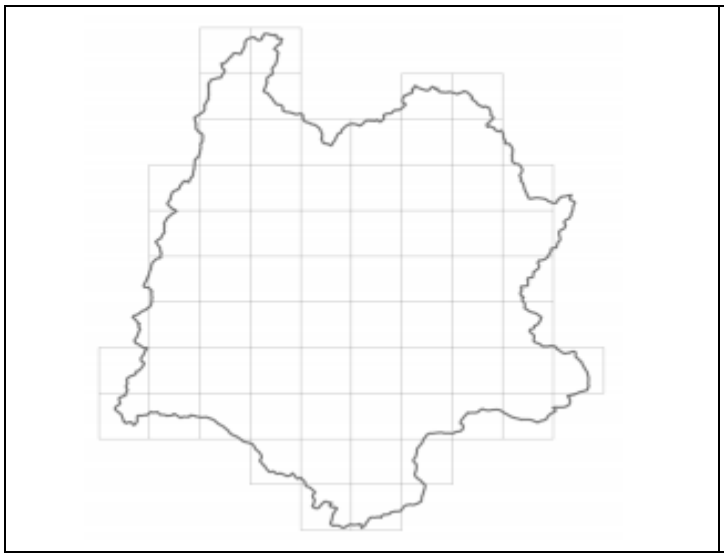


Fig2.1 Discretization of space in FDM(Elango, 2005)

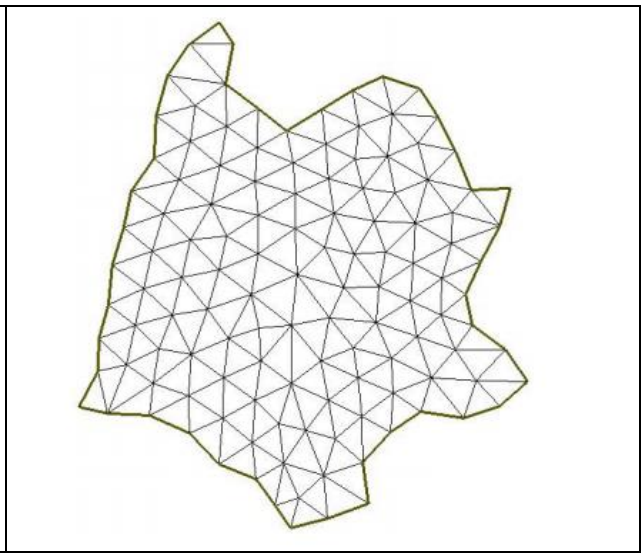


Fig 2.2 Discretization of space in FEM

2.2 Finite element method

FEM was initially used to solve problems related to civil engineering and construction. In this method, the grid is discretized into a network of nodes that form the vertices of irregularly/regularly-sized triangular elements. These elements and nodes are used to estimate the value of the field parameters like the solute concentration and hydraulic head. The elements may have different sizes, orientations and spatial dimensions. The mesh design is fundamental in the FEM because it greatly affects the solution's accuracy and convergence. The different approaches to the finite element method are Galerkin's method, weighted residuals, essential functions and variational principle (Pinder G., 1978). The key drawbacks of the method of finite elements over finite difference techniques are their complexity to formulate, difficulty to solve, and time to use. Conversely, to approximate a geometrically irregular mesh, these methods have more versatility than finite difference methods. However, to estimate an unconditionally irregular scheme, a large amount of calculation and storage memory for software simulation is required.“Advantages of this method include: a better mesh configuration, which suites irregular model boundaries, anisotropy is well incorporated, the governing system of equations is symmetric and irregular shapes can represent elements” (Baalousha, 2005). Discretization of the area in the finite element method within the domain of the model is shown in fig 2.2.

3. Modelling approach

Identification of model objectives is the first step in any modelling. A critical concern in the modelling process is data collection and processing. However, model conceptualization is the most fundamental step in modelling. After completion of the model and the first run, calibration, verification, and sensitivity analysis can be performed. The steps involved in groundwater modelling are shown in Figure 3.1.

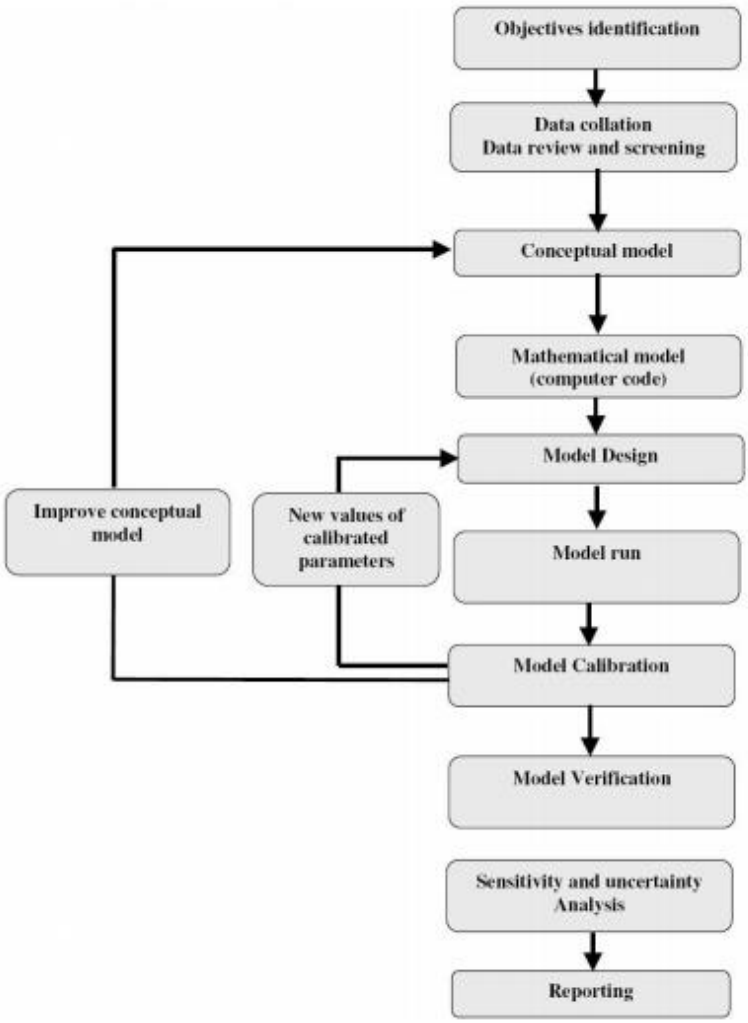


Fig 3.1 Modelling process flow chart(Baalousha, 2005).

3.1. Formulating modelling objectives

Objectives of modelling define the purpose of modelling. It guides the overall working of a model. Generally, management decisions on the quality or quantity of groundwater are based on the groundwater models. The model scope, methodology, and model type can differ depending on the modelling objectives. Groundwater models can be interpretive, predictive or generic. To study a specific case and to evaluate groundwater flow or the transport of pollutants, interpretive models are used. Predictive models, on the other hand, are used to know the difference in solute concentration or potential groundwater head. Different problems in the management of the water resources or remediation operations are studied using generic models.
Groundwater modelling objectives may be the measurement of groundwater flow and groundwater head, the investigation of the groundwater abstraction effect on the flow system at a well, the analysis of the effects on groundwater quality due to human activities (e.g. landfills, wastewater discharge, agricultural processes), or solute transport estimations. Modelling objective helps in model approach selection and collection of data needed.

3.2 Conceptual Model

Model conceptualization is a qualitative description of a ground-water system that includes the information of the water balance, hydrological and geological conditions. Model conceptualization is the most important part of groundwater modelling. After the identification of model objectives, it is the next step in the model development. Data about Geology, hydrology, boundary conditions and hydraulic parameters are needed to

develop a conceptual model. A good conceptual model must be able to represent real-world situations in a simple manner that meets modelling goals and management requirements. In the case of groundwater quality modelling, it must outline our interpretation of contaminant transport or water flow.
The key factors that the conceptual model should contain are:

1. Model domain area
2. Aquifer's geometry
3. Sources and sinks identification
4. Groundwater recharge
5. Boundary conditions
6. Aquifer properties like Porosity, Hydraulic conductivity, Storativity
7. Water balance

The mathematical model can be created after the completing model conceptualization. The mathematical model illustrates the assumptions made in the mathematical equations as well as the conceptual model. We can solve these equations either numerically or analytically.

3.2.1. Boundary Value Problem

The water balance theory underpins all mathematical models of groundwater flow. "In isotropic, homogeneous porous media, the general equation that governs three-dimensional groundwater steady-flow is

$$\frac{d^2h}{dx^2} + \frac{d^2h}{dy^2} + \frac{d^2h}{dz^2} = 0 \quad \text{.....eq. (3.1)}$$

where *h* is the groundwater head". This equation is also known as the Laplace equation. The governing equation for groundwater flow is created by combining the "mass balance equation and Darcy's Law". To get a unique solution to an equation, we need to understand the boundary conditions. As a result, this equation is referred to as a "boundary value problem". So, the conditions at boundaries define the field or domain in which the "boundary value problem" can be solved.

3.2.2. Boundary Conditions

The first step in model conceptualization is to identify conditions at the boundaries. To get a peculiar solution to partial differential groundwater flow equations, boundary conditions must be identified. Improper boundary condition detection has an impact on the solution and can lead to completely incorrect output.
Three main types of boundary conditions are as follow:

1. Specified flow (also known as type II boundary or Neumann). Mathematically, ∇h $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{T})$ = constant.
2. Specified head (also known as type I boundary or Dirichlet). Mathematically, it can be expressed as h $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{T})$ = constant.
3. Head dependent flow (also known as type III boundary or Cauchy). Mathematically it can be written as: ∇h $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{T})$+a*h = constant (where "a" is a constant). Natural boundaries should be identified carefully.

For example, Hydraulic boundaries, such as groundwater divides, may shift location as field conditions alter.

3.3. Selection of modelling software

The software for the modelling is selected after the hydrogeological characterization of the concerned area has been completed and the conceptual model has been created. The chosen model should have the potential to simulate the conditions at a given location. When field evidence shows that transport processes or

groundwater flow are comparatively straightforward, analytical models can be used. Similarly, we can make use of model conceptualization and hydrogeological characterization to pick one-dimensional, two-dimensional, or three-dimensional transport and groundwater flow models.

3.4. Model Design

Model designing includes all the input parameters that are needed to calibrate the model are included in the designing model. Degradation rate coefficients, dispersion coefficients transient or steady-state modelling, recharge, hydraulic conductivity/transmissivity, boundary conditions, layer elevations, model grid size and spacing, and any additional model data are all examples of input parameters.

3.5. Model Calibration

Calibration of the model is a method of fine-tuning the model results to fit the field measurements. Model results can vary from field measurements after the first run because modelling is merely a generalisation of the real world and estimations. So, errors in the computation are bound to occur. Model calibration is done by altering the model input parameter values to meet field conditions under certain limits. Field conditions at a site must be correctly defined before a model can be calibrated. A model calibrated to a set of conditions that do not represent real field conditions may result from a lack of proper site characterization. A model can be calibrated either automatically or manually. “PEST (Doherty et. Al. 1994) and UCODE (Poeter and Hill 1994) are the software used for automatic calibration”.

3.6. Model Verification

To meet historical field conditions, selected values of hydrogeological parameters, boundary conditions, sources, and sinks are used by a calibrated model. The model could be calibrated or refined further as a result of the verification process. Model verification is used to see if the calibrated model performs satisfactorily on any dataset or not. Since the calibration phase entails adjusting various parameters including recharge, pumping rate, hydraulic conductivity among other factors. As a result, different values for these parameters can yield a similar output. The model is ready for predictive simulations once it has successfully replicated measured changes in field conditions.

3.7. Sensitivity Analysis

There are a lot of uncertain parameters in regional groundwater models. It takes a long time and a lot of work to deal with these uncertainties. Sensitivity analysis identifies which parameters or a parameter has the greatest impact on the final result. The process of changing values of input parameters of a model over a fair range (limits of uncertainty in the values of the model parameters) and examining the respective difference in model output is known as a sensitivity analysis. Observed changes in the contaminant transport, flow rate, or hydraulic head are usually recorded. Instead of parameters for which the model is comparatively insensitive, parameters for which the model is relatively responsive will necessitate future characterization.

3.8. Uncertainty Analysis

Any type of modelling is no better than an estimate due to uncertainties and errors in contaminant transport modelling groundwater flow modelling. For a variety of factors, uncertainty in groundwater modelling is unavoidable. Aquifer heterogeneity is one cause of uncertainty. Uncertainty exists in field data as well. “Mathematical modelling implies many assumptions and estimations, which increase the uncertainty of the model output”(Baalousha & Köngeter, 2006).

Any potential groundwater flow or contaminant transport conditions can be predicted using a model. The model can also be used to evaluate various problem mitigation options. As a result, all model predictions must be manifested as a set of potential outcomes, taking into account the uncertainties and assumptions made in model parameters and model input data.

4. Migration of Solutes in Groundwater

From a variety of sources, contaminants are mobilised and released into the groundwater. Advection, advection-dispersion, and advection-dispersion-chemical/biological reaction models are typical solute transport models (Saatsaz, 2020). Advection models are defined as the solute's movement through groundwater flow, where there is no change in the solute concentration with distance. "Advection is primarily due to the effective porosity, hydraulic conductivity, and hydraulic gradient of the medium, according to Darcy's law of groundwater flow". When solute concentration changes in time and space due to diffusion and dispersion, advection-dispersion models are used to explain solute transport. The diffusion dispersion principle is used in these models. The word "dispersion" refers to a mechanical mixing process induced by the medium's heterogeneity. Solutes can spread in both longitudinal and transverse directions. The rate at which solute mixes by convergence and divergence of molecular diffusion and flow paths is determined by the saturation degree and the value of water velocity in porous media. When individual solutes shift slowly from a higher concentration region to a lower concentration region is called diffusion. Random molecular motions caused by the concentration gradient, with or without water movement, carry out this operation.
Models of advection-dispersion-chemical/biological reactions are by far more complicated. When chemical or biological reactions are introduced into advection-dispersion systems, these models are often used to measure the change in contaminants' concentration.

4.1. Governing Equations

The governing partial differential equations of contaminants transport can be expressed as follows (Zheng & Wang, 1998).

$$\frac{\partial}{\partial x}\left(wD_{xx}\frac{\partial c}{\partial x}+wD_{xy}\frac{\partial c}{\partial y}+wD_{xz}\frac{\partial c}{\partial z}\right)+\frac{\partial}{\partial y}\left(wD_{yy}\frac{\partial c}{\partial y}+wD_{yx}\frac{\partial c}{\partial x}+wD_{yz}\frac{\partial c}{\partial z}\right)+\frac{\partial}{\partial z}\left(wD_{zz}\frac{\partial c}{\partial z}+wD_{zy}\frac{\partial c}{\partial y}+wD_{zx}\frac{\partial c}{\partial x}\right)-\left(q_x\frac{\partial c}{\partial x}+q_y\frac{\partial c}{\partial y}+q_z\frac{\partial c}{\partial z}\right)=\rho_b\frac{\partial s}{\partial t}+\frac{\partial wc}{\partial t}+K_m wC^m \pm R \quad \text{.....eq. (4.1)}$$

Where w is the moisture content, D_{ij} $(ij = x, y, z)$ is the hydrodynamic dispersion coefficient, c is the concentration of contaminants dissolved in water, q_x , q_y , q_z , are the groundwater velocities in x,y,z directions respectively and

$$q_x = -K_x\frac{\partial h}{\partial x}, q_y = -K_y\frac{\partial h}{\partial y}, q_z = -K_z\frac{\partial h}{\partial z}$$

h is the hydraulic head, ρ_b is the bulk dry density of porous media, s is the weight of the adsorbed water per unit area of porous media, m is the order of chemical/biological decay, K_m is the contaminant concentration decay coefficient, R is the retardation coefficient and

$$\mathrm{R} = 1 + \frac{\rho_b}{n}K_d \quad \text{.....eq. (4.2)}$$

Where n is the effective porosity and K_d is the partition coefficient. Hydrodynamic dispersion coefficient for isotropic porous media can be calculated by equations given below

$$D_{xx} = \varepsilon_L\frac{v_x^2}{v} + \varepsilon_T\frac{v_y^2}{v} + \varepsilon_T\frac{v_z^2}{v} + D^* \quad \text{.....eq. (4.3)}$$

$$D_{yy} = \varepsilon_L \frac{v_y^2}{v} + \varepsilon_T \frac{v_x^2}{v} + \varepsilon_T \frac{v_z^2}{v} + D^* \text{.....eq. (4.4)}$$

$$D_{zz} = \varepsilon_L \frac{v_z^2}{v} + \varepsilon_T \frac{v_x^2}{v} + \varepsilon_T \frac{v_y^2}{v} + D^* \text{.....eq. (4.5)}$$

$$D_{xy} = D_{yx} = (\varepsilon_L - \varepsilon_T) \frac{v_x v_y}{v} \text{.....eq. (4.6)}$$

$$D_{xz} = D_{zx} = (\varepsilon_L - \varepsilon_T) \frac{v_x v_z}{v} \text{.....eq. (4.7)}$$

$$D_{yz} = D_{zy} = (\varepsilon_L - \varepsilon_T) \frac{v_z v_y}{v} \text{.....eq. (4.8)}$$

Where D_{xx} , D_{yy} , D_{zz} = Principal components of dispersion coefficient, D_{xy} , D_{yx} , D_{xz} , D_{zx} , $D_{zy} D_{yz}$ = Cross-terms of dispersion coefficient, ε_L = Longitudinal dispersivity, ε_T = Transeverse dispersivity, D^* = Effective molecular diffusion coefficient, v_x , v_z , v_y are velocity vector components along x,y and z axes, and $v^2 = v_x^2 + v_y^2 + v_z^2$; v is the magnitude of the velocity vector. The equation of the movement of groundwater through porous media is represented by Darcy's law as follow

$$\frac{\partial}{\partial x}(K_{xx} \frac{\partial h}{\partial x}) + \frac{\partial}{\partial y}\left(K_{yy} \frac{\partial h}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_{zz} \frac{\partial h}{\partial z}\right) - Q = S_s \frac{\partial h}{\partial t} \text{.....eq. (4.9)}$$

Where, h = hydraulic head, S_s = specific storage (volume of water released from the storage per unit volume of porous material per unit change in the head (flow area × Thickness), and K_{xx} , K_{yy} and K_{zz} are the hydraulic conductivities along the x, y and z axes. Q is the volumetric water flux (recharge/discharge) per unit volume of porous media (sources and sinks).

$$Q = \sum_{i=1}^{r} Q_i \, S(x - x_i)(y - y_i)(z - z_i) \text{.....eq. (4.10)}$$

Where, S = number of sources and sinks along the x, y, z axes, Q_i is the volumetric injection/discharge rate at point (x_i , y_i , z_i). The value of Q_i in withdrawal/discharging points is positive and for injection/recharging points is negative. More details about the above equations can be found in (Zheng & Wang, 1998) and (Harbaugh, 2005).

5. Simulator software for transport and groundwater flow modelling

I. MODFLOW Software (USGS Modular 3-Dimensional Groundwater Flow Mode)

The MODFLOW is the "most commonly used numerical groundwater flow model" (Kumar, 2012). The US Geological Survey created this three-dimensional model. It is used for saturated zones only and is based on the finite difference method. Numerous data preparation facilities, simple data exchange in standard form, extensive global experience, continuous growth, a relatively low price and source code availability are the advantages of MODFLOW(Kumar, 2003).

MODFLOW employs a block-centred finite-difference technique in the simulation of groundwater flow within the aquifer. The layers may be confined, unconfined or composition of both. Stresses due to external forces including evapotranspiration, areal recharge, flow to wells, flow to drains, and flow through riverbeds can all be modelled. However, unsaturated flow and surface runoff are not included in MODFLOW. So, MODFLOW cannot be used in the problems where the "flux at the groundwater table is dependent on the measured head and we do not know the function in advance" (Harbaugh, 2005).

Visual MODFLOW is one of the user-friendly and most important simulation environments to simulate three-dimensional contaminant transport and groundwater flow. The software's instructions were simple, and it

could only simulate groundwater flow below the water table, according to the instructions. MODFLOWSURFACT, MODFLOW, ZoneBudget, MODPATH, MGO WinPEST and MT3D/RT3D, are all included in the Visual MODFLOW software kit with graphical user interface (Fouad et al., 2018).
The MT3DMS is a three-dimensional numerical model that can be used to simulate solute transport in complicated hydrogeologic settings. It is MODFLOW's 3-D advection, dispersion, sorption, and reaction simulation module. It can be used in both saturated and unsaturated areas. MT3DMS can also simulate heat transfer in the presence of minor buoyancy and viscosity changes. In MT3DMS, chemical reactions such as advection-dispersion and diffusion are involved. It was created using the Finite Difference Method by (Chunmiao Zheng, 1990). Specifically designed to manage "advectively-dominated transport problems without the need to construct refined models for solute transport".
MODFLOW includes another module called RT3D, which is a three-dimensional solute reactive transport model (Clement, 1997). RT3D is used to solve linked partial differential equations describing the transport of multiple and immobile contaminants or species in saturated groundwater systems. Battelle Pacific Northwest National Laboratory developed RT3D (Reactive Transport Modelling in 3D with GMS) employs the FDM (Finite Difference Method). It's a tweaked version of MT3DMS that incorporates several reactive chemical packages. RT3D is capable of accurately simulating natural dilution and bioremediation.

II. FEFLOW (finite element subsurface and transport system)

The federal institute of geoscience (BGR) collaborated with the Institute of Hydromechanics (University of Hannover) to create a fractured rock simulation in the mid-1980s. To implement the flow process in complicated geological structures, RockFlow was created using a multi-dimensional finite element method. The Academy of Science (Chemnitz) created the FEFLOW code for "density-dependent flow processes in porous media" at the same time. At the time, FORTRAN code was used to implement RockFlow and FEFLOW. Groundwater flows and contaminant transport can be simulated in 2D and 3D using a finite element subsurface and transport system (FEFLOW)(Koukidou & Panagopoulos, 2010). Also, it can simulate flow for groundwater age, heat and mass transport in both transient and steady-state modes. FEFLOW is well-suited to salinity and heat-dependent transport. It is a powerful tool for describing the spatial and temporal distribution and reactions of contaminants in groundwater(*FEFLOW 7.3*, n.d.).
It is capable of computing:

1. The processes related to geothermal energy
2. The total period and travel duration of the contaminant in aquifers
3. Ground-water systems with and without free surfaces like perched aquifers, phreatic aquifers, moving meshes
4. Thermohaline flows (Both temperature and salinity dependent transport phenomena)
5. Saturated-unsaturated zone problems
6. Complex parametric and geometric situations.

This program is completely interactive and graphics-based. The pre-, main-, and postprocessing stages are all combined. A data interface and a programming interface are available for GIS (Geographic Information System).
The workflow of FEFLOW and MODFLOW based on the conceptual model is shown in figure 5.1.

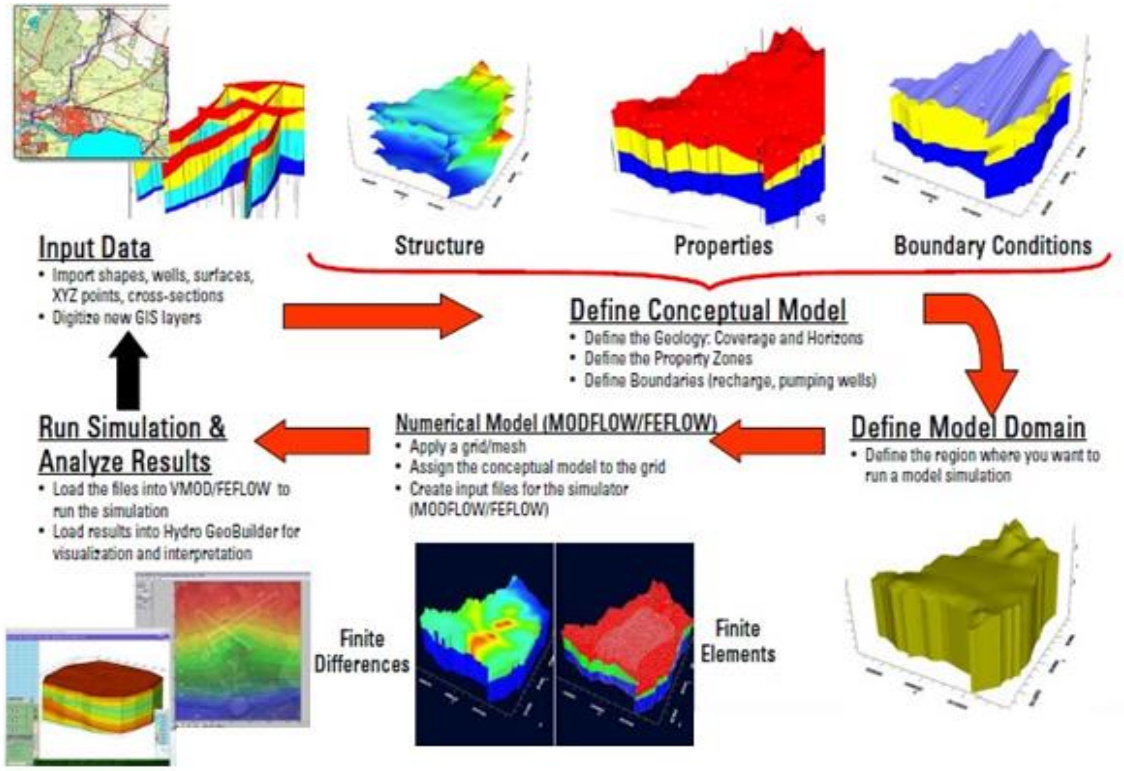


Fig 5.1 Workflow of FEFLOW and MODFLOW(Fouad et al., 2018)

III. HST3D (3-D Heat and Solute Transport Model)

HST3D (Heat and Solute Transport Model) is used to simulate groundwater flow and related heat and solute transport in 3D form. It is an effective and user-friendly interface built into the Argus ONE (Argus Open Numerical Environments) modelling environment. Users can input all spatial data, run HST3D graphically, and visualize the output using HST3D. Within one comprehensive graphical user interface (GUI) Argus ONE incorporates GIS, CAD, Database, Geostatistics, Conceptual Modelling, Automated Grid and Mesh Creation, and Scientific Visualization(Kipp, 1997). The HST3D model can be used to examine issues such as – freshwater recharge and recovery, contaminant plume movement, subsurface waste injection, landfill leaching, geothermal water systems, hazardous waste disposal, the intrusion of saltwater, subsurface energy storage and other transportation-related issues (Kumar, 2012). To run HST3D, you'll need the Argus ONE GIS and Grid Modules. Only the heat- or solute-transport equation is solved in combination with the ground-water flow in most applications.

IV. SUTRA(3-D Saturated/Unsaturated Transport Model)

“SUTRA is a finite-element simulation model for saturated - unsaturated, fluid-density-dependent ground-water flow with energy transport or chemically-reactive single-species solute transport with energy transport”(*Summary of SUTRA*). For saturated groundwater flow zones, SUTRA may be used for cross-sectional and areal modelling and cross-sectional modelling of unsaturated flow systems. “Solute transport simulation with SUTRA can be used to model natural or man-made chemical species transport, including solute sorption, production and decay processes, as well as to analyse ground-water pollutant transport problems and aquifer reclamation designs”. SUTRA can model “variable density leachate movement” and “cross-sectional modelling of saltwater intrusion in aquifers” at regional scales and near-well, with either dispersed or rather sharp transition zones between saltwater and freshwater.

To solve the governing equations describing the two interdependent processes that are simulated, the model uses a “three-dimensional hybrid finite element and integrated finite difference technique”. The approximation

is done for the governing equations of unsaturated or saturated ground-water flow dependent on fluid density, and either
(a) transport of a contaminant in the groundwater, where contaminant can be subjected to equilibrium adsorption on the porous matrix with both first-order and zero-order production or decay, or
(b) transport of thermal energy in the groundwater and solid matrix of the aquifer.

V. Use of GIS and Meshfree techniques

With the aid of various geo analysts and spatial statistical methods, a large number of GIS-based systems have been used in recent years to classify regions vulnerable to contamination(Shrimali et.al,. 2019). ArcGIS Model Builder and Visual Basics (VB) software can be used to develop a GIS-based Contaminant Transport Model (GIS-CONTRAM) to demonstrate the fate of contaminants transport in the subsurface (Menezes, 2009). When checked with Visual MODFLOW, the CONTRAM model achieved fair results. The element free Galerkin method (EFGM) is a sufficiently accurate and simple method for numerically simulating contaminant transport in two dimensions through saturated homogeneous porous media and landfill liners. In the EFGM, an estimated solution is designed entirely in terms of a series of nodes, and no characterization of the nodes interrelationship is needed. The EFGM approximates the function with moving least square approximants and imposes critical boundary conditions with the Lagrange multiplier process. The EFGM findings and the data from the field investigation are in good agreement(Praveen Kumar & Dodagoudar, 2009).

6. CONCLUSION

This study indicates that the knowledge of the basics of groundwater modelling, selection of the right numerical or analytical model and clear model objective will solve any problem related to contaminant transport in groundwater. The selection of appropriate models is governed by factors such as aquifer properties, hydrogeology, type of pollutants, and the solute transport process that occurs in one's study area. Model calibration is also an important step in model development and it can be done manually or automatically. The finite difference method is used by MODFLOW-based software, while the finite element method is used by software like FEFLOW. No method is ideal, and each method has its own set of advantages and drawbacks. For numerically simulating contaminant transport in two dimensions, the mesh-free design is a sufficiently accurate and simple process. The outcomes of the EFGM and the evidence from the field investigation are quite similar. New methodologies, such as Arc Hydro Groundwater executable programming code and GIS CONTRAM, can be built by merging GIS and GMS.

REFERENCES

Baalousha, H. (2009). Fundamentals of groundwater modelling. *Groundwater: Modelling, Management and Contamination; Konig, LF, Weiss, JL, Eds*, 149-166.

Baalousha, H., & Köngeter, J. (2006). Stochastic modelling and risk analysis of groundwater pollution using FORM coupled with automatic differentiation. *Advances in Water Resources*, *29*(12), 1815–1832. https://doi.org/10.1016/j.advwatres.2006.01.006

Clement, T. (1997). A Modular Computer Code for Simulating Reactive Multi-Species Transport in 3-D Groundwater Systems. PNNL-11720. *Pacific Northwest National Laboratory, Richland, WA. 59 pp.*

Shrimali N. J, Daksh H Soni, M. A. M. (2019). *A Study of Various Approaches for Groundwater Contaminant*. http://iwwa-ahmedabad.org/images/downloads/papers/2019/1913.pdf

Elango, L. (2005). Numerical Simulation of Groundwater Flow and Solute Transport. *Allied Publishers Pvt. Ltd.pp.151-164. January 2005*.

FEFLOW 7.3 Documentation - Introduction. (n.d.). Retrieved June 30, 2021, from http://www.feflow.info/html/help73/feflow/01_Introduction/intro.htm

Fouad, M., Hussein, E. E., & Jirka, B. (2018). Assessment of Numerical Groundwater Models. *International Journal of Scientific & Engineering Research*, *9*(6), 951–974. http://www.ijser.org

Harbaugh, A. W. (2005). The U.S. Geological Survey Modular Ground-Water Model — the Ground-Water Flow Process. *U.S. Geological Survey Techniques and Methods 6-A16*, 253.

Kipp, K. L. (1997). Guide to the revised heat and solute transport simulator: HST3D- Version 2. *U.S. Geological Survey Water-Resources Investigations*, *97–4157*, 1–158.

Koukidou, I., & Panagopoulos, A. (2010). Application Of FEFLOW For The Simulation Of Groundwater Flow At The Tirnavos (Central Greece). *Bulletin of the Geological Society of Greece*, *43*(2010), 1747–1757.

Kumar, C. P. (2003). Introduction to Groundwater Modelling. (Vol. 247667). http://nihroorkee.gov.in/sites/default/files/uploadfiles/CPK_GW_Modelling.pdf

Kumar, C. P. (2012). Groundwater Modelling Software – Capabilities and Limitations. *IOSR Journal of Environmental Science, Toxicology and Food Technology*, *1*(2), 46–57. https://doi.org/10.9790/2402-0124657

Pinder G., G. W. (1978). Finite element simulation in surface and subsurface hydrology. *Advances in Water Resources*, *1*(3), 175. https://doi.org/10.1016/0309-1708(78)90029-5

Praveen Kumar, R., & Dodagoudar, G. R. (2009). Modelling of contaminant transport through landfill liners using EFGM. *International Journal for Numerical and Analytical Methods in Geomechanics*, *34*(7), n/a-n/a. https://doi.org/10.1002/nag.817

Summary of SUTRA. (n.d.). Retrieved March 11, 2021, from https://water.usgs.gov/cgi-bin/man_wrdapp?sutra(1)

Wang, D., Wu, C., Huang, W., & Zhang, Y. (2019). Vibration investigation on fluid-structure interaction of AP1000 shield building subjected to multi earthquake excitations. *Annals of Nuclear Energy*, *126*, 312–329. https://doi.org/10.1016/j.anucene.2018.11.021

Zheng, C., & Wang, P. (1998). MT3DMS: A modular three-dimensional multispecies transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater systems. *US Army Corps of Engineers Engineer Research and Development Center*, *June*, 239. http://www.geo.tu-freiberg.de/hydro/vorl_portal/gw-modellierung/MT3DMS Ref Manual.pdf

2021 2nd International Conference for Emerging Technology (INCET)
Belgaum, India. May 21-23, 2021

Object Tracking in a Zone using DeepSORT, YOLOv4 and TensorFlow

Dr. Rajni Jindal
Delhi Technological University,
Delhi, India
rajnijindal@dce.ac.in

Aditya Panwar
Delhi Technological University,
Delhi, India
adityapanwar_2k17co27@dtu.ac.in

Nishant Sharma
Delhi Technological University,
Delhi, India
nishantsharma_2k17co214@dtu.ac.in

Aman Rai
Delhi Technological University,
Delhi, India
amanrai_2k17co45@dtu.ac.in

2021 2nd International Conference for Emerging Technology (INCET) | 978-1-7281-7029-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/INCET51464.2021.9456443

Abstract*: In this paper, we are watching out for the issue of different article track-ing in a single packaging. To oblige the course of action, we at first recognize all of the articles in the packaging and subsequently designate wonderful ID's to all of them until they move out of the edge of reference. We proposed to deal with this issue in three stages, which are according to the accompanying: perceiving all of the articles in the current edge followed by identi-fying them, and thereafter finally, we intend to follow them. We took this issue announcement considering the way that in spite of the way that we have various estimations identifying with this situation, the MOT-A figure is not adequate. So to manage on something practically the same, we hope to use Kalman channels for redesigned results and differentiated our MOT-An estimation and right now open works in this space and the results were adequate. The key computation used by us was YOLO after which the articles are recognized and organized under various classes. After this, for settled mo-tion figure and feature extraction, Kalman channels were used to show these states and as of late perceived articles are settled of a current edge, behind the linking which is created for new acknowledgment. It is done using the Deep SORT algo-rithm, which is mainly a Deep alliance metric using a SORT computation. We used kalman channels because they improve the precision of our model and besides redesigned the result for a most part. On practically identical lines, we have used YOLO to do entity disclosure as well as ID, both at the similar timing. It can in like manner be treated as an identifier, which when is applied a lone neu-ral association can expect the hopping box and do multiple class gathering. Our proposed plan can have distinctive multi-reason uses in our consistently life like it can hold people back from getting kept at a singular spot, especially during COVID times and rais-ing the alert to caution people. We can similarly use this approach to manage address prob-lems concerning manage the chiefs and perception zoo/biodiversity stops if there ought to emerge an event of contentions between different animals where checking each and every transgression gle area is no basic task.

Keywords: Kalman Filters, Multiple Object Tracking, SORT, YOLOv4.

I. INTRODUCTION

The issue of Multiple article following is a standard issue since today is the need vital. This is because consistently, we see various articles everywhere, be it roads, parks, film passages or markets. We are endeavoring to handle this issue using the top tier development with least information sources and endeavor to achieve the best results. Regardless, saying this and doing these things are two particular things. It is fairly perplexed, and it may make a greater number of counterfeit positives. For the abatement of problematic acknowledgments, we have earlier needed to use simply the SORT computation, but at this point we will overall use the Deep SORT estimation which uses Kalman channels for better gauge [1]. We have surveyed our model on the MOT18 dataset, with the progress in this field of learning estimation and colleague of new pos-sibilities with our regular multi-object following issue and besides to think about a couple of changes in the development front. What about we start with the beginning of such models. As of now we will analyze in a word about the MOT issue and we ought to appreciate what absolutely this MOT issue is? Different Object Tracking (MOT) issue can be simply put as looking at the picture once and perceiving all are the articles in the pic-ture. We go over various articles in a singular edge and perceiving is the first and the chief development, which requires a huge load of challenge and a while later after is another task. So we devised disconnecting the issue into two segments: acknowledgment and connection. The model is furthermore prepared to perform perception and moving article acknowledgment and following consistently.

By and by let us bounce into how is object following and revelation being performed? In this computation, we have used the SORT algo-rithm with a development, for instance the deep association grid is used that improve the organizing with procedure and reduce the amount of character switche by using the procedure which is to add visually appearing descriptor as well as appearance features. This part includes 2 segments which are first thing recognizing the things and a while later causing a leaping to limit round it every bit of packaging of the system and thereafter following the articles in the ricocheting box till it moves out from the degree of the packaging. Regardless, it was one critical issue which the philosophy due to which results were not adequate and the clarification was hindrance, i.e., our structure couldn't perform well when there was an article in the perspective on another thing since SORT relies upon a fundamental development code and it couldn't work outstandingly as indicated by needed results. Subsequently, we picked the Deep SORT computation for the same. Hence, it had the choice to per-structure better and give needed results. We could moreover betters the precision by using development assessment and through appear incorporates too.

For following of vehicles, we used the segregation estimation. It addresses Simple Online Real-time i.e. SORT which follows the going with progresses:

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:26:45 UTC from IEEE Xplore. Restrictions apply.

- It makes use of Hungarian strategy and kalman filtering to manage development predict as well as data alliance
- Entity model contains scale, entity position, skipping boxes extent as well as development gauge for the accompanying edge
- Arranging tends to the data relationship by finding out bobbing Intersection box over affiliation (IOU) distance
- The last errand is done based on the loU distance
- It has high precision and accuracy and is speedy and fundamental

How might we show up toward the completion of perceiving any article? Starting, an all out game plan of things in the packaging of the picture are recognized. Picture division is moreover an im-portant part of thing detection[2]. At the point when the disclosure is more than, an organizing with plan is set performed is done for tantamount acknowledgments inside the edge of the skipping box and they are followed till they stay inside it. The accompanying request arises concerning how are we getting along object following dynamically? For that, we from the outset made a lot of article detec-tions. By then we made a wonderful ID for all of the acknowledgments. After this, we followed the article all through some time frame till it stays in the packaging. Finally, we kept up the ID till the thing is in the packaging. At the point when it moves out of the edge, the article's ID was vanished, and this is how we are following the things.

Now let us understand the work of the Deep SORT technique as well as YOLO:

- Object ID as well as affirmation happen through YOLO Technique, for instance bouncing boxes are first made and a while later they are put under the recognized classes.
- Motions Predicting and features age happen which an appraisal model is made and kalman channels are use to exhibit this state's so we can follow them till their augmentation in the edge.
- Finally, following occurs with the help of Kalman stations in the past diagram and as of late recognized things are set in the current packaging, after which a connection is made for new revelation and this is cultivated for discovering them.

We utilized Kalman channels as they improved the exactness of our model as they yielded better outcomes. Our calculation can besides go presumably as a locater, which by apply a particular neural affiliation can expect the ricocheting box and can perform multiple class gathering. On equivalent line, that utilized YOLO can perform object ID and confirmation meanwhile. Further, we will talk about past related work in the field prior to examining our procedure and clarifying the idiosyncrasy in the equivalent.

II. Related Work

Past works has shown that the Neural Networking is proving to have a surprising computation for faster acknowledgment and request of authentic articles. Some typical applications of CNN are Decoding Facial Recognition, recognizing vehicles, Ancient and Environment groups, and vehicles as well as fauna on roads [3]. CNN modelling is called Single Shot Multi-box. Detecting make the recognizable proof communication essentially speedier [4].

For making a precise neural association, the going with progress is following:

- Capturing of picture
- Pre-setting up the image
- Segmentation of picture
- Extraction of features
- Distinguishing the image - accomplice the image which the class to has a spot

YOLO model as of late has gotten mainstream for object location and it is not quite the same as different strategies as it utilizes a solitary CNN that can predict a diverse bouncing box around the article into a picture and discovers different probabilities for partner various classes to various jumping boxes. Profound Learning is one of the calculations that has demonstrated its significance over and over in the space of item recognition, however YOLO improve induction rate for the identification of genuine articles as well as it makes the article discovery favorable to cess quicker with colossal exactness and continuous vigor and more reasonable to use, in actuality, scenarios[5].

In the MOT which is Multiple Object Tracking issue we endeavor to screen multi-ple objects that are there in the different edges in an image course of action. Pushing ahead to the second supporter of the problem, i.e entity following which in it-self is a troublesome endeavor, in fact, similarly as during video assessment where we need to screen a thing in different resulting edges by making a bobbing box around it and allocate an intriguing character to the article across all of the edges of that comparable thing which appears in the image gathering which can be either consistent or can be recorded video. It is easier when we have quite recently each article in the edge thusly anyway the issue becomes flighty when we have various things at a comparable second in an edge. Saying perceives all of the things in the packaging and apportions an exceptional character to each dissent and uses these characters to screen different articles in the packaging over time[6]. The article track-ing system has a stream wherein they play out the task of discovering various things of an image course of action examination which can be either steady or recorded video. First thing an article is perceived and a while later the improvement of the thing is continued in following frames[7].

Here in this endeavor, we do the thing tracking machine to distinguish multi-ple entities in an area inside an edge of picture game plans and keep the incorporate of the outright article in the edge similarly as the hard and fast item in a particular zone demonstrated [8]. Thusly, its show can be smoothed out beginning to end clearly and hence handling the info data quickly. For Tracking, we use the Extension of SORT methodology which is Simple Online and Realtime Tracking and when it is getting together with Deep Association Metric called Deep SORT estimation. To recognize and follow various things dependent on various qualities of the articles like the size of article, the speed of thing, shape and genuine appearance of the article, etc we

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:26:45 UTC from IEEE Xplore. Restrictions apply.

use YOLO and Deep SORT[9]. Kaman segregators are used when there is a hindrance in the edge. Kalman filters work on consistent speed and straight discernment and assessments the method of the article when that thing isn't perceptible due to blockages like smoke, murkiness, flying bird, etc In entity disclosure as well as worldwide situating systems, Deep SORT is one of the huge estimations which is astoundingly speedy as well as accu-rate [10].

III. METHODOLOGY

The undertaking's work process mostly comprises of three stages, for example location of an item, following of article and including of the article in the zone. For the main stage, for example detec-tion of an article, we utilized YOLOv4 [11] with pre-prepared loads. YOLOv4 is an expansion in YOLOv3, with better speed and precision Article recognition models are prepared to characterize the items in a picture. . The principle justification utilizing YOLOv4 in our task is that it is effective for continuous article discovery, which is fundamental for our venture. The upside of constant recognition models is that they are not difficult to employ by all designers. YOLOv4 focuses on it for ongoing article location. Moreo-ver, the creators' planned for PC vision architects and designers to utilize their YOLOv4 structure in custom spaces without any problem.

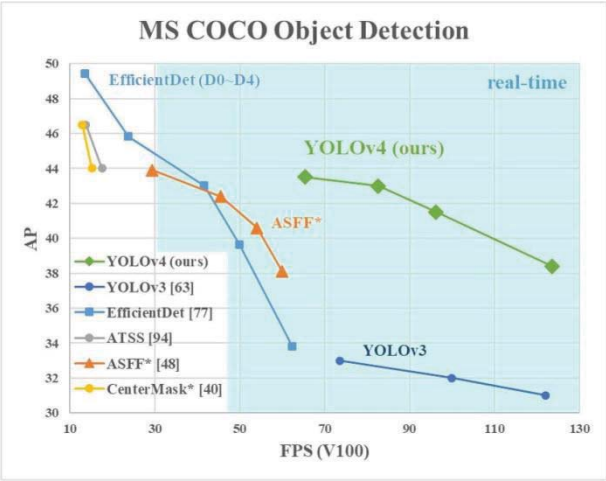


Fig. 1. Other detectors comparison vs YOLO

For the subsequent stage, for example following of items, we utilized the Deep SORT calculation. It is an augmentation to the SORT which is Simple Real-Time Tracker calculation. The DeepSORT calculation utilizes the kalman channels and it is necessary segment to gauge the roads of existing articles in the current edge. It has expressed that it contains eight factors; (u', u, v, 'v, a', a, h', h) where:

- (u,v) – focal points of the boxes bouncing
- h – Image height
- a – perspective extent
- variable other than this are specific there velocity

The Deep SORT uses a general and clear sort of Kalman channels which uses consistent rates as well as straight discernments. Exactly when each new edge comes, the circumstance of each and every track is surveyed dependent on its past regions. It isolates features so much that the features with different characters are at a critical distance isolated, and the features with a comparative character are almost each other. The new ID outcomes are connected with each incoming edges' present tracks utilizing a appearance incorporate vector and the track expected position. Simply the space info is used for track appraisal. By then we used the appearance incorporate vector to portray all of the features of a given picture. This vector gains the appearance info of recognizable pieces of proof by removing features and following pictures from past diagrams. Recognizable pieces of proof with low conviction regard than the edge are filtered through.

The vector is convolution neural association arranged on an enormous re-conspicuous verification dataset. New detec-tions need to pass this breaking point to transform into the contenders of data association. For the last development, which is checking the things, we have counted the amount of articles with the fascinating ID for calculating the amount of things in the zone. By then a cutoff is set for the zone. From the outset, the area is of green tone, and in case the amount of articles outperforms the cutoff, the alarm is ringed and the shade of the area changes to red.

IV. EXPERIMENTS AND RESULTS

This is a tremendous dataset con-taining unique, containing 16658 edges, 2412 tracks and 572834 bouncing boxes. We utilized the MOT18 dataset [12] for assessing the model.

Assessing Multiple Tracking of Objects messes some up as datasets are challeng-ing to build up. Different Object Tracking (MOT) calculations are subject to the indicator. The analysts don't generally concur on the convention to follow, i.e., regardless of whether to assess the tracker with a similar locator or assess the entire framework (indicator in addition to tracker) together? Albeit some presentation measures, for example, MOTA (Multi-ple Object Tracking Accuracy) are ordinarily utilized, a few analysts actually don't cling to this measurement.

In this, we have utilized Multiple Object Tracking Accuracy which is MOTA for assessing our results. MOTA is the most utilized rundown number in MOT. It can be characterized as:

Here we have ,

- FPt = ghost trajectories (No. of positive which are false)
- IDSt = At time t, identity switches number
- FNt = missed targets (No. of negatives which are false)
- GTt = truth of ground

Note: If ground truth is less than threshold, a target is assumed missed

TABLE I. MOTA SCORES

Sequence	Avg. Density	MOTA	FPS
MOT17-03	69.8	61.03%	12.61
MOT17-04	45.3	60.17%	19.43

We surveyed the model using the MOT18 dataset . It was done using Jetson Xavier NX. While using open area to use

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:26:45 UTC from IEEE Xplore. Restrictions apply.

the MOT18 dataset, the MOTA scores are close to bleeding edge trackers. Following rate can be reached upto 31 FPS (traces each second of time) dependent upon the article thickness. On a work territory FPS , GPU/CPU can be significantly higher. Regardless of the way that the tracker runs significantly faster, especially ac-priest. Lightweight finders, for instance, YOLOv4-little can be used to get greater fps up to 45(fps).

V. DISCUSSIONS

Following works so much that when it separates a specific thing, it will relegate a particu-lar exceptional id number to that specific article, and consequently, this intriguing character will change into the fundamental focal point for observing that specific article. Assigning out the remarkable IDs makes it all the more clear for the tracker to screen protests all through the video or imaging gathering (see Fig. 2).

Fig. 2. Object detection having distinguishable ids

During the article disclosure and following Process, there may happen a situation when a specific article is impeded by another article, which may really be the inspiration driving the assessment and can cause the thing in the assessment to vanish from the view totally or fairly (see Fig. 3). By applying the Kalman channels, this issue of hindrance is re-handled.

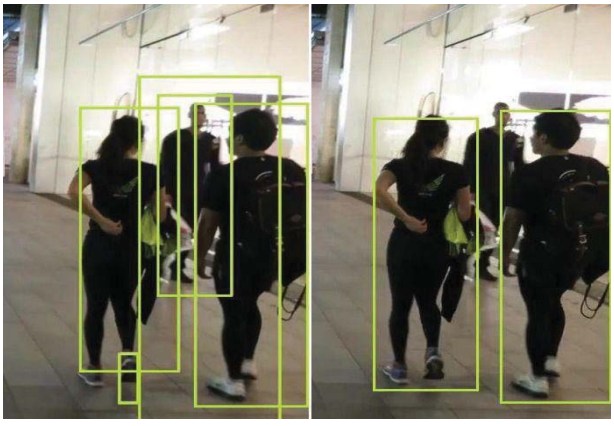


Fig. 3. Detections

The balance of the spread of Covid by placing this device in the Coronavirus space of interest domains and setting up a couple of zones, and raising alarms when the amount of people in the zone gets more unmistakable than the breaking point regard is one of the essential ap-plications of this article disclosure and worldwide situating structure in a specific zone. Exactly if the amount of person in the area is under 5 (edge regard), the area gets green (see Fig. 4), and if the amount of persons is more conspicuous than or identical to 5, by then it be-comes red (see Fig. 5).

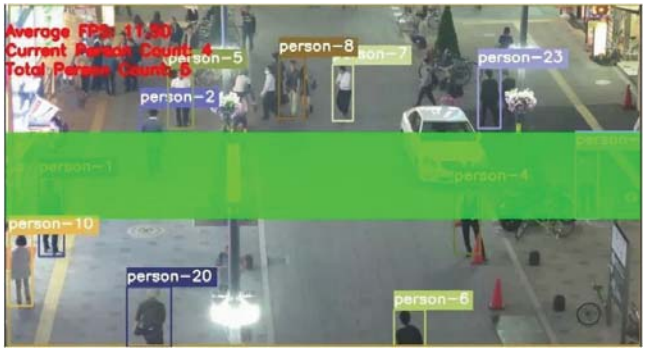


Fig. 4. Object count less than 5 makes zone green

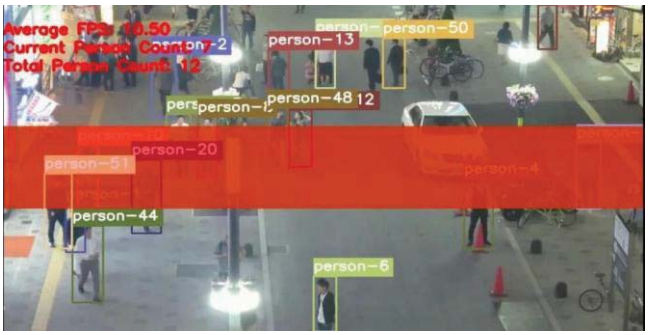


Fig. 5. Objects count greater than 5 make zone red

VI. CONCLUSION

With this, we are going towards the completion of the paper. Besides, it will in general be used to check the portion subtleties and see the quantity of peo-ple are moving beginning with one spot then onto the following, and through this data, govt. Ensuing to perceiving all of the articles in a given packaging, we will follow them until they are in the edge.

The experiment can have many businesses. For example, Tracing of vehicles can have many uses, for instance, checking the quantity of people are there in the leaving of a space. For significant and improved result, we have made use of kalman filters, and a previous SORT estimation was used, which addresses SORT ie simple online realtime tracking, but at this point we have used Deep SORT, which is a SORT computation with a deep connection figures. At last, we made use of the MOT-A figure as well as had the choice to make approx 60% accu-disgusting on area.

To summarize what we have done in this paper, we should rehash the troublesome we are overseeing: object recognizing verification and Tracking in a specific region where we will at first separate the thing using the YOLO show and see something comparative. Can give better inspirations and organizations to such people, and the public authority can similarly isolate the spots of interest and fix the security.

At long last, if there should arise an occurrence of any burglary in a specific district, it will in general be used to track, and we can alert the different experts included with the objective that the case can be dealt with even more quickly and without any problem. It will in general be used to screen

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:26:45 UTC from IEEE Xplore. Restrictions apply.

encroachment of travel guidelines, for instance, way advancing, speeding, and rash driving. It can in like manner be used to decrease time-handling gridlocks.

Because of specific disadvantages, for instance, awful environment, ill-advised lighting conditions during the evening, and less dataset availability, we couldn't achieve unprecedented results. Regardless, for extra degree, we can regardless improve the precision of the model by means of train-ing the model on a greater dataset and from this time forward extending the efficiency of the experiment. For further references, we should use the data which is stored for data assessment and make use of it to outfit the notice ministration with much better outcomes for figuring the leave plan in the era of crisis.

REFERENCES

[1] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," Proceedings of the IEEE, vol. 105, no. 10, pp. 1865–1883, 2017.

[2] S. Shinde, A. Kothari and V. Gupta, "YOLO based Human Action Recognition and Localization," Procedia Computer Science, vol. 133, pp. 831-838, 2018

[3] S. Kanimozhi, G. Gayathri, and T. Mala, "Multiple Real-time object identification using Single shot Multi-Box detection," in 2019 International Conference on Computational In- telligence in Data Science (ICCIDS), (2019).

[4] G. Swalaganata, Mun-iri and Y. Affriyenni, "Moving obje-ct tracking using hybrid method," 2018 International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, 2018, pp. 607-611, (2018).

[5]] D. Wen, X. Huang, L. Zhang, and J. A. Benediktsson, "A novel automatic change detection method for urban high-resolution remotely sensed imagery based on multiindex scene representation," IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 1, pp. 609–625, 2015

[6] P. SHARMA, "A Step-by-Step Introduction to the Basic Object Detection Algorithms (Part 1)," Analytics Vidhya, 2019. [Online]. Available: https://www.analyticsvidhya.com/blog/2018/10/a-step-bystep-introduction-to-the-basic-object-detection-algorithms-part-1/.

[7] Dumitru Erhan, Christian Szegedy, Alexander Toshev, "Scalable Object Detection using Deep Neural Networks", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 2147-2154.

[8] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," 2014.Karen Simonyan, Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", published in Computer Vision and Pattern Recognition (cs.CV).

[9] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep as- sociation metric," in 2017 IEEE International Conference on Image Processing (ICIP), (2017).

[10] YOLO Juan Du1,"Understanding of Object Detection Based on CNN Family",New Research, and Development Center of Hisense, Qingdao 266071, China.

[11] Y. Hu, J. Chen, D. Pan and Z. Hao, "Edge-Guided Image Object Detection in Multiscale Segmentation for High-Resolution Remotely Sensed Imagery," in IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 8, pp. 4702-4711, Aug. (2016).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:26:45 UTC from IEEE Xplore. Restrictions apply.

Journal of the Iranian Chemical Society
https://doi.org/10.1007/s13738-021-02335-4

REVIEW

Oxadiazoles: moiety to synthesis and utilize

Poonam[1] · Geetika Bhasin[1] · Richa Srivastava[1] · Ram Singh[1]

Received: 3 November 2020 / Accepted: 1 July 2021
© Iranian Chemical Society 2021

Abstract
Oxadiazoles are five-membered heterocyclic molecules possessing two carbon, two nitrogen, and one oxygen atom. They exist in four regioisomeric forms and established their potential for a wide range of applications. They have been successfully utilized as an essential part of the pharmacophore or as a flat, aromatic linker to place required substituents as per potential applications. These molecules have also been established themselves as potential high-energy core, and their derivatives have shown favorable oxygen balance and positive heat of formations. The current compilation covers the synthesis of all regioisomeric forms taking representative examples, medicinal applications of oxadiazoles as an anticancer, vasodilator, anticonvulsant, antidiabetic, and other miscellaneous applications. This article also discusses the energetic behavior possessed by oxadiazole derivatives.

Keywords Energetic materials · Heterocyclic ring · High energy molecules · Oxadiazole · Pharmaceutical molecule

Introduction

Oxadiazoles are organic compounds that possess a five-membered heterocyclic ring system. The ring system consists of two nitrogen atoms, one oxygen atom, and two carbon atoms Fig. 1. The oxadiazoles, 1,2,5-oxadiazole (**1**), 1,2,4-oxadiazole (**2**), 1,3,4-oxadiazole (**3**), and 1,2,3-oxadiazole (**4**), are regioisomeric forms Fig. 1. They are of considerable importance in different fields such as material science, medicinal chemistry, and high energy molecules [1–3]. The presence of the two carbon atoms allows two substituents in the oxadiazole ring, whose positions cause differences in their properties due to variation in the electronic environment. The variation in the properties of oxadiazoles (**1**–**4**) allows these molecules to be utilized as high energy molecules or energetic materials, ionic salts, and pharmaceutical compounds [4–10].

The 1,2,5-regioisomer (**1**) is also known as furazan, whose mono and disubstituted derivatives have long been known [3]. Depending on the electron-withdrawing and donating substituents in the 3 and 4 positions of 1,2,5-oxadiazole (**1**), a wide range of compounds has been designed, synthesized, and subjected to various applications, including pharmaceuticals and energetic materials [11–14]. These compounds range from relatively inert to extremely sensitive in terms of their response to heat and impact. Furthermore, the ring has been linked together either directly or with a functional group bridge, thus allowing for a replication of the ring structure and construction of higher molecular weight "oligomers" and macrocycles [4–11]. The 1,2,4-oxadiazole (**2**) derivatives are part of natural products and many biologically important molecules [2]. The 1,3,4-oxadiazole (**3**) derivatives showed their utility for an array of bioactivities [15]. The 1,2,3-oxadiazole isomer (**4**) is not very stable in cyclic forms and exists mostly in tautomeric form as diazoketone (**5**) through 1,5-electrocyclic ring-opening Fig. 2 [16].

The present review article covers the synthesis of oxadiazoles, their medicinal and energetic applications. The medicinal applications like anticancer, vasodilator, anticonvulsant, antidiabetic, and other miscellaneous utilization have been covered. The article furthers covers the energetic behavior possessed by oxadiazole derivatives as high energy molecules.

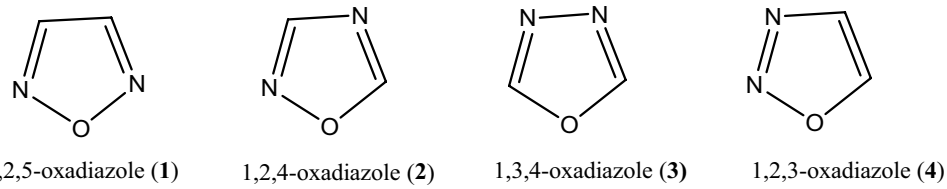
✉ Richa Srivastava
richa.dtu@gmail.com

✉ Ram Singh
ramsingh@dtu.ac.in

[1] Department of Applied Chemistry, Delhi Technological University, Delhi-110 042, Delhi, India

Published online: 06 July 2021

Fig. 1 Regioisomers of oxadiazoles



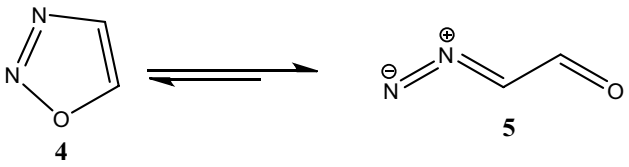


Fig. 2 1,2,3-Oxadiazole and its isomeric form

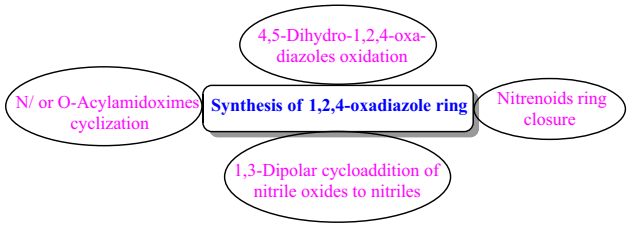


Fig. 4 Important synthetic methods for 1,2,4-oxadiazole ring [14, 15]

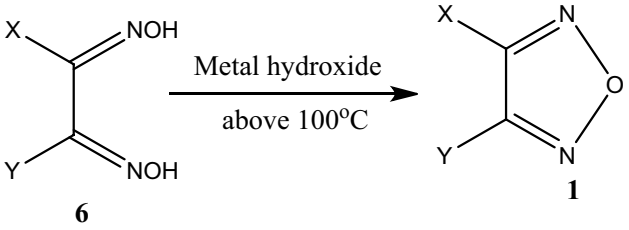


Fig. 3 Synthesis of 1,2,5-oxadiazoles (**1**)

Synthesis of oxadiazoles

Various groups have designed and performed the synthesis of oxadiazoles [17–25]. This section will discuss some of the representative examples for the synthesis of oxadiazoles so that the chemists shall design their synthetic methods based on the given literature collection. 1,2,5-Oxadiazole (or furazan, **1**) has been synthesized from vicinal bisoximes (**6**) through dehydrative cyclization at temperatures greater than 100 °C using metal hydroxides Fig. 3 [17]. Efforts were made to simplify the reaction using several activating reagents attached to one oxygen of bisoximes. Neel and Zhao published the synthesis of **1** at a temperature lower than 100 °C using 1,1′-carbonyldiimidazole (CDI) as a dehydrating agent and bisoximes as a reactant [17]. The method showed better functional group toleration at mild reaction conditions. For the synthesis of 1,2,5-oxadiazoles, the use of metal hydroxides affected the sensitive functional groups like ester [17].

The important synthetic methods of the 1,2,4-oxadiazole ring (**2**) have been compiled in Fig. 4 [18, 19]. The O-acylamidoximes were prepared from the condensation reaction of amidoximes [20] with derivatives of carboxylic acids or acid chlorides [18, 21, 22]. Adib et al. reported 3,5-disubstituted 1,2,4-oxadiazoles (**7**) synthesis through condensation between nitriles (**8**), hydroxylamine (**9**), and aldehydes in 92–97% yields under microwave irradiation and solvent-free conditions Fig. 5 [19]. This method tolerated varied substitutions in the oxadiazole 3- and 5-positions. In this case, the proposed mechanism showed the formation of amidoximes **10** in situ, which further reacts with aldehydes to form the imine derivative **11**. The cyclization and further oxidation gave the desired 1,2,4-oxadiazoles (**7**) Fig. 6 [14].

Poulain et al. used the uronium activation method for the synthesis of 1,2,4-oxadiazoles from carboxylic acids and amidoximes in 66–96% yield [23]. They have used 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate (TBTU) as an activating agent of the carboxylic acid function for the O-acylation step and support the use of sterically hindered acids also [23]. The reaction of nitrile oxide (**12**) with common aliphatic nitriles or electron-rich aromatic nitriles (**13**) also gave 3-functionalized 1,2,4-oxadiazoles (**14**) through 1,3-dipolar cycloaddition Fig. 7 [26]. The intermediate nitrile oxide was obtained from easily available 2-methyl-4-nitro-3-isoxazolin-5(2H)-one with THF.

The 2,5-disubstituted 1,3,4-oxadiazoles (**16**) have been synthesized through cyclization of acylhydrazines using dehydrating agents like zirconium chloride [27], phosphorus oxychloride [28], triflic anhydride [29]. Another popular method includes oxidative cyclization of N-acylhydrazones (**15**) using reagents such as trichloroisocyanuric acid [30] and ceric ammonium nitrate [31]. Pardeshi et al. synthesized through oxidative cyclization in the presence of N-chlorosuccinimide (NCS) [32] and 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) [33, 34] in CH_2Cl_2 [35] Fig. 8. The 1,3,4-oxadiazol-2-amines (**17**) have been obtained from oxidative

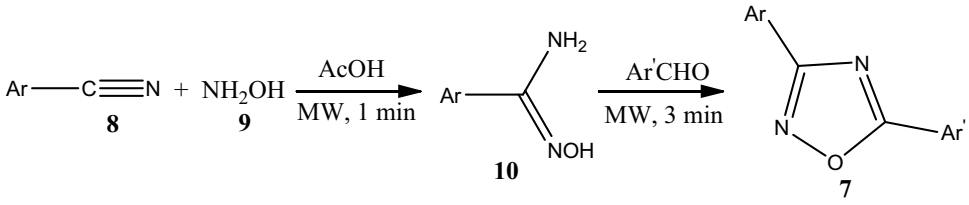


Fig. 5 Synthesis of 3,5-disubstituted 1,2,4-oxadiazoles

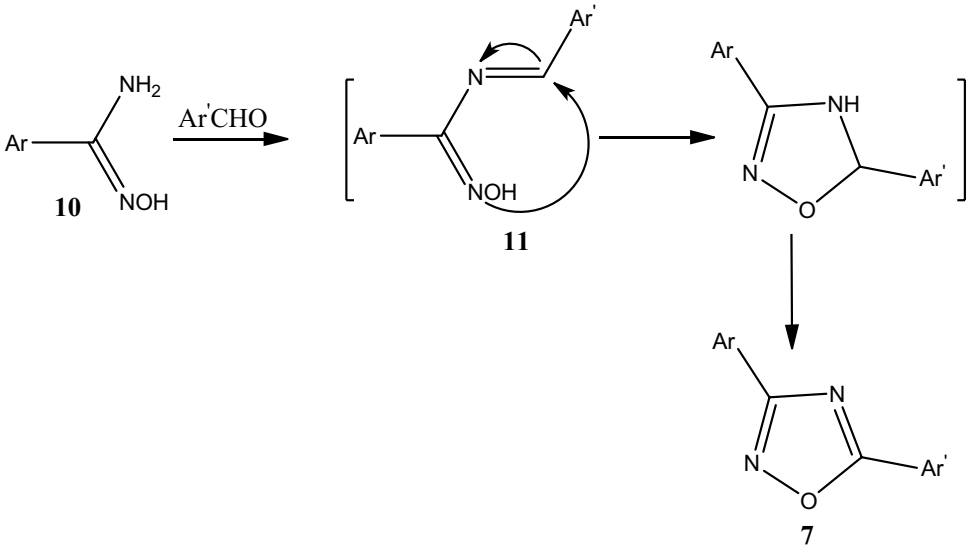
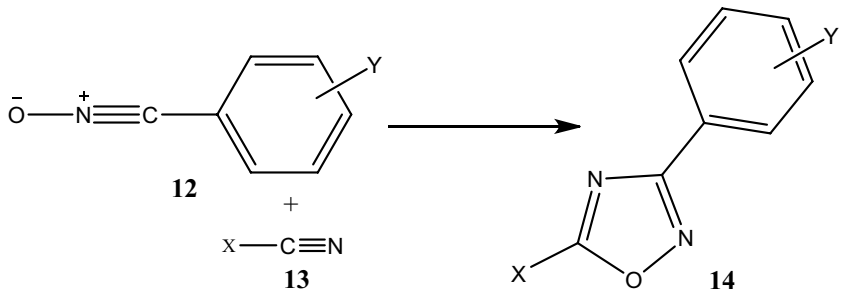


Fig. 6 Proposed mechanism to synthesize 3,5-disubstituted 1,2,4-oxadiazoles



X: electron withdrawing group
Y: electron donating group

Fig. 7 Synthesis of 1,2,4-oxadiazoles

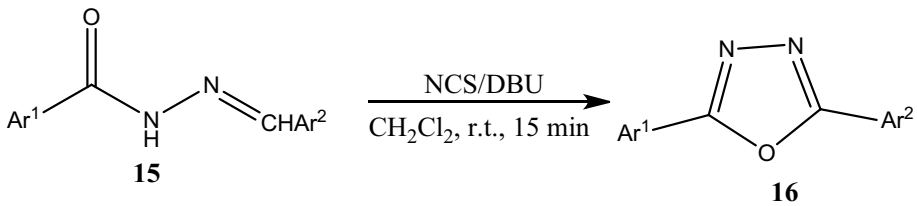


Fig. 8 Synthesis of 2,5-disubstituted 1,3,4-oxadiazoles

cyclization of semicarbazones (**18**) Fig. 9 [36]. This was the ultrasonic reaction in the presence of NBS and sodium acetate (NaOAc).

Zhao et al. synthesized 2-imino-1,3,4-oxadiazoline derivatives (**21**) from hydrazides (**19**) with isothiocyanates (**20**) under metal-free conditions Fig. 10 [37]. This

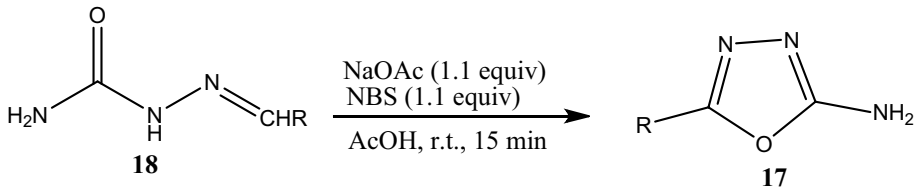


Fig. 9 Synthesis of 1,3,4-oxadiazol-2-amines

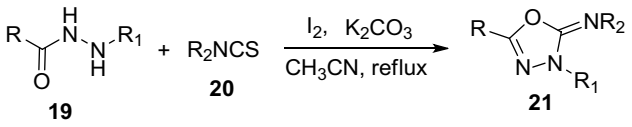


Fig. 10 Synthesis of 2-imino-1,3,4-oxadiazoline

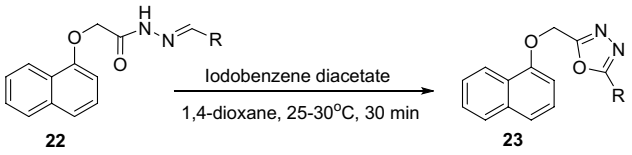


Fig. 11 Synthesis of 2,5-disubstituted-aryl-1,3,4-oxidazole

was an annulation reaction, followed by desulfurization/intramolecular rearrangement, and gave a maximum of 94% yield. This method tolerated many substituents; however, the yield for the substituents like $R = R_1 = Ph$ and $R_3 =$ cyclohexyl obtained in only 12% yield. Jha et al. also followed the similar synthetic protocol and synthesized 11 derivatives of 2-imino-1,3,4-oxadiazolines having an ester functional group ($R_1 = Ar$, $R_2 = CH_2CH_2COOR_4$, and $R_3 = Ph$) [38]. The reactions took time 4–20 h to give the product in 40–90% yields. The low yield for some substrates and longer reaction time remains a limitation for the developed method. Recently, Viswanathan et al. synthesized a new class of 2,5-disubstituted-aryl-1,3,4-oxidazoles (**23**) from 2-(naphthalen-4-yloxy)acetohydrazide derivatives (**22**) in 93–95% yield in the presence of iodobenzene diacetate in the solvent dioxane at a temperature of 25–30 °C Fig. 11 [39].

Different synthetic methods discussed above give a piece of wide information about the synthesis of oxadiazoles. Some of the limitations observed are harsh reagents like sodium hydride [40] and reaction conditions like metal hydroxide at high temperature [17], and lack of wide acceptability of substrates [37]. Also, the use of microwaves is always appreciated, but still, their large-scale industrial application is its limitation [41, 42]. Due to the above-reported drawbacks from synthetic methods, still, various groups are working on this useful ring system. Any research group can take up the oxadiazoles synthesis involving sustainable methods and principles of green chemistry that can be commercially exploited.

Oxadiazole-based pharmaceutical molecules

Oxadiazoles have immensely attracted the interests of the scientific community for over two decades due to their pharmaceutical importance [2, 43]. They have shown capability to bind with different receptors or targets with suitable metabolic profile [15, 44–46]. A number of oxadiazole derivatives are either in the final stages of clinical trials or have been launched as effective drugs, including ataluren (**24**) [47], used against cystic fibrosis, raltegravir (**25**) [48] has been launched as an antiretroviral drug helping in the treatment of HIV and zibotentan (**26**) [49] Fig. 12 is a potential anticancer agent. Various other oxadiazole-based drug discovery plans for various diseases are going on throughout the globe, such as obesity [50], cancer [51], diabetes [52], and inflammation [53]. This heterocyclic moiety has been found as an aromatic linker to provide appropriate orientation to substituents [54] and as a necessary part of pharmacophore [55]. Among oxadiazole compounds, 1,3,4-oxadiazole derivative has been explored more due to their ability to undergo various chemical functionalization [56].

Anticancer agent

Mamatha et al. have reported the synthesis of morpholine and piperidine tagged oxadiazole compounds as a potential anticancer agent [57]. Anticancer screening of the synthesized derivatives was performed on MCF-7 (breast cancer), KB (oral cancer), L929 (normal cell line for cytotoxicity). The compound, (2-chlorophenyl){3-methyl-4-[(5-((2-morpholinoethyl) thio)-1,3,4-oxadiazol-2-yl) methoxy]phenylmethanone (**27**) Fig. 13 showed better result by MTT assay with IC_{50} values to be 112.6 µg/mL and 126.7 µg/mL against the MCF-7 and KB cell lines, respectively [57]. This compound also showed activity toward the normal cell line (L292) with IC_{50} value of 256.7 µg/mL against the control paclitaxel 298.6 µg/mL [57].

Mohan et.al. have reported 1,3,4-oxadiazole derivatives as effective molecules for the blocking of NF-κB signaling pathway [58]. NF-κB is inflammatory mediators linked with the progression of malignant hepatocellular

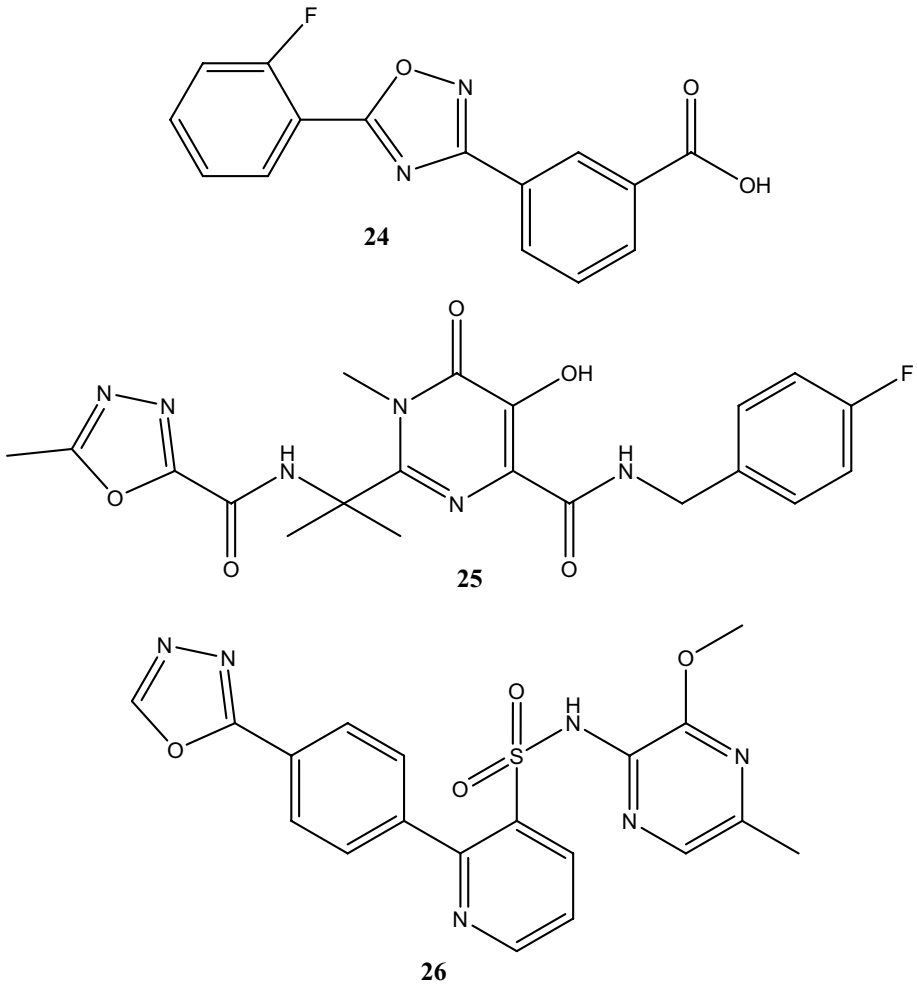


Fig. 12 Some examples of oxadiazole-based pharmaceutical molecules

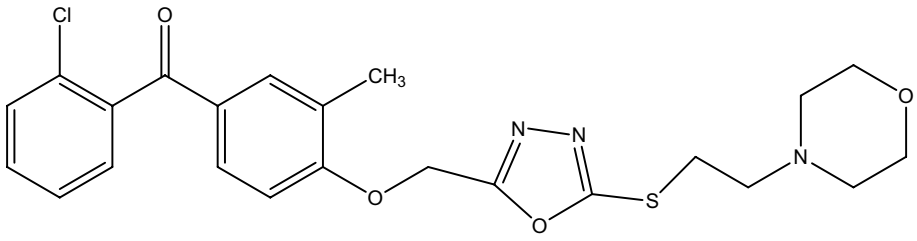


Fig. 13 Chemical structure of compound **27**

carcinoma (HCC) [59, 60]. Thereby, inhibition of the signaling pathway of this mediator served as the basis for the synthesis. The derivatives were screened against HCC cell line, and 2-(3-chlorobenzo[b]thiophen-2-yl)-5-(3-methoxyphenyl)-1,3,4-oxadiazole (28) Fig. 14 was identified as the most potent antiproliferative agent (IC_{50} of 27.5 µM against HCCLM3 cells) [58].

Akhtar et al. synthesized benzimidazole linked 1,3,4-oxadiazole derivatives for the inhibition of EGFR and erbB2 receptor via employing gefitinib as a positive control [61]. EGFR is an epidermal growth factor receptor also known as erbB1 and erbB2 [62]. Their overexpression and inadequate

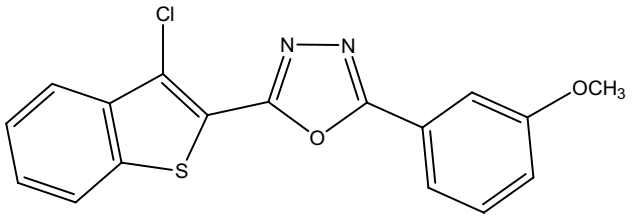


Fig. 14 Chemical structure of **28**

activation result in uncontrolled cell growth, which leads to the development of cancer. EGFR activation is linked with the development of various human tumors such as lung, skin, breast, and liver; based on this fact, they conducted cytotoxicity studies on five cancer cell lines, including lung cancer (A549), breast cancer (MCF-7, MDA-MB231), liver cancer (HepG2) and skin cancer (HaCaT) using the synthesized compounds. The compound 2-((1H-benzo[d]imidazol-2-yl)methyl)-5-(4-chlorophenyl)-1,3,4-oxadiazole (**29**) (Fig. 15) inhibited EGFR and erbB2 receptor at 0.081 and 0.61 mM concentration, respectively [61]. This compound induced apoptosis and resisted the cell cycle at G2/M phase and hence inhibited the MCF-7 cancer cells. This was also observed that the presence of *p*-methoxyphenyl at the 5th position (**30**) Fig. 15 also showed effectiveness toward cancer cell line [61].

Another series, 1,2,5-oxadiazole-3,4-diamide derivatives, were tested for growth-inhibitory potency in U-87 MG human glioblastoma and SW480 human colon adenocarcinoma cell lines [63]. The assay results showed that compound (**31**) Fig. 15 having two chloroacetyl moieties possessed a significantly lower IC_{50} in both cell lines (IC_{50} = 17.1 and 7.4 l M) [63].

Patridge et al. reported effective furazan that function by disturbing protein–protein interactions as 7-nitro-4-(phenylthio)benzofurazan (**32**) Fig. 15 [64]. This compound carries out redox-cycling by accepting electrons readily, and the nitro group of the compound gets reduced in a similar manner as that of aromatic nitro compounds. The superoxide and hydrogen peroxide observed as products of its redox-cycling activity, along with the rapid rate of ROS production [64]. Zhang et al. has reported the development of furoxan-based oleanolic acid derivatives, which functions as nitric oxide-releasing groups [65]. The screening of synthesized compounds found the compound **33** Fig. 15 most potent. In vitro studies found that the compound shows appreciable cytotoxicity against human hepatocellular carcinoma (HCC). Also, it showed inhibition of the growth of HCC tumors in vivo. Though, it's less solubility in aqueous media and toxic nature due

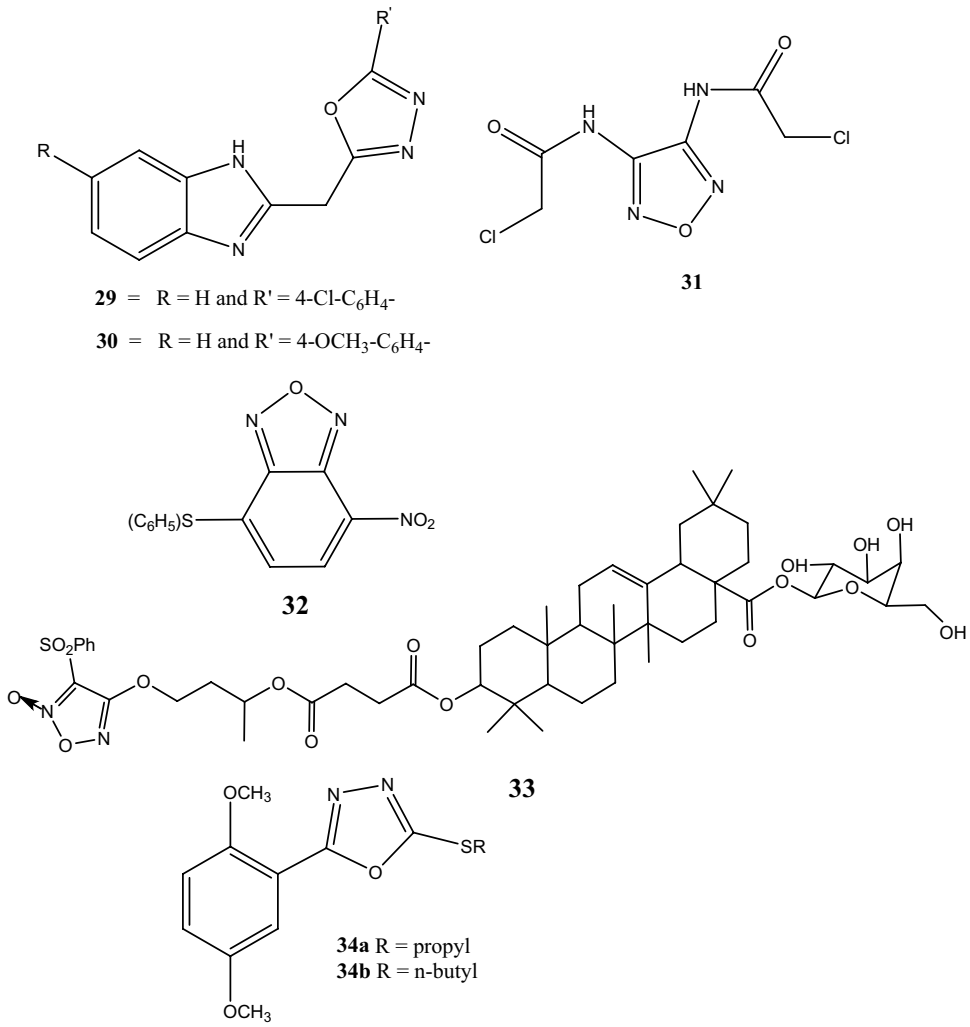


Fig. 15 Chemical structures of oxadiazole derivatives **29–34**

to the fast release of nitric oxide (NO) in blood defied its formulation [65].

The 1,3,4-oxadiazole analogs (**34**) Fig. 15 evaluated against a range of human cancer cell lines [66]. The compounds 34a and 34b gave better result against breast cancer cell lines. This may be due to the better interaction of the receptors with thioether moieties. The compound 34a gave good results against all the cell lines. The study was done using the drug doxorubicin as a reference [66].

Vasodilator

Vasodilators are medications that help in the easy flow of blood through your vessels. Numerous literature states that furoxan derivatives produce NO or related N-oxides species in a controlled manner [67, 68]. The released NO plays a vital role in maintaining micro and macrovascular homeostasis in the cardiovascular system via various mechanisms, including variation of monocyte, and platelet adhesion, vascular relaxation, inhibiting the platelet aggregation and proliferating the smooth muscle cell [67]. It also serves in the modulation of gastric mucosal defense. Any abnormality or impairment in its production leads to various cardiovascular diseases such as atherosclerosis and atrial fibrillation [68]. The compounds 35 and 36 Fig. 16 are notable examples of furoxan derivatives acting as potent vasodilators [69, 70].

Lolli et al. have developed a series of bisphosphonates linked with a furoxan ring (1,2,5-oxadiazole-2-oxide) serving as nitrogen-containing NO-donor [71]. The compounds were evaluated for vasodilator activities and their affinities for hydroxyapatite. The reported results showed that the compounds efficiently constrained osteoclastogenesis on RAW 246.7 cells [71]. Most potent compounds obtained through screening were evaluated on rat microsomes and human PBMC cells. The synthesized compounds have shown relaxation of contracted vascular tissue. The activity of the compounds was found to work in a concentration-dependent manner [71].

Another class of hypoglycemic compounds working as NO-donors has been reported by Tambolia et al. [72]. Synthesis of these compounds was designed by linking 1,2,5-oxadiazole 2-oxide functioning as N donor with sulfonylurea moiety showing hypoglycemic properties. The compounds were studied for insulin release stimulation from pancreatic b-cells (INS-1E line) of cultured rats [72]. The other studies included in vitro vasodilator and anti-aggregatory activities. Among all synthesized compounds, the compounds 37 and 38 Fig. 16 showed excellent activity.

Anticonvulsant activity

Anticonvulsant agents are used in the treatment of epileptic seizures [73]. The oxadiazole derivatives possessing suitable substituents like thiol, hydroxy, and phenoxy groups showed potential toward anticonvulsant activity [73, 74]. Mashayekh et al. synthesized 2-amino-5-(2-phenoxybenzyl)-1,3,4-oxadiazole series and evaluated them as anticonvulsant agents [75]. These molecules showed relative binding affinity (RBA) toward $GABA_A$ receptors and exhibited IC_{50} values of 1.90, 1.75, and 1.88 1.9 nM for the compounds 39, 40, and 41, respectively Fig. 17 (diazepam (IC_{50} = 2.9 nM)). The binding affinity was also superior to the diazepam, especially the chloro- as well as chloro- and fluoro- on phenyl ring substitution of 2-phenoxy moiety [75].

Harish et al. synthesized 1-[5-(4-methoxyphenyl)-1,3,4-oxadiazole-2-yl]-piperazine derivatives and evaluated for anticonvulsant activity with maximal electroshock test (MES) model using male winster rats model [76]. The results were compared with the standard drug phenytoin, which showed maximum protection of 75.88%. The compounds 42–45 Fig. 17 possessing different nature of groups in sulfonyl moiety exhibited a profound anticonvulsant activity giving the % protection of 77.64, 74.31, 73.47, and 71.25%, respectively. The study concluded that the group(s) attached with the sulphonyl moiety play an important role toward efficacy. In the rotarod test, these compounds showed no neurotoxicity at the maximum dose administered (100 mg/kg) [76].

The other oxadiazole series of compounds evaluated as anticonvulsant activity includes 2-[4-substituted benzylideneamino-5-(N^9-carbazolylmethylene)]-1,3,4-oxadiazoles (46), 9-[(4-substituted phenyl)-3-chloro-2-oxo-azetidin-1-yl-1,3,4-oxadiazol-5-yl]methylene carbazoles (47), and 9-[(4-substituted phenyl)-2-oxothiazolidin-1-yl-1,3,4-oxadiazol-5-yl]methylene carbazoles (48) which showed moderate

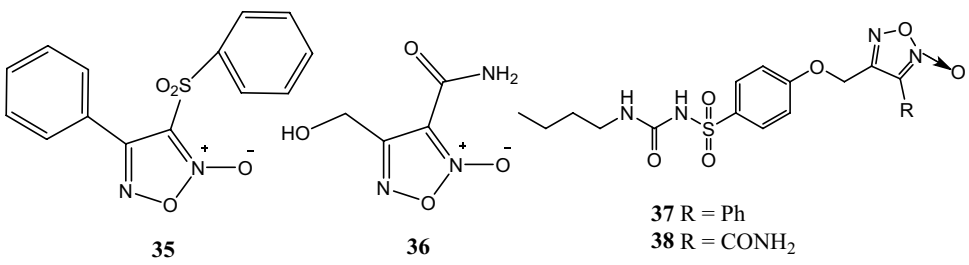


Fig. 16 Oxadiazole derivatives **35–38** evaluated as vasodilators

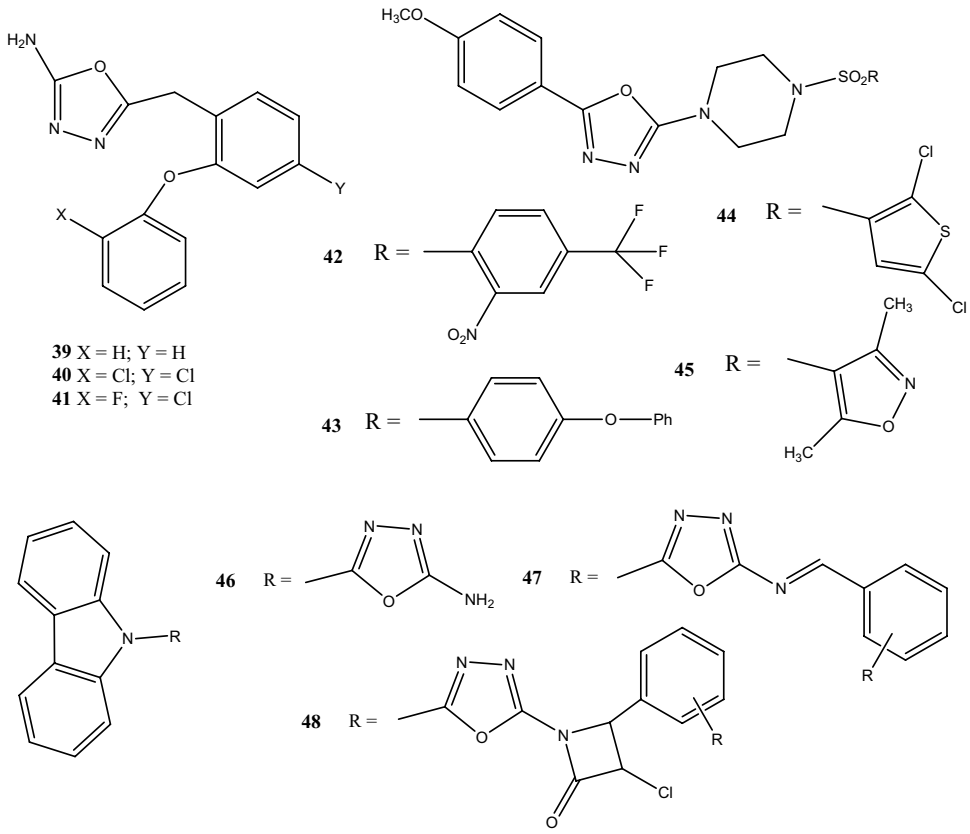


Fig. 17 Oxadiazole derivatives **39–48** evaluated for anticonvulsant activity

potential in comparison with their thiadiazole counterparts Fig. 17 [77].

Antidiabetic activity

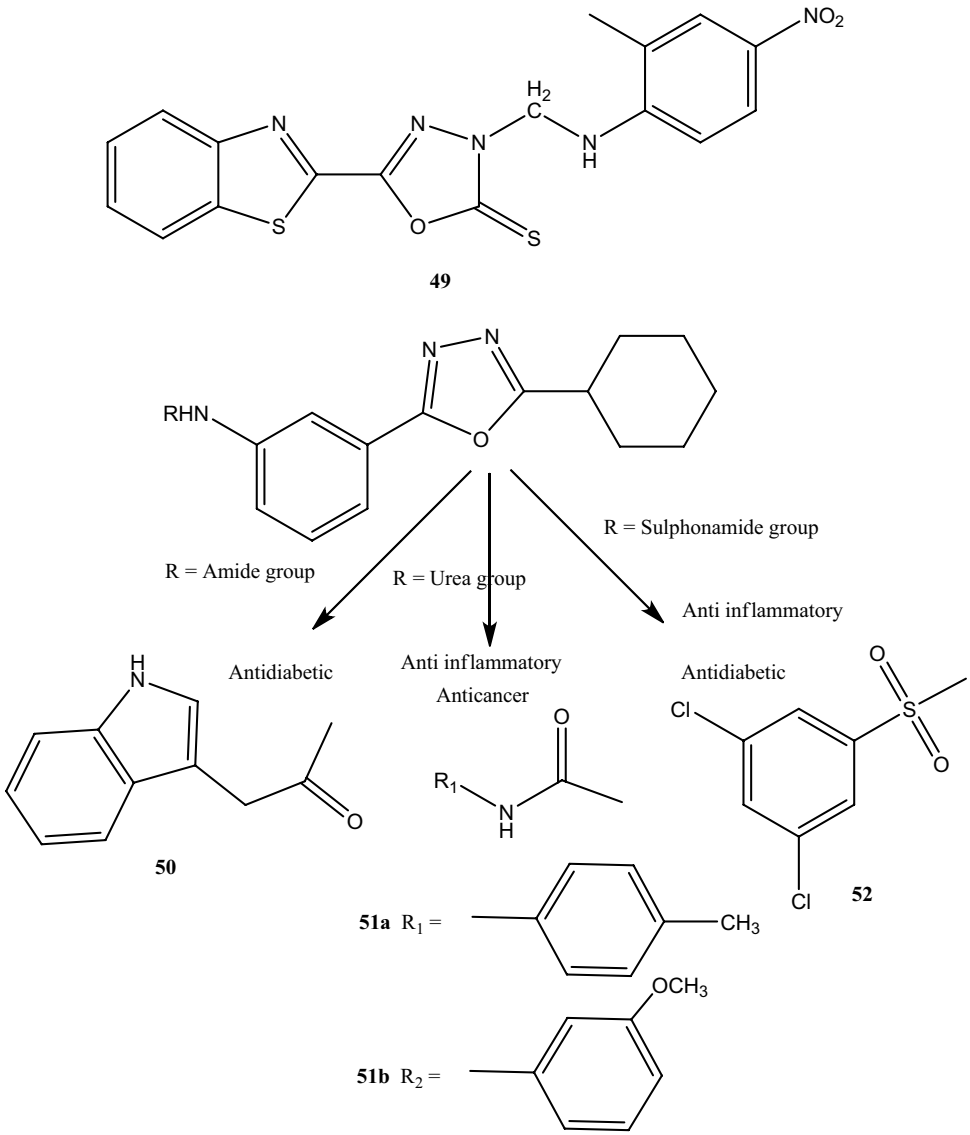
Bhutani et al. synthesized and evaluated the in vivo antidiabetic activity of benzothiazole-1,3,4-oxadiazole conjugates using Oral Glucose Tolerance Test (OGTT) in normal rats followed by Streptozotocin (STZ)-induced diabetes [78]. The compound 5-(benzothiazol-2-yl)-3-[(2-methyl-4-nitrophenylamino)methyl]-1,3,4-oxadiazol-2(3H)-thione (**49**) showed better results by reducing the glucose level as compared to the control drug glibenclamide [78].

The compound 5-benzyl-1,3,4-oxadiazole-2-thiol also showed antidiabetic activity [79]. Another series of oxadiazoles, 3-(5-cyclohexyl-1,3,4-oxadiazo-2-yl)N-substituted anilines, fifteen compounds (**50–52**) Fig. 18, have been synthesized and evaluated in vitro α-amylase inhibition activity (antidiabetic effect), anti-inflammatory, and anticancer by Selvaraj et al. [80]. The compounds having amide and sulphonamide group in oxadiazole enhance antidiabetic activities as observed from N-[3-(5-cyclohexyl-[1,3,4] oxadiazol-2-yl)-phenyl]-2-(1H-indol-3-yl)-acetamide (**50**) and 3,5-dichloro-N-[3-(5-cyclohexyl-[1,3,4]oxadiazol-2-yl)-phenyl]-benzenesulphonamide (**52**) which showed more than 70% inhibition of α-amylase [64]. The compounds 1-[3-(5-cyclohexyl-[1,3,4]oxadiazol-2-yl)-phenyl]-3-p-tolyl-urea (**51a**) and 1-[3-(5-cyclohexyl-(1,3,4)-oxadiazol-2-yl)-phenyl]-3-(3-methoxyphenyl)-urea (**51b**) also showed anticancer activity against MCF-7 and HeLa cells having IC_{50} values in the range of 23.5–45.6 µM using MTT assay [80].

Iqbal et al. synthesized oxadiazole derivatives and evaluated for in vivo hypoglycemic and hypolipidemic activity on alloxan-induced diabetic male wistar rats [81]. The authors studied decrease in plasma glucose (PG) and triglyceride (TG) to improve insulin sensitivity. In the compounds, the presence of -OCH_3 group at 4-position of phenyl ring present at C-5 position of triazole ring, 4-pyridyl ring, and 3-pyridyl ring at C-5 position of the oxadiazole ring showed increased insulin sensitivity [81].

Miscellaneous applications

Apart from the above examples, the oxadiazole molecules have also been widely studied for different other biological activities [82]. The 1,2,5-oxadiazole derivatives showed potential toward inhibition of *Acinetobacter baumannii* [82]. The 1,2,5-oxadiazole-3-carboximidamide derivatives were studied for the inhibition of indoleamine-2,3-dioxygenase 1 (IDO1) to enhance anti-tumor potency [83]. Wang et al. have synthesized two series of 1,3,4-oxadiazole derivatives and evaluated antidepressant activity in vivo [84]. The



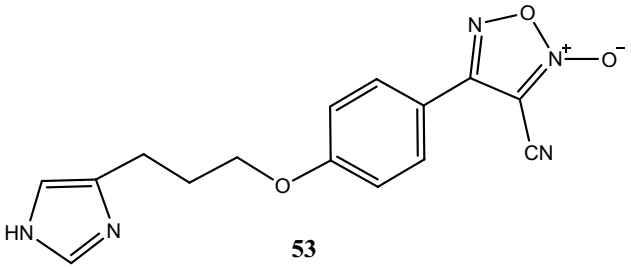


Fig. 19 Chemical structure of furoxan derivative **53**

compound N-(3-((5-((4-chlorobenzyl)thio)-1,3,4-oxadiazol-2-yl)methoxy)phenyl)acetamide gave good results, similar to the drug fluoxetine. Tosco et al. designed selective histamine H_3-receptor antagonists endowed with NO-donor properties, having 1,2,5-oxadiazole as core moiety [85]. The cyano substituted furoxan derivative **53** (Fig. 19) was able to relax electrically contracted guinea-pig ileum to display H_3-antagonist activity at the concentration used in the assay [85]. The oxadiazole derivatives have been successfully being used in place of functional groups like ester, carbamates, and amides [86]. The 1,3,4-oxadiazole derivatives **23** on the molecular docking study showed a strong binding affinity with the targeted COX-2 enzyme [39].

Farshori et al. evaluated 2,5-disubstituted-1,3,4-oxadiazoles for in vitro antimicrobial activities using disc diffusion method taking ciprofloxacin as positive control [87]. The antibacterial activity was evaluated against *Streptococcus pyogenes*, *S. aureus*, *Escherichia coli*, *Klebsiella pneumoniae*, and *Pseudomonas aeruginosa*. The antifungal activity was evaluated in DMSO against Trichophytonmentagrophytes (recultured), *Candida albicans*, *Penicillium marneffei*, and *Aspergillus fumigatus* taking a standard drug,

greseofulvin [87]. Chromene-based oxadiazole derivatives were evaluated against pathogenic bacterial strains, *E. coli* and *K. pneumoniae*, and the standard drug, gentamicin [88]. The 5-phenyl-2-substituted-1,3,4-oxadiazole derivatives were evaluated for their antioxidant and antimicrobial activity [89].

Oxadiazoles as high-energy molecules (HEMs)

The molecular structure of high-energy molecules or energetic materials possess chemical energy inside that can be released due to external stimulations [90], and their design is a hot research topic [91]. The HEMs are very useful for military, space, and civilian purposes, and their incorporation is increasing day-by-day [92]. These types of molecules usually possess large positive heat of formation and N–N and C-N bonds. The compactness of the compound, good oxygen balance, and large moles of gaseous products per gram of the molecule, along with a low percentage of carbon and hydrogen, are some of the important requirements for any molecule to become energetic material or high energy molecules [93, 94]. The weakest bond in the molecule like, R–NO_2 (R = C, N or O) plays important role in these molecules. There is a constant search for low-sensitive and high-performance HEMs, the heterocyclic compounds possessing nitrogen/oxygen showed their potential toward this [92, 95, 96]. To provide better detonation properties, these molecules possess high nitrogen content, high density due to planarity, and high enthalpy of formation and become favorable to the field [97, 98].

Among the oxadiazoles (**1–4**), the 1,2,5-oxadiazole (furazan) have been studied at large for HEMs due to better heat of formation (216.9 kJ mol^{-1}) in comparison with 1,2,3-oxadiazole (158.1 kJ mol^{-1}), 1,2,4-oxadiazole (100.6 kJ mol^{-1}), and 1,3,4-oxadiazole (72.2 kJ mol^{-1}) [95, 99, 100]. The furazan ring has been established as a potential high energy core, and its derivatives have shown favorable oxygen balance. The 1,2,5-oxadiazole subunit along with various cyclic/acyclic explosophoric groups or nitrogen/oxygen-rich groups makes excellent high-energy molecules [101–103]. A series of 1,2,5-oxadiazole derivatives have been synthesized since the work done by Coburn at the Los Alamos Laboratory in 1968 that set up basis for this research [102]. The mono or diaminofurazans are the precursors for the synthesis of corresponding nitrofurazans by oxidation, which have been studied as HEMs [103]. The 1,2,5-oxadiazole derivatives, dinitraminofurazan, and their salts showed the properties of HEMs [104, 105]. The introduction of nitro group along with an additional oxygen atom to the oxadiazole moiety stabilizes it and also enhances density and performance [104]. The compound, 3,3′-dinitroamino-4,4′-azoxyfurazan (**54**) Fig. 20 and its dicationic salts were synthesized and evaluated for HEMs by Zhang and Shreeve [104]. This has been observed that the hydroxylammonium salt (**55**) Fig. 20 is the most potential HEM possessing good detonation pressure (42.2 GPa), acceptable sensitivities (19 J, 120 N), velocity (9511 m s^{-1}), and high density (1.90 g cm^{-3}) [104]. These examples clearly explain the energetic potential of 1,2,5-oxadiazoles. The compounds like 4,4-dinitro-3,3-diazenofuroxan (**56**), possessed density and VOD exceeding that of CL-20 (**57**) Fig. 20; however, they showed limitation with respect to decomposition temperature and high impact sensitivity for use [106].

The binuclear and fused furazans have also been studied and showed their potential toward energetic molecules [107, 108]. Furazan ring fused to a six-membered ring containing two heteroatoms in positions 1 and 4 exists in four structural types, namely furazano[3,4-b]pyrazine (**58**),

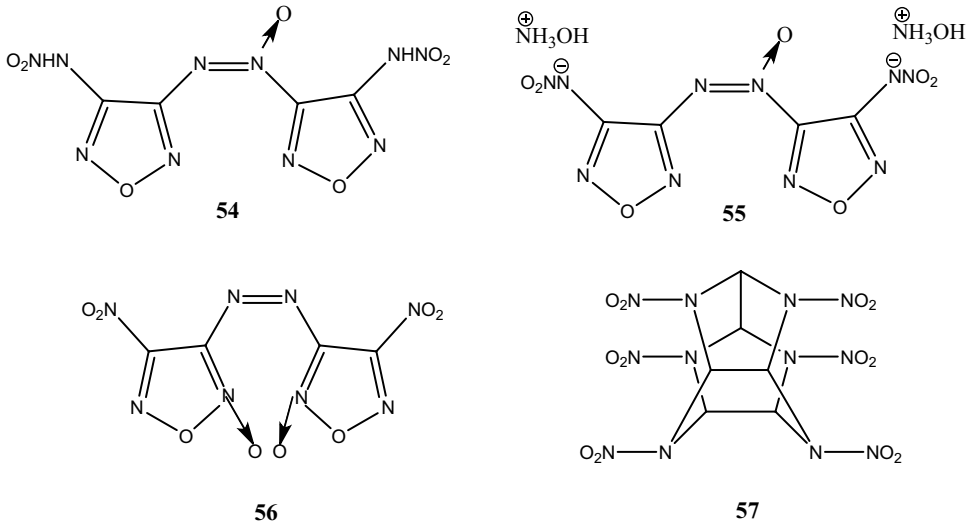


Fig. 20 Chemical structure of oxadiazole derivatives **54–57**

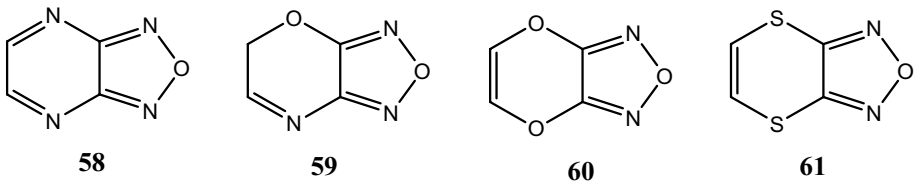


Fig. 21 Furazan ring fused to a six-membered heterocyclic ring

furazano[3,4-b]oxazine (**59**), 1,4-dioxino[2,3-c]furazan (**60**) and 1,4-dithiino[2,3-c]furazan (**61**) Fig. 21.

Out of these four analogs, furazanopyrazine (**58**) have been most studied as energetic molecules [108]. Some examples of studied compounds include, 4H,8H-bis(furazano)[3,4-b:3',4'-e]pyrazine (**62**) [109], 1,4,5,8-tetranitro-1,4,5,8-tetraazadifurazano[3,4-c:3',4'-h]decalin (**63**) [110], 1,2,3-triazolo[4,5-e]furazano[3,4-b]pyrazine 6-oxide (**64**) Fig. 22 [111]. Jin et al. also studied 1,2,5-oxadiazole-1,2,3,4-tetrazole derivatives as high energy materials [112]. The molecules like 1,3,4-oxadiazole-bridged furazan derivatives were also studied for their thermal stability, detonation velocity, friction sensitivity, other high energy parameters [113].

The above examples along with the available literature about the utility of oxadiazole derivative as high energy molecules (HEMs) show their potential towards the field. The requirement for today is HEMs with higher safety because in general contradiction existed between energy and safety of these molecules, they are inversely proportional. Still a lot of work is required to achieve this formidable task of high energy and high safety.

Summary

Oxadiazoles, five-membered heterocyclic molecules, exist in different regioisomeric forms possessing various properties, charge distributions, and applications. This article covered brief information about the medicinal and high-energy properties of oxadiazole moiety. This will help and motivate the researchers working toward the synthesis and application of heterocyclic molecules for the development of new oxadiazole-based molecules, and it also exposed the reputation of the nucleus. The oxadiazole scaffold attracted various medicinal chemists to design and develop biologically active molecules possessing pharmacological activities like anticancer, vasodilators, anticonvulsant, antidiabetic, anti-inflammatory, antimicrobial, and so on. The oxadiazole and fused oxadiazoles showed their utility as HEMs. The discussed content is enough to suggest the broad applicability and acceptability of oxadiazoles.

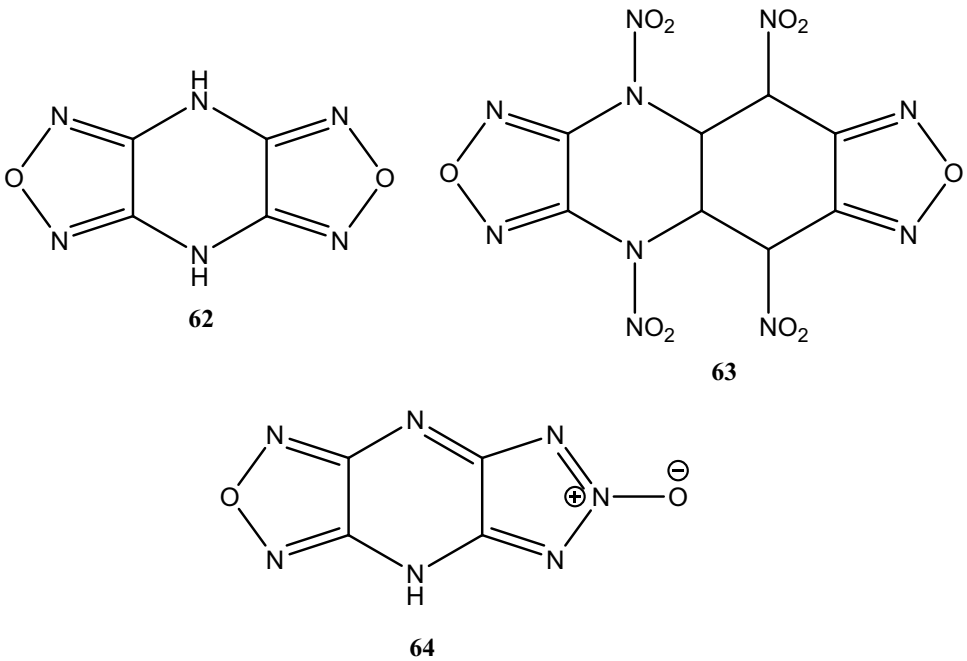


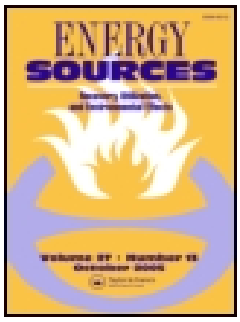
Fig. 22 Furazan ring fused high energy molecules

Acknowledgements The author Poonam is thankful to Delhi Technological University (DTU) for financial support. Other authors are thankful to DTU for infrastructure support.

References

1. A. Pace, P. Pierro, Org. Biomol. Chem. **7**, 4337 (2009)
2. J. Boström, A. Hogner, A. Llinàs, E. Wellner, A.T. Plowright, J. Med. Chem. **55**, 1817 (2012)
3. S. Cicchi, F.M. Cordero, G. Donatella, Prog. Heterocycl. Chem. **18**, 288 (2007)
4. A. Vaidya, S. Jain, P. Jain, P. Jain, N. Tiwari, R. Jain, R. Jain, A.K. Jain, R.K. Agrawal, Mini Rev. Med. Chem. **16**, 825 (2016)
5. A. Vaidya, D. Pathak, K. Shah, Chem. Biol. Drug Design. **97**, 572 (2021)
6. F.M. Cordero, D. Giomi, L. Lascialfari, Prog. Heterocycl. Chem. **29**, 353 (2017)
7. M.A. Prezent, S.V. Baranin, Chem. Heterocycl. Comp. **55**, 1131 (2019)
8. J. Han, J. Mater. Chem. C **1**, 7779 (2013)
9. K. Santhanalakshmi, R.P. Jacquline, S. Muthukumar, Res. J. Chem. Environ. **24**, 157 (2020)
10. A. Singhai, M.K. Gupta, S. Jain, S.P. Chaturvedi, World J. Pharmacy Pharmaceut. Sci. **9**, 886 (2020)
11. M. Xiao, X. Jin, J. Zhou, B. Hu, New J. Chem. **43**, 11610 (2019)
12. P.S. Paulo, S.S. Vitor, M.E.F. Lima, J. Braz. Chem. Soc. **29**, 435 (2018)
13. G. Bhasin, R. Srivastava, R. Singh, Org. Prep. Proced. Int. **49**, 370 (2017)
14. R. Singh, G. Bhasin, R. Srivastava, Mor. J. Chem. **5**, 677 (2017)
15. G. Verma, M.F. Khan, W. Akhtar, M.M. Alam, M. Akhter, M. Shaquiquzzaman, Mini-Rev. Med. Chem. **19**, 477 (2019)
16. J.A. Joule, K. Mills, *Heterocyclic Chemistry* (Wiley, New York, 2013)
17. A.J. Neel, R. Zhao, Org. Lett. **20**, 2024 (2018)
18. J.C. Jochims, A.R. Katritzky, C.W. Rees, E.V.F. Scriven, Comprehensive Heterocyclic Chemistry II. Pergamon Press: London **4**, 179 (1996)
19. M. Adib, A.H. Jahromi, N. Tavoosi, M. Mahdavi, H.R. Bijanzadeh, Tetrahedron Lett. **47**, 2965 (2006)
20. S. Baykov, T. Sharonova, A. Osipyan, S. Rozhkov, A. Shetnev, A. Smirnov, Tetrahedron Lett. **57**, 2898 (2016)
21. Y. Wang, R.L. Miller, D.R. Sauer, S.W. Djuric, Org. Lett. **7**, 925 (2005)
22. V. Santagada, F. Frecentese, E. Perissutti, D. Cirillo, S. Terracciano, G. Caliendo, Bioorg. Med. Chem. Lett. **14**, 4491 (2004)
23. R.F. Poulain, A.L. Tartar, B.P. Deprez, Tetrahedron Lett. **42**, 1495 (2001)
24. B.K. Banik, B.M. Sahoo, B.V.V.R. Kumar, K.C. Panda, J. Jena, M.K. Mahapatra, P. Borah, Molecules **26**, 1163 (2021)
25. F.E. Teslenko, A.I. Churakov, A.A. Larin, I.V. Ananyev, L.L. Fershtat, N.N. Makhova, Tetrahedron Lett. **61**, 151678 (2020)
26. N. Nishiwaki, K. Kobiro, S. Hirao, J. Sawayama, K. Saigo, Y. Ise, Y. Okajima, M. Ariga, Org. Biomol. Chem. **9**, 6750 (2011)
27. G.V.M. Sharma, A. Begum, P.R.K. Rakesh, Synth. Commun. **34**, 2387 (2004)
28. F.N. Hayes, B.S. Rogers, D.G. Ott, J. Am. Chem. Soc. **77**, 1850 (1995)
29. S. Liras, M.P. Allen, B.E. Segelstein, Synth. Commun. **30**, 437 (2000)
30. D.M. Pore, S.M. Mahadik, U.V. Desai, Synth. Commun. **38**, 3421 (2008)
31. M. Dabri, P. Salehi, M. Baghanzadeh, M. Bahramnejad, Tetrahedron Lett. **47**, 6983 (2006)
32. W.M. Gołebiewski, M. Gucma, Synthesis **2007**(23), 3599–3619 (2007)
33. D. Mishra, R. Singh, C. Rout, Int. J. Chemtech. Res. **10**, 365 (2017)
34. B. Veer, R. Singh, P. Roy, Soc. A **475**, 20190238 (2019)
35. S.P. Pardeshi, S.S. Patil, V.D. Bobade, Synth. Commun. **40**, 1601 (2010)
36. A.F. Borsoi, M.E. Coldeira, k. Pissinate, F.S. Macchi, L.A. Basso, D.S. Santos, P. Machado, , Synth. Commun. **47**, 1319 (2017)
37. Q. Zhao, L. Ren, J. Hou, W. Yu, J. Chang, Org. Lett. **21**, 210 (2019)
38. A.K. Jha, S.E. Sarita, Tetrahedron Lett. **69**, 152971 (2021)
39. V. Viswanathan, M. Govindhan, K. Subramanian, D. Velmurugan, R. Padmanaban, S.M. Mahalingam, N. Arumugam, A.I. Almansour, R.S. Kumar, J. Mol. Str. **1229**, 129752 (2021)
40. A.R. Katritzky, P.P. Mohapatra, L. Huang, ARKIVOC **2008**(9), 62–68 (2008)
41. P. Priecel, J.A. Lopez-Sanchez, A.C.S. Sust, Chem. Eng. **7**, 3 (2019)
42. R. Nagahata, K. Takeuchi, Chem. Rec. **19**, 51 (2019)
43. R. Bhutani, D.P. Pathak, G. Kapoor, R. Kant, M. Datt, Indian J. Pharm. **53**, s1 (2019)
44. A. Benassi, F. Doria, V. Pirota, Int. J. Mol. Sci. **21**, 8692 (2020)
45. S.G. Nayak, B. Poojary, Chem. Afr. **2**, 551 (2019)
46. G. Chawla, Mini-Rev. Med. Chem. **18**, 1536 (2018)
47. A.M. Jones, J.M. Helm, Drugs **69**, 1903 (2009)
48. V. Summa, A. Petrocchi, F. Bonelli, B. Crescenzi, M. Donghi, M. Ferrara, F. Fiore, C. Gardelli, O.G. Paz, D.J. Hazuda, P. Jones, O. Kinzel, R. Laufer, E. Monteagudo, E. Muraglia, E. Nizi, F. Orvieto, P. Pace, G. Pescatore, R. Scarpelli, K. Stillmock, M.V. Witmer, M. Rowley, J. Med. Chem. **51**, 5843 (2008)
49. N.D. James, J.W. Growcott, Drugs Future **34**, 624 (2009)
50. S.H. Lee, H.J. Seo, S.H. Lee, M.E. Jung, J.H. Park, H.J. Park, J. Yoo, H. Yun, J. Na, S.Y. Kang, K.S. Song, M.A. Kim, J. Med. Chem. **51**, 7216 (2008)
51. H.Z. Zhang, S. Kasibhatla, J. Kuemmerle, W. Kemnitzer, O.M. Kristin, L. Qiu, C.C. Grundy, B. Tseng, J. Drewe, S.X. Cai, J. Med. Chem. **48**, 5215 (2005)
52. R.M. Jones, J.N. Leonard, D.J. Buzard, J. Lehmann, Expert Opin. Ther. Pat. **19**, 1339 (2009)
53. D.M. Cottrell, J. Capers, M.M. Salem, K.D.L. Fradley, S.L. Croft, K.A. Werbovetz, Bioorg. Med. Chem. **12**, 2815 (2004)
54. M. Ono, M. Haratake, H. Saji, M. Nakayama, Bioorg. Med. Chem. **16**, 6867 (2008)
55. K. Ohmoto, T. Yamamoto, T. Horiuchi, H. Imanishi, Y. Odagaki, K. Kawabata, T. Sekioka, Y. Hirota, S. Matsuoka, H. Nakai, M. Toda, J.C. Cheronis, L.W. Spruce, A. Gyorkos, M. Wieczorek, J. Med. Chem. **43**, 4927 (2000)
56. D.S. Musmade, S.R. Pattan, M.S. Yalgatti, Int. J. Pharm. Chem. **5**, 11 (2015)
57. S.V. Mamatha, S.L. Belagali, M. Bhat, S.N. Appl, Sci. **2**, 882 (2020)
58. C.D. Mohan, N.C. Anilkumar, S. Rangappa, M.K. Shanmugam, S. Mishra, A. Chinnathambi, S.A. Alharbi, A. Bhattacharjee, G. Sethi, A.P. Kumar, K.S.R. Basappa, Front. Oncol. **8**, 42 (2018)
59. R.G. Baker, M.S. Hayden, S. Ghosh, Cell Metab. **13**, 11 (2011)
60. E.Z.P. Chai, K.S. Siveen, M.K. Shanmugam, F. Arfuso, G. Sethi, Biochem J. **468**, 1 (2015)
61. M.J. Akhtar, A.A. Siddiqui, A.A. Khan, Z. Ali, R.P. Dewangan, S. Pasha, M.S. Yar, Eur. J. Med. Chem. **126**, 853 (2017)
62. T. Glomb, K. Szymankiewicz, P. Swiatek, Molecules **23**, 3361 (2018)
63. W.S. Li, S.V. More, C.H. Wang, Y.C. Jen, C.F. Yao, T.F. Wang, C.C. Hung, S.C. Jao, Bioorg. Med. Chem. Lett. **20**, 1148 (2010)

64. E.V. Patridge, E.S. Eriksson, P.G. Penketh, R.P. Baumann, R. Zhu, K. Shyam, L.A. Eriksson, A.C. Sartorelli, Arch. Toxicol. **86**, 1613 (2012)
65. J. Zhang, Y. Gao, F. Su, Z. Gong, Y. Zhang, Chem. Pharm. Bull. **59**, 734 (2011)
66. N. Polkam, S. Malthum, J.S. Anireddy, U. Brahma, G.M.N. Vegi, Russ. Chem. Bull. Int. Ed. **70**, 580 (2021)
67. J.F. Kerwin Jr., J.R. Lancaster Jr., P.L. Feldman, J. Med. Chem. **38**, 4343 (1995)
68. J.L. Wallace, Br. J. Pharmacol. **152**, 421 (2007)
69. C. Medana, G. Ermondi, R. Fnittero, C. Fruttero, P. Gasco, J. Med. Chem. **37**, 4412 (1994)
70. M. Amir, A.M. Waseem, T. Sana, K. Somakala, Int. Res. J. Pharm. **6**, 585 (2015)
71. M.L. Lolli, B. Rolando, P. Tosco, S. Chaurasia, A. Di Stilo, L. Lazzarato, E. Gorassini, R. Ferracini, S. Oliaro-Bosso, R. Fruttero, A. Gasco, Bioorg. Med. Chem. **18**, 2428 (2010)
72. Y. Tamboli, L. Lazzarato, E. Marini, S. Guglielmo, M. Novelli, P. Beffy, P. Masiello, R. Fruttero, A. Gasco, Bioorg. Med. Chem. Lett. **22**, 3810 (2012)
73. S. Nazar, N. Siddiqui, O. Alam, Arch Pharm. **353**(7), e1900342 (2020)
74. P. Singh, P.K. Jangra, Der Chem. Sin. **1**, 118 (2010)
75. S. Mashayekh, N. Rahmanipour, B. Mahmoodi, F. Ahmadi, D. Motaharian, S. Shahhosseini, H. Shafaroodi, H.R. Banafshe, A. Shafiee, L. Navidpour, Bioorg. Med. Chem. **22**, 1929 (2014)
76. K.P. Harish, K.N. Mohana, L. Mallesha, B.N. Prasanna Kumar, Eur. J. Med. Chem. **65**, 276 (2013)
77. H. Kaur, S. Kumar, P. Vishwakarma, M. Sharma, K.K. Saxena, A. Kumar, Eur. J. Med. Chem. **45**, 2777 (2010)
78. R. Bhutani, D.P. Pathak, G. Kapoor, A. Husain, R. Kant, M.A. Iqbal, Bioorg. Chem. **77**, 6 (2018)
79. S. Qamar, K. Hussain, N.I. Bukhari, S.Z. Siddique, A. Rehman, M.A. Abbasi, S. Parveen, A. Latif, A. Qamar, E. Ali, N. Shehzadi, M. Islam, S. Naheed, Trop. J. Pharm. Res. **18**, 1095 (2019)
80. K. Selvaraj, K. Kulanthai, G. Sadhasivam, Saudi Pharm. J. **25**, 337 (2017)
81. M.A.K. Iqbal, A.Y. Khan, M.B. Kalashetti, N.S. Belavagi, Y.D. Gong, I.A.M. Khazi, Eur. J. Med. Chem. **53**, 308 (2012)
82. R.M. Christoff, G.L. Murray, X.P. Kostoulias, A.Y. Peleg, B.M. Abbott, Bioorg. Med. Chem. **2**, 6267 (2017)
83. X. Song, P. Sun, J. Wang, W. Guo, Y. Wang, L.H. Meng, H. Liu, Eur. J. Med. Chem. **189**, 112059 (2020)
84. S. Wang, L. Qi, H. Liu, K. Lei, X. Wang, R. Liu, RSC Adv. **10**, 30848 (2020)
85. P. Tosco, M. Bertinaria, A. Di Stilo, C. Cena, G. Sorba, R. Fruttero, A. Gasco, Bioorg. Med. Chem. **13**, 4750 (2005)
86. G.A. Patani, E.J. LaVoie, Chem. Rev. **96**, 3147 (1996)
87. N.N. Farshori, A. Rauf, M.A. Siddiqui, E.S. Al-Sheddi, M.M. Al-Oqail, Arab. J. Chem. **10**, S2853 (2017)
88. N. Baral, S. Mohapatra, B.P. Raiguru, N.P. Mishra, P. Panda, S. Nayak, S.K. Pandey, P.S. Kumar, C.R. Sahoo, J. Heterocycl. Chem. **56**, 552 (2019)
89. A.S. Madhurao, Y.D. Bodke, K. Rajappa, S. Telkar, B.M. Vinoda, Ind. J. Adv. Chem. Sci. **4**, 257 (2016)
90. J.P. Agrawal, *High Energy Materials: Propellants, Explosives and Pyrotechnics* (Wiley, New York, 2010)
91. G. Zhao, C. He, P. Yin, G.H. Imler, D.A. Parrish, J.M. Shreeve, J. Am. Chem. Soc. **140**, 3560 (2018)
92. L.L. Fershtat, N.N. Makhova, ChemPlusChem **85**, 13 (2020)
93. M.B. Talawar, R. Sivabalan, T. Mukundan, H. Muthurajan, A.K. Sikder, B.R. Gandhe, A.S. Rao, J. Hazard Mater. **161**, 589 (2009)
94. Z. Xu, G. Cheng, H. Yang, J. Zhang, J.M. Shreeve, Chem. Eur. J. **24**, 10488 (2018)
95. M.B. Talawar, R. Sivabalan, M. Anniyappan, G.M. Gore, S.N. Asthana, B.R. Gandhe, Combus. Explos. Shock Waves **43**, 62 (2007)
96. Z. Xu, G. Cheng, H. Yang, X. Ju, P. Yin, J. Zhang, J.M. Shreeve, Angew. Chem. Int. ed. **56**, 5877 (2017)
97. H. Gao, J.M. Shreeve, Chem. Rev. **111**, 7377 (2011)
98. Q. Zhang, J.M. Shreeve, Chem. Rev. **114**, 10527 (2014)
99. Y. Qu, S.P. Babailov, J. Mater. Chem. A **6**, 1915 (2018)
100. N.N. Makhova, L.L. Fershtat, Tetrahedron Lett. **59**, 2317 (2018)
101. H. Xiong, H. Yang, C. Lei, P. Yang, W. Hu, G. Cheng, Dalton Trans. **48**, 14705 (2019)
102. M.D. Coburn, J. Heterocycl. Chem. **5**, 83 (1968)
103. P.F. Pagoria, G.S. Lee, A.R. Mitchell, R.D. Schmidt, Thermochim Acta **384**, 187 (2002)
104. J. Zhang, J.M. Shreeve, J. Am. Chem. Soc. **136**, 4437 (2014)
105. Y. Tang, J. Zhang, L.A. Mitchell, D.A. Parrish, J.M. Shreeve, J. Am. Chem. Soc. **137**, 15984 (2015)
106. I.V. Ovchinnikov, N.N. Makhova, L.I. Khemelniskii, V.S. Kuzman, L. Akimova, V. Pepkin, Doklady Chem. **359**, 67 (1998)
107. A.B. Sheremetev, I.L. Yudin, Russ. Chem. Rev. **72**, 87 (2003)
108. A.B. Sheremetev, V.L. Korolev, A.A. Potemkin, N.S. Aleksandrova, N.V. Palysaeva, T.H. Hoang, V.P. Sinditskii, K.Y. Suponitsky, Asian. J. Org. Chem. **5**, 1388 (2016)
109. A.B. Sheremetev, I.L. Yudin, Mendeleev Commun. **6**, 247 (1996)
110. R.L. Willer, D.W. Moore, J. Org. Chem. **50**, 5123 (1985)
111. I.B. Starchenkov, V.G. Andrianov, A.F. Mishnev, Chem. Heterocycl. Compd. **33**, 1355 (1997)
112. X. Jin, M. Xiao, J. Zhou, B. Hu, Beilstein Arch. 202080 (2020)
113. J. Ma, A.K. Chinnam, G. Cheng, H. Yang, J. Zhang, J.M. Shreeve, Angew. Chem. Int. Ed. **60**, 5497 (2021)



Energy Sources, Part A: Recovery, Utilization, and Environmental Effects

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ueso20

Performance enhancement of savonius turbine with the application of reorienting blade mechanism

Prashant Chandra Pujari, Amit Jain, Devang S. Nath & Naveen Kumar

To cite this article: Prashant Chandra Pujari, Amit Jain, Devang S. Nath & Naveen Kumar (2021): Performance enhancement of savonius turbine with the application of reorienting blade mechanism, Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, DOI: 10.1080/15567036.2021.1948635

To link to this article: https://doi.org/10.1080/15567036.2021.1948635

Published online: 12 Jul 2021.

Submit your article to this journal

Article views: 24

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ueso20

ENERGY SOURCES, PART A: RECOVERY, UTILIZATION, AND ENVIRONMENTAL EFFECTS
https://doi.org/10.1080/15567036.2021.1948635

Taylor & Francis
Taylor & Francis Group

Check for updates

Performance enhancement of savonius turbine with the application of reorienting blade mechanism

Prashant Chandra Pujari, Amit Jain, Devang S. Nath, and Naveen Kumar

Department of Mechanical Engineering, Delhi Technological University, Delhi, India

ABSTRACT

The present study aims to the efficiency of conventional Savonius turbines by virtue of active aerodynamics, and a novel reorienting blade mechanism is designed and implemented on the existing turbine. To analyze the performance of the current design, five different blade configurations have been studied in compliance with varying wind speed and tip speed ratios. Two-dimensional CFD simulations are performed by applying the URANS model along with a Realizable k-ε turbulence model. The available experimental data are used to validate the results obtained from the numerical simulations. The findings indicated a considerable improvement in the coefficient of power for blade angles 10° and 20° in comparison with the standard Savonius turbine (0°). The maximum increase of 9.5% and 7.6% in the power coefficient is obtained for blades at 10° and 20° positions. It is also found that during extreme wind conditions, turbine blade configurations of 30° and 40° are best suited to prevent the turbine from failing. The proposed design enhances the power coefficient and ensures the Savonius turbine's safety by changing the blade orientation according to the wind conditions.

ARTICLE HISTORY
Received 10 March 2021
Revised 8 June 2021
Accepted 22 June 2021

KEYWORDS
Wind energy; Savonius turbine; CFD; VAWT; Renewable energy

Introduction

The rise in the global population has contributed to a substantial increase in energy consumption in recent decades (Edelenbosch et al. 2017), which has exhausted the nonrenewable resources and increased the rate of global warming and greenhouse gas production (Bonalumi and Giuffrida, 2016). For more sustainable growth, renewable energy such as solar energy and wind energy may be adopted as solutions to minimize the usage of fossil fuels and limit its adverse effects on the environment. Wind power is an abundant, clean, and cheap renewable energy source (Yannopoulos et al. 2015). Today, wind energy is developed in electrical power generation as a mainstream source of energy.

Wind turbines are primarily classified on the basis of the rotation axis as HAWTs and VAWTs (Sharma and Sharma 2017; Wang, Yabo, Ling Huang, and Raymond Lau 2016). While HAWTs are used to produce power at a very large scale, VAWTs are viable during light wind speed for small-scale power generation (Mari, Venturini, and Beyene 2017; Rezaeiha, Kalkman, and Blocken 2017). VAWTs have, therefore, recently earned an increasing interest in extracting energy in urban environments because of their uncomplicated design, low production, installation, and servicing expenses (Battisti et al. 2011; Mandal, Rana, and Tripathi 2020; Pourhoseinian, Sharifian, and Asasian-Kolur 2021). Besides, VAWTs can harness wind energy under highly unstable and turbulent urban environments (Chen et al. 2017; Persico et al. 2017). The comparison between the C_p and λ for various kinds of wind-driven turbines as shown in Figure 1. VAWTs are further classified into the Savonius turbine and

CONTACT Prashant Chandra Pujari prashantpujari05@gmail.com Department of Mechanical Engineering, Delhi Technological University, Delhi110042, India

© 2021 Taylor & Francis Group, LLC

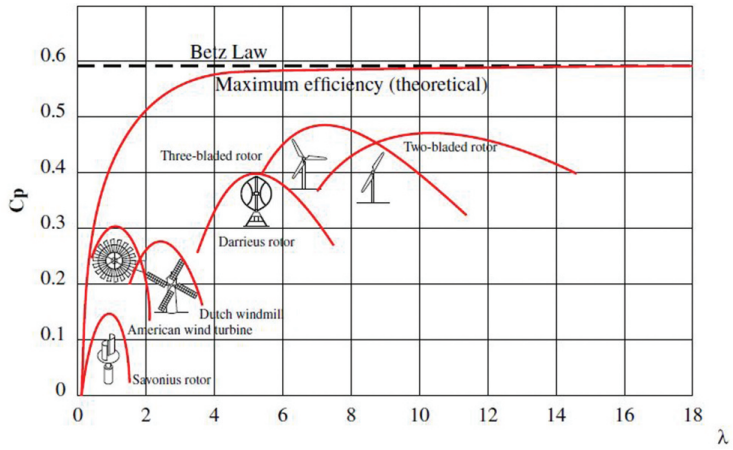


Figure 1. Tip speed ratio vs power coefficient for various types of wind turbines (Wood 2011).

Darrieus turbine (Roshan, Sagharichi, and Maghrebi 2020). Savonius turbine generates high positive torque due to its symmetrical design in any wind direction compared to the Darrieus turbine. It also operates quietly at a lower λ (Goodarzi and Keimanesh 2015; Laws et al. 2020). Despite these advantages, there are various issues that need to be addressed to improve the performance of the Savonius turbines. Researchers have dived into various factors such as the effect of overlap ratio, number of blades, number of stages, blade designs, guide vanes, etc., to enhance the efficiency of Savonius turbines.

A study for an OL = 0.15 and AR of 1.0 and 1.5 is conducted to test the efficiency of the Savonius turbine (Sheldahl, Blackwell, and Feltz 1978). Their findings indicated that the torque and power coefficient for AR = 1.5 was marginally higher than that for 1.0. Another three-dimensional flow study on the Savonius rotor for 0.2 overlap ratio is performed with four different aspect ratios of 0.55, 0.78, 1.10, and 1.66 (Ferrari et al. 2017). It was observed that with the increase in the aspect ratio, the power coefficient followed an increasing trend. In comparison to those without overlap, traditional Savonius turbines having overlapping blades are reported to have better performance and self-starting capacity (El-Askary et al. 2020). The concave region of the advancing blade allows the wind to travel into the overlap region formed by the semi-circular curvature of the blades. The pressure on the downstream face of the returning blade increases due to the overlap area, and the pressure difference between the concave and convex sides decreases in return. This leads to a decrease in the resultant torque produced by the preceding blade and an increase in the total torque acting on the turbine (Zhao et al. 2009).

As the number of blades increased, the variations in the coefficients of static and dynamic torque of a standard Savonius turbine are observed to be lowered, resulting in the reduction of the power coefficient (Saha, Thotla, and Maity 2008). To determine the optimum number of blades, (Mahmoud et al. 2012) experimentally tested the output of single and double-stage Savonius rotors without considering any overlap. It was found that for both single- and double-stage, two-bladed rotors are more efficient than multi-bladed rotors.

The drawbacks of the traditional Savonius turbine incorporated considerable variations in the value of torque coefficient, producing negative values for some blade angles (Akwa, Vielmo, and Petry 2012). To enhance the efficiency of the turbine and reduce the thrust load, an experimental study was conducted with numerical validation considering the various stages of the turbine (Mohan Kumar et al. 2019). The results showed that the three stages of the Savonius turbine achieved greater self-starting capacity and a substantial 58% decrease in thrust load. However, compared with a single-stage rotor, a significant decrease in the coefficient of power was observed by the application of a multi-stage turbine.

After observing all the aforementioned factors affecting the performance and safety of the turbines, there is still an unexplored area of application of active aerodynamics in the Savonius turbines. Active Aerodynamic control systems can adaptively adjust the wind turbine's aerodynamic characteristics

under varied operating conditions to achieve the desired performance. The present study introduces a new concept of reorienting mechanism as an application of active aerodynamics on the Savonius turbine. The real-time value of the wind speed recorded by the anemometer is used to actuate the blades of the turbine to orient themselves at the most favorable position to extract maximum available wind power without compromising the safety of the system. Based on this concept, a reorienting Savonius turbine is designed in which different blade configurations can be obtained with blade positions varying from 0° to 40°. A two-dimensional, incompressible URANS model in compliance with the Realizable k–ε turbulence model is developed. The reorienting turbine's performance is analyzed for various blade angles (θ) and wind speed (V) at different λ. The aerodynamic performance is assessed using pressure distribution and velocity flow fields around the turbine. Subsequently, the results obtained are mentioned for potential implementation in the real world.

Design and mechanism

The geometry of the turbine is adapted from a traditional single-stage Savonius rotor, which consists of multiple semi-circular blades. From the experimental investigation performed by (Saha, Thotla, and Maity 2008), it is being observed that a two-bladed rotor provided a better performance as compared to a multi-bladed system. The various parameters of the rotor are derived from Hayashi et al. (Hayashi, Yan, and Hara 2005). Figure 2 presents the design and specifications of the turbine. This design is modified by adding a reorienting mechanism, as shown in Figure 3. Table 1 depicts the list of the parameters used for designing the new Savonius turbine. The overlap ratio (OL) is taken to be 0.2, which ranges from 0.15 to 0.25 (Sheldahl, Blackwell, and Feltz 1978; Ushiyama, Nagai, and Shinoda 1986). The swept area (A) is calculated to be 0.0759 m^2.

A few modifications were implemented in the design in SOLIDWORKS 2020 to obtain the desired operation of the blade rotation around their axis. Firstly, the bottom cover plate was replaced with a semi-circular internal spur gear; secondly, a center shaft is placed on top of the transmission shaft on which a spur gear is mounted, as shown in Figure 3, which is used to control the movement of the blades with the rotation of the center shaft. The top cover plate was also modified to prevent the

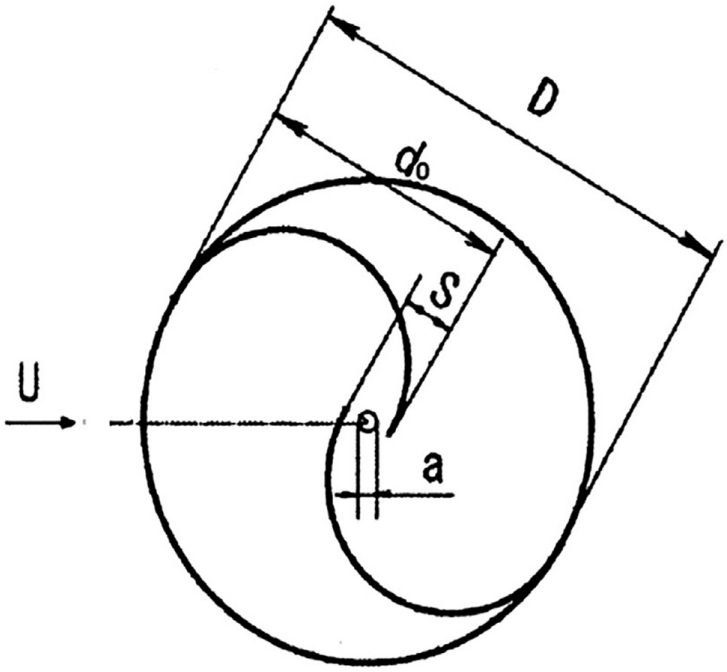


Figure 2. Geometry of the standard Savonius rotor (Hayashi, Yan, and Hara 2005).

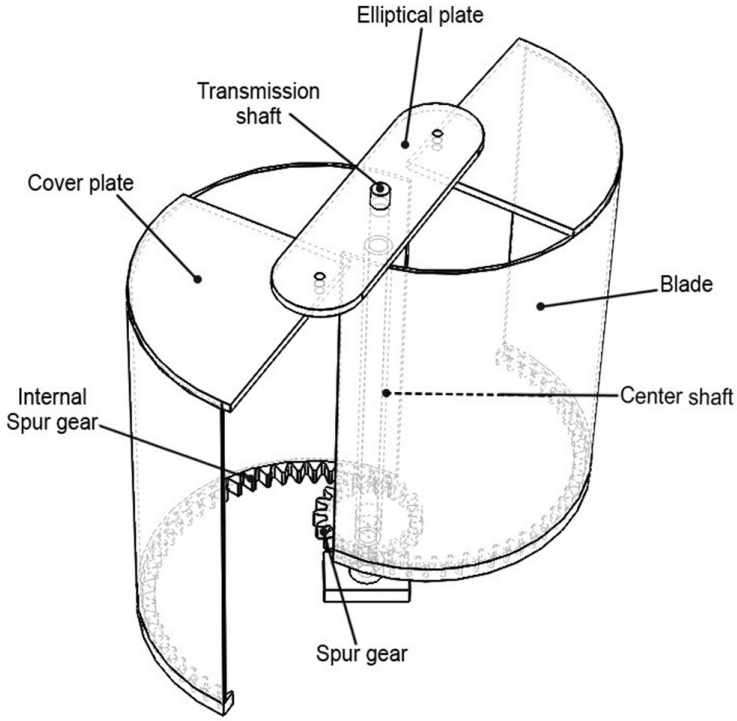


Figure 3. Reorienting Savonius Turbine.

Table 1. Design Parameters.

D	0.33 m
S	0.066 m
H	0.23 m
b	0.01 m
t	0.002 m
a	0.015 m
d_o	0.198 m
A	0.0759 m^2
m	3 mm
w	10 mm
Φ	20°
t_i	62
t_c	19

interference of the plates, and rotational unbalance of the turbine. Furthermore, an elliptical plate is used to join the blades, and the transmission shaft generated the power at the desired orientation. The blade rotation mechanism was controlled by servo motors to restrict the motion between 0° and 40°. For the sake of time and resources, the simulation was limited to five different orientations as shown in Figure 4.

Numerical simulation

Boundary conditions

The simulation consists of two main parts, the inner and outer domains. The first one simulates the rotor mechanism with the change in blade angles, and hence it is a rotating domain in a circular shape. The size of the inner domain is 1.75 times larger than the diameter of the rotor. The second part is appropriately sized to evade any effects of boundary conditions on the simulation results. It is a stationary domain and square in shape. Figure 5 shows the relationship between the minimum

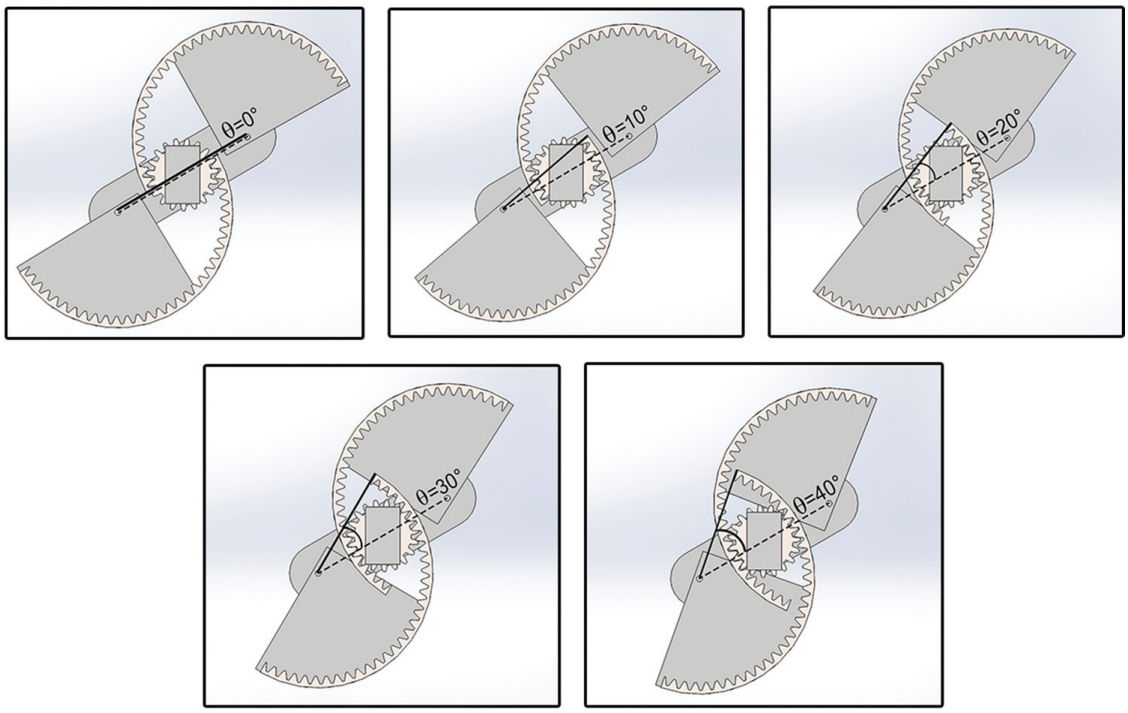


Figure 4. Different blade positions of the reorienting turbine.

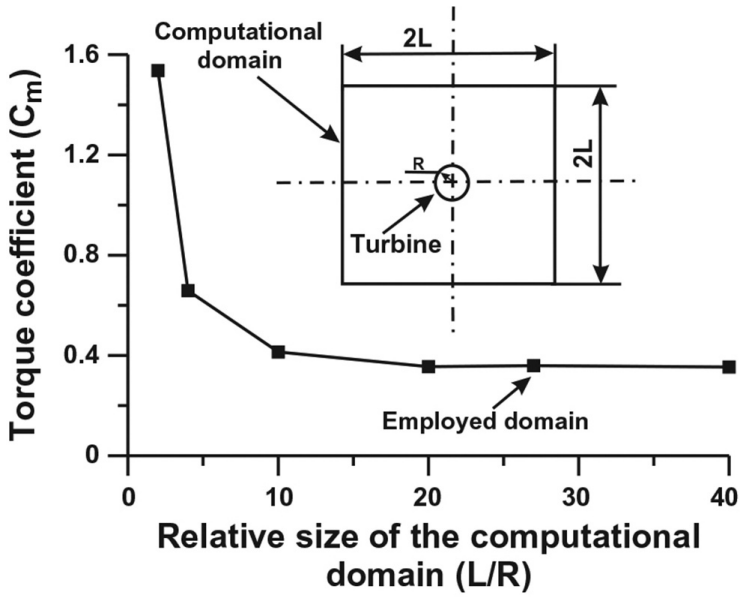


Figure 5. The variation in torque coefficient due to domain size (Mohamed et al. 2011).

relative size of the outer domain and the torque coefficient. In this current numerical analysis, the length, L of the domain, is kept as 12D.

The simulation domain is partitioned into two parts, the rotating zone around the rotor and the stagnant zone, as shown in Figure 6. An interface segregates these two zones to validate the steadiness of the flow field. The stagnant outer domain consists of a velocity inlet and pressure outlet at the two ends. In the velocity inlet, the speed of the wind ranges from 6 m/s to 15 m/s in the x-direction, and the

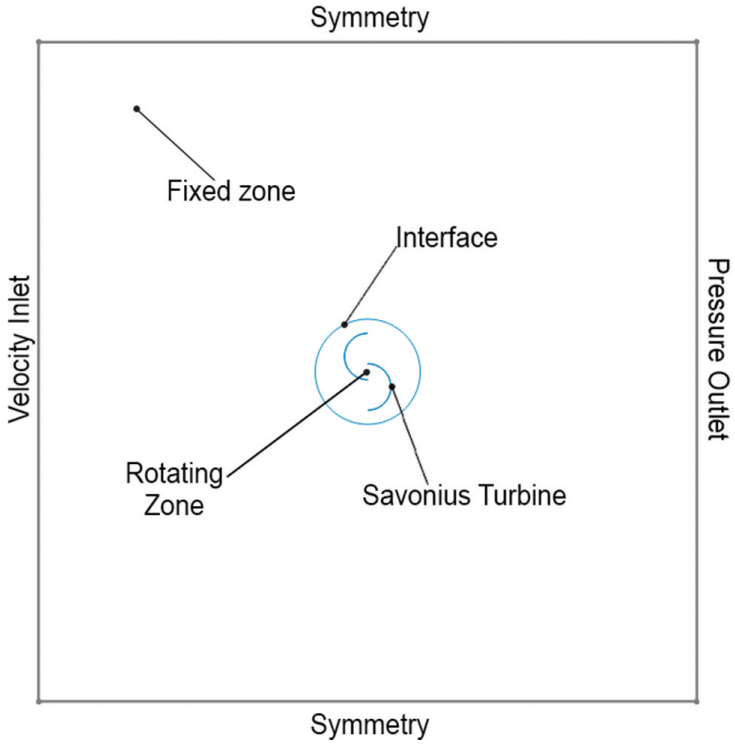


Figure 6. Computational Domain.

outlet is maintained at standard atmospheric pressure. The walls of the outer domain at the top and bottom have symmetric boundary parameters. The inner domain rotates in a counter-clockwise direction with a time step size of 0.0005 s for all 2D simulations.

The angular velocity of the revolving inner zone is calculated by the term λ, obtained by the ratio between the tangential velocity of the tip of the blade and the undistributed wind speed. The following equation is used to calculate λ:

$$\lambda = \frac{V_t}{V} = \frac{\omega \times R}{V} \tag{1}$$

Savonius turbine is driven by various λ for various wind speed V and the resulted angular velocity ω for each case is shown in Table 2. In the present simulation, six different values of λ from 0.5 to 1 are used for four different wind speed at the inlet, which resulted in 24 cases for each blade angle and, hence, 120 cases for five different blade angles 0°, 10°, 20°, 30°, and 40°.

Table 2. Angular velocity for different wind speed and tip speed ratio.

λ / V	0.5	0.6	0.7	0.8	0.9	1
6	18.18	21.82	25.45	29.09	32.73	36.36
9	27.27	32.73	38.18	43.64	49.09	54.55
12	36.36	43.64	50.91	58.18	65.45	72.73
15	45.45	54.55	63.64	72.73	81.82	90.91

Mesh Generation

A grid convergence study is conducted to determine an efficient mesh for the simulation. This is to ensure that the further refinement of the present mesh does not significantly influence the results. Three different 2-Dimensional mesh models, namely coarse, medium, and fine, are developed to study

mesh independence. The refinement is employed based on size reduction of the inner rotating and outer static domains to obtain acceptable mesh quality. The URANS realizable k-ε turbulence model with the standard wall function is used. The details of all the three grids used in the grid convergence study are shown in Table 3.

Table 3. Mesh sizing.

	Refinement		
	Coarse	Medium	Fine
Number of Elements	83572	139863	300977
Number of Nodes	84377	140859	302340
Rotating Domain	2.32E-02	1.52E-02	1.02E-02
Static Domain	1.21E-01	9.09E-02	6.06E-02
Y+ Value		150	

The simulation is carried out for five complete rotations of the turbine at 0-degree blade position with a time step size of 0.0005 s. The inlet wind speed of 9 m/s is used, and λ vary from 0.5 to 1. The results obtained for all three grids are shown in Figure 7. It is evident from the graph that the values of the power coefficient do not change significantly as the mesh is refined. The average difference between the medium and fine mesh results is only 1.50% across the λ; therefore, the medium mesh sizing is used for all the 2-D simulations to obtain acceptable results with the advantage of less computation time.

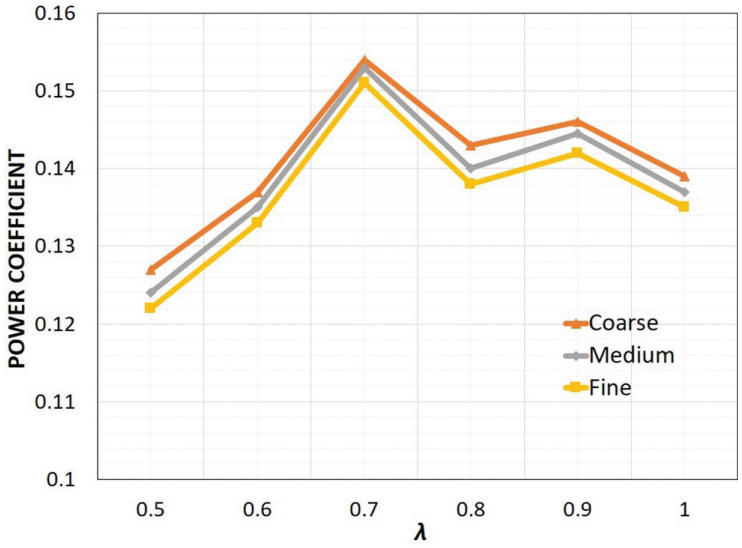


Figure 7. Results of the 2-D grid convergence study.

Using a medium unstructured grid with triangular and quadrilateral elements, the whole computational domain mesh is constructed. The moving mesh model is chosen in the present analysis to depict the movement of the Savonius turbine. The inner rotating domain has finer mesh compared to the outer stationary domain, as shown in Figure 8. The uniform mesh element size is used along with the interface of these two domains. It is observed that there is no deviation in the flow at the rotor–stator interface due to a difference in mesh sizing. The inflation is applied around the turbine blades to accommodate the variations due to the boundary layer gradient. Figure 9 shows the 20 layers of inflation with a growth rate of 1.2 used around nine turbine edges. The mesh for the realizable $k-\epsilon$

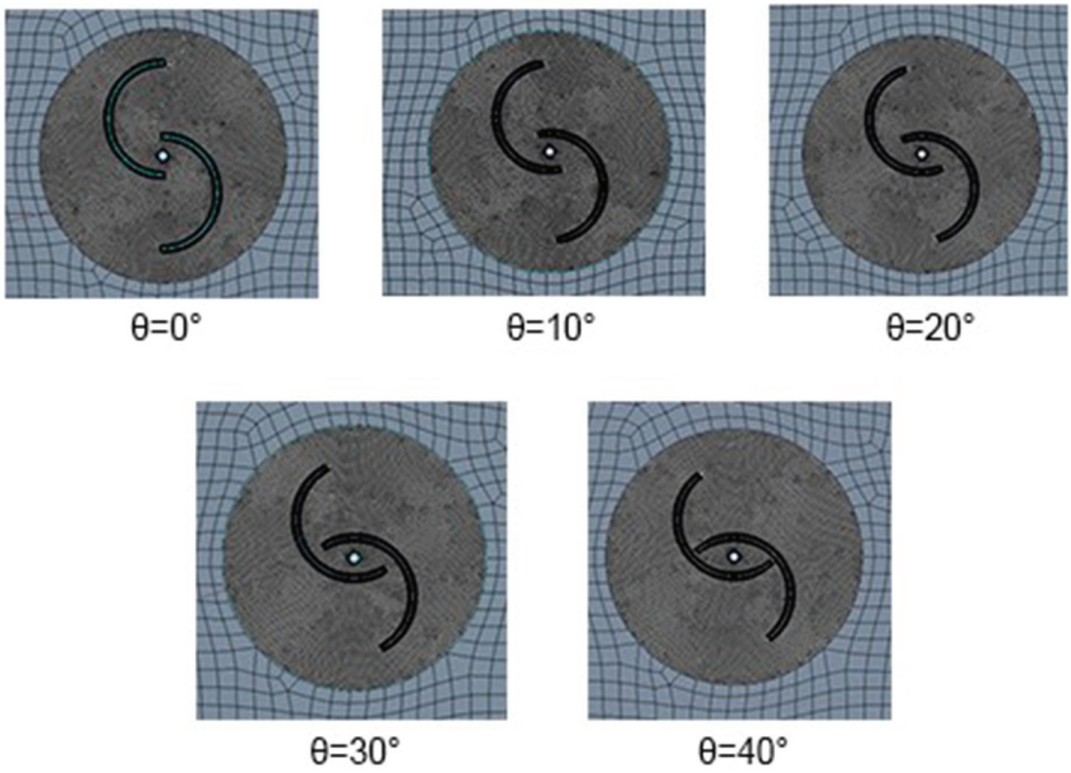


Figure 8. Two-dimensional meshing of the turbine.

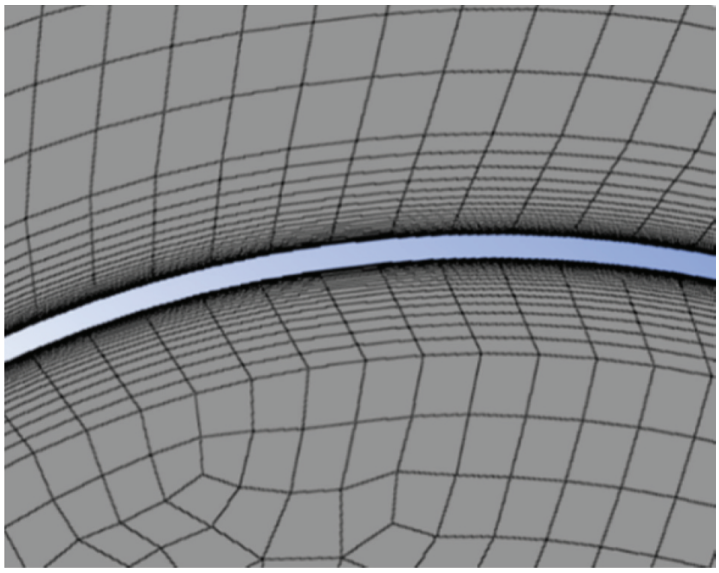


Figure 9. Inflation layers around turbine blades.

model with standard wall function has a target y+ in the range of 30 to 300 for the buffer layer and more than 300 for the outer layer (Larin, Paraschivoiu, and Aygun 2016). The y+ value under the present study has been taken as 150 to calculate the height of the first layer from the wall of the turbine. Indeed, a mesh independence study is presented to increase the sincerity of the numerical model.

Turbulence modeling

The Reynolds-averaged Navier-Stokes (RANS) model is employed, and the governing equations for unsteady Newtonian incompressible for turbulent flow are numerically simulated using ANSYS

FLUENT R2020. The SIMPLE algorithm is used for pressure–velocity coupling (Saad et al. 2019). Moreover, for all the turbulence, momentum, and pressure equations, the second-order upwind spatial discretization algorithm is used, and the least-squares cell-based algorithm is set for gradients to obtain maximum accuracy (Saeed et al. 2019). Previous research suggested that a turbulence intensity value equal to or less than 1% is considered low, whereas a value greater than 10% is generally considered high (Kumar and Saini 2017a, 2017b). Thus, the default values of 5% and 10% are used for the turbulence intensity and viscosity ratio, respectively, at the inlet and outlet boundary conditions (Mosbahi et al. 2019). The convergence criteria for all residuals equations of continuity, momentum, and turbulence are fixed equal to 10^{-5}. The simulation is run for five complete rotations of the turbine with a time step size of 0.0005 s. It corresponds to the rotor rotation 1° for each time step, which is employed with 30 iterations per time step.

The moment coefficient is calculated from the moment value obtained from the rotating turbine using the equation:

$$C_m = \frac{PitchingMoment}{DynamicMoment} = \frac{M}{0.5\rho ARV^2} \tag{2}$$

The coefficient of power of the Savonius turbine is determined as the ratio of the rotor power to the available wind power (Akwa, Vielmo, and Petry 2012) and expressed by:

$$C_p = \frac{P_{rotor}}{P_{wind}} = \frac{P}{0.5\rho AV^3} \tag{3}$$

The coefficient of power can also be derived from the moment coefficient using the following relation:

$$C_p = C_m \times \lambda \tag{4}$$

Validation

Turbulence model validation

A comparative analysis between the results obtained and the existing experimental data is used to validate the turbulence model to determine the best possible turbulence model for the study. Figure 10 shows the simulated results for different turbulence models and the experimental findings of Hayashi et al. (Hayashi, Yan, and Hara 2005). The turbulence models used are the SST k-ω, the standard k–ε,

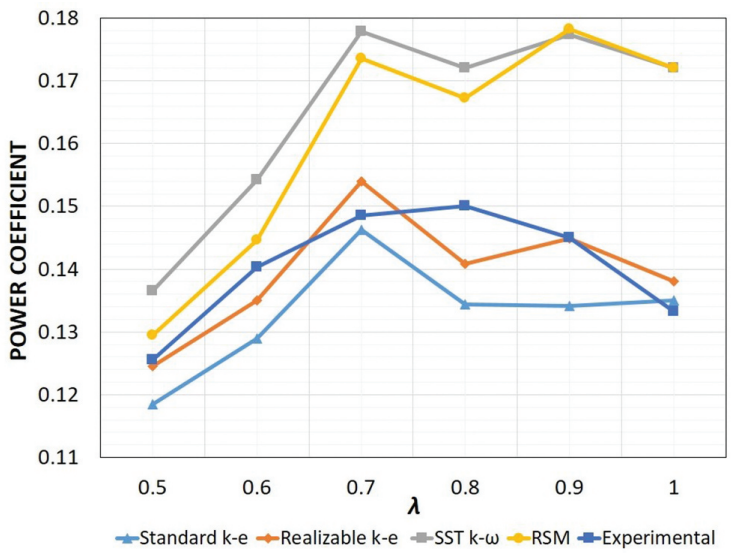


Figure 10. Validation of turbulence models.

the RSM, and the realizable k–ε. The simulated and experimental results are compared for 0° blade angle at 9 m/s wind speed for $\lambda = 0.7$.

The simulated findings are compliant with the experimental results of Hayashi et al. (Sheldahl, Blackwell, and Feltz 1978), and the correlation coefficient (R^2) (Tahani et al. 2017) is applied to determine which turbulence model has more accuracy in contrast with the experimental findings. Table 4 shows the values of the correlation coefficient for the four turbulence models used. It is found that the realizable k–ε turbulence model gives the best results among other models having the highest R^2 of 0.828. Hence, a realizable k–ε turbulence model is employed for the entire simulation.

Table 4. The turbulence models and their correlation coefficient (R^2).

Turbulence Model	R^2
SST k–ω	0.775
Standard k–ε	0.763
RSM	0.688
Realizable k–ε	0.828

2-D model validation

To study the differences in 2-D simulation results with the 3-D model, a three-dimensional CFD model of the turbine is simulated in the domain, as shown in Figure 11. The turbine with 0-degree blade orientation is placed inside the inner domain, which is programmed as a moving mesh, while the outer domain is taken sufficiently large enough so that the effects of walls do not influence the results. The standard medium-mesh sizing is applied in the domain with inflation near the blades, resulting in 1.5 million elements. The URANS realizable k–ε turbulence model with standard wall function is employed. The simulation is carried out for five complete rotations of the turbine using a time step size of 0.0005 s with the variable number of time steps according to λ. The inlet wind speed of 9 m/s is used in addition to λ varying from 0.5 to 1.

The results of the 3-D simulations are plotted against experimental investigations of Hayashi et al. with the findings of the 2-D CFD model as shown in Figure 12. The values of power coefficients are observed to be comparable with the experimental findings, which justified the turbulence model. The above CFD results of the 3-D model are only shown for five cases, while in total, 120 cases need to be simulated for five different orientations and four different wind speeds. The comparison of power coefficients obtained by CFD simulation for both 2-D and 3-D models is shown in Table 5. It can be seen that the 3-D CFD model predicts lower values of power coefficient than the 2-D model. This is due to the presence of endplate friction in the 3-D model, which has not been taken into account during the 2-D CFD simulation. The presence of the endplate has a negative effect on the power coefficient, which can be seen from Table 5, but the effect is minimal and can be subdued (Marinić-Kragić, Vučina, and Milas 2020). It is also inferred that the implementation of 2-D models is

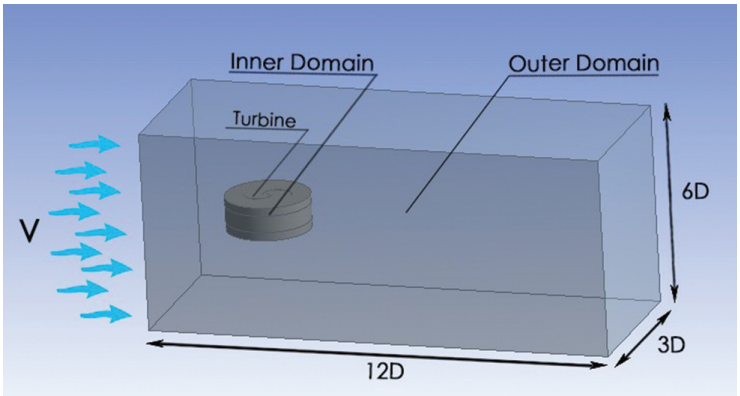


Figure 11. 3-D computational domain

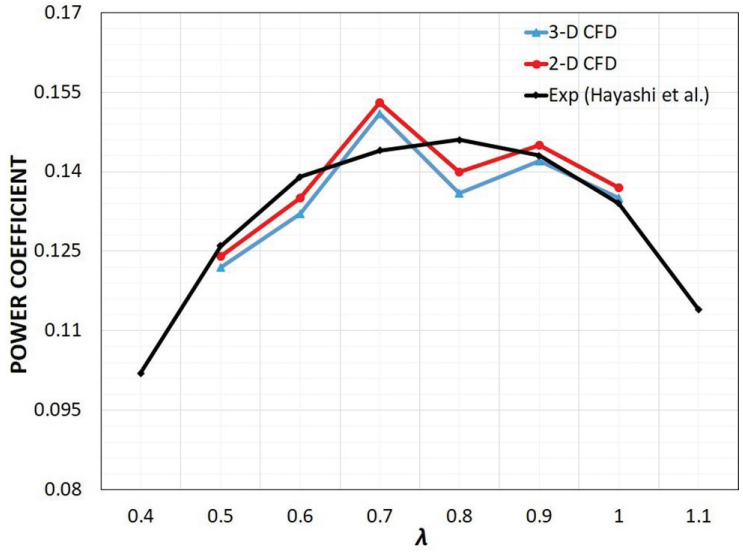


Figure 12. Comparison of power coefficient of 2-D and 3-D CFD models with experimental data.

Table 5. 2-D Simulation Results Validation.

λ	Results		
	3-D CFD	**2-D CFD**	**% Error (-)**
0.5	0.122	0.124	1.64
0.6	0.132	0.135	2.27
0.7	0.151	0.153	1.32
0.8	0.136	0.14	2.94
0.9	0.142	0.145	2.11
1	0.135	0.137	1.48

permissible with the additional benefits of less computational time and resource requirements compared to 3-D simulations (Rezaeiha, Montazeri, and Blocken 2019).

Figure 13 shows the moment coefficient variation with flow time for 9 m/s wind speed at $\lambda = 0.7$. The simulation for each case was run for five complete rotations. It is witnessed that at the later part of the second period, the moment coefficient values become approximately periodically sustained. The average values of the moment coefficient for the subsequent cycles came out to be nearly equal, and hence the third rotation is used to calculate the values of the moment coefficient for each case, and further, these values are used for calculating the power coefficient using Eq. (4).

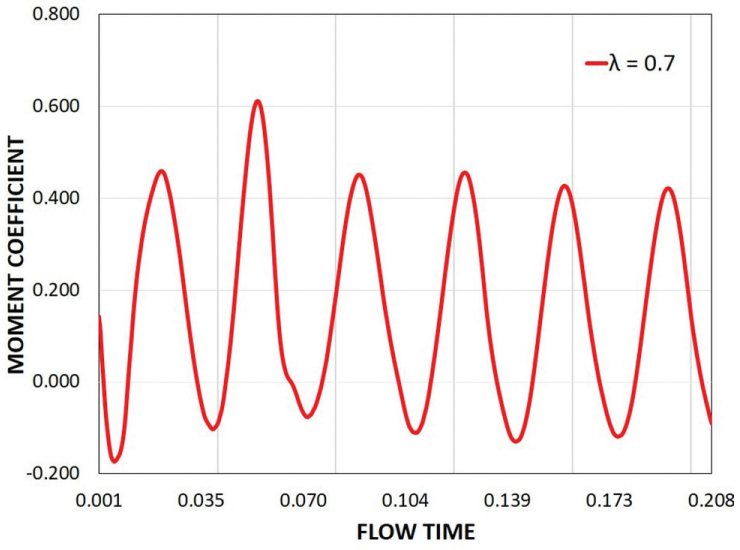


Figure 13. Moment coefficient vs flow time.

Results and discussion

The performance parameter of the Savonius wind turbine depends on the coefficient of power of the turbine. Thus, the effects of the geometrical parameter that is blade angles (θ) ranged from 0° to 40°, and wind speed (V) ranging from 6 to 15 m/s along with tip speed ratio (λ) varied from 0.5 to 1 on the output parameters are discussed below:

Effect of blade angle

Variation in the power coefficient for various blade angles with λ at wind speed ranged from 6 m/s to 15 m/s are presented in Figure 14. With the rise in wind speed in the range of 0.5 to 0.7 tip speed ratio, the value of C_p continues to increase. For λ = 0.7, the highest value of C_p is obtained for all blade configurations at a wind speed of 15 m/s. The maximum value of C_p is observed to be 0.169 for θ = 10° for V = 15 m/s, which shows a significant increment of 9.5% in power coefficient compared to

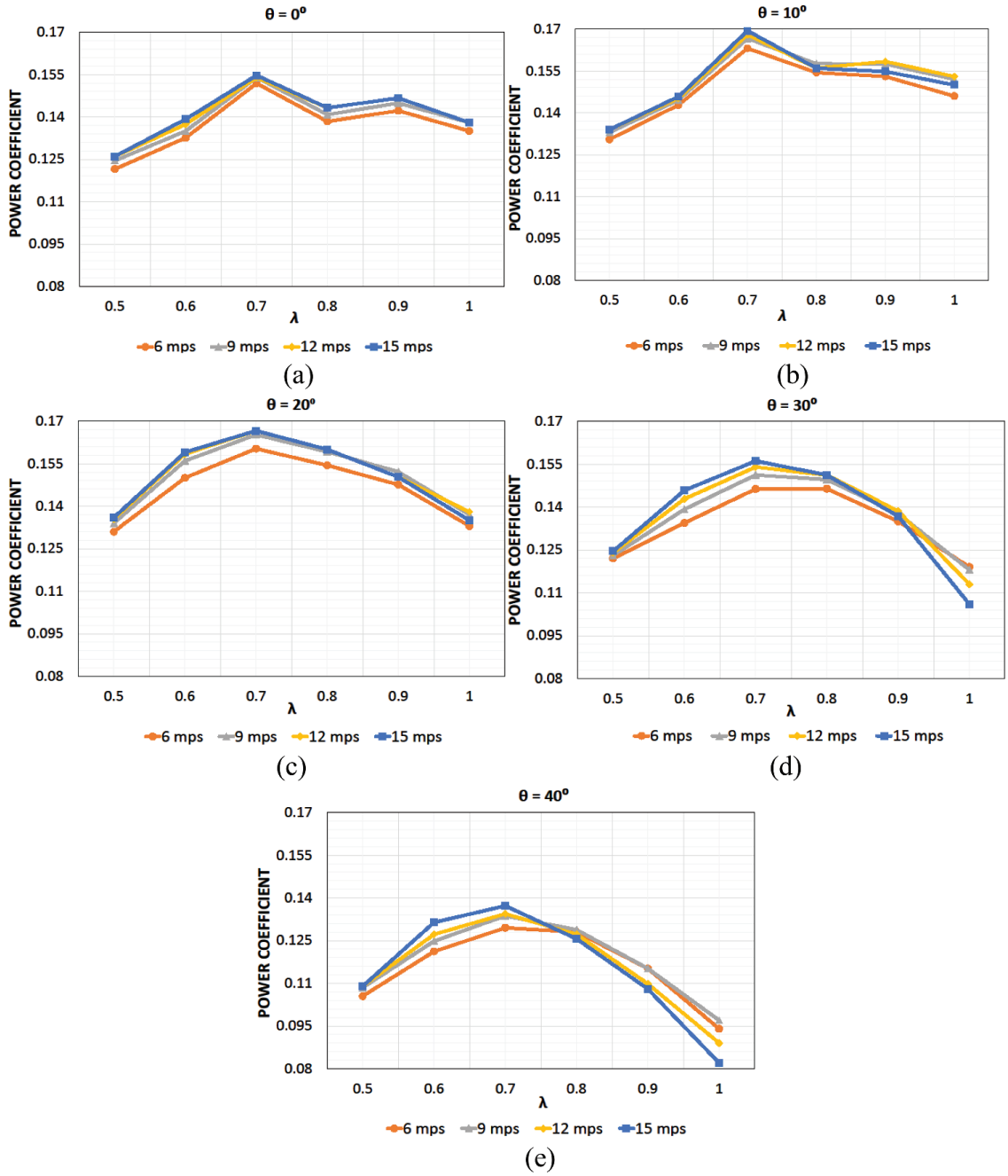


Figure 14. The trend of power coefficient with wind speed for different blade angles (a) θ = 0° (b) θ = 10° (c) θ = 20° (d) θ = 30° (e) θ = 40°.

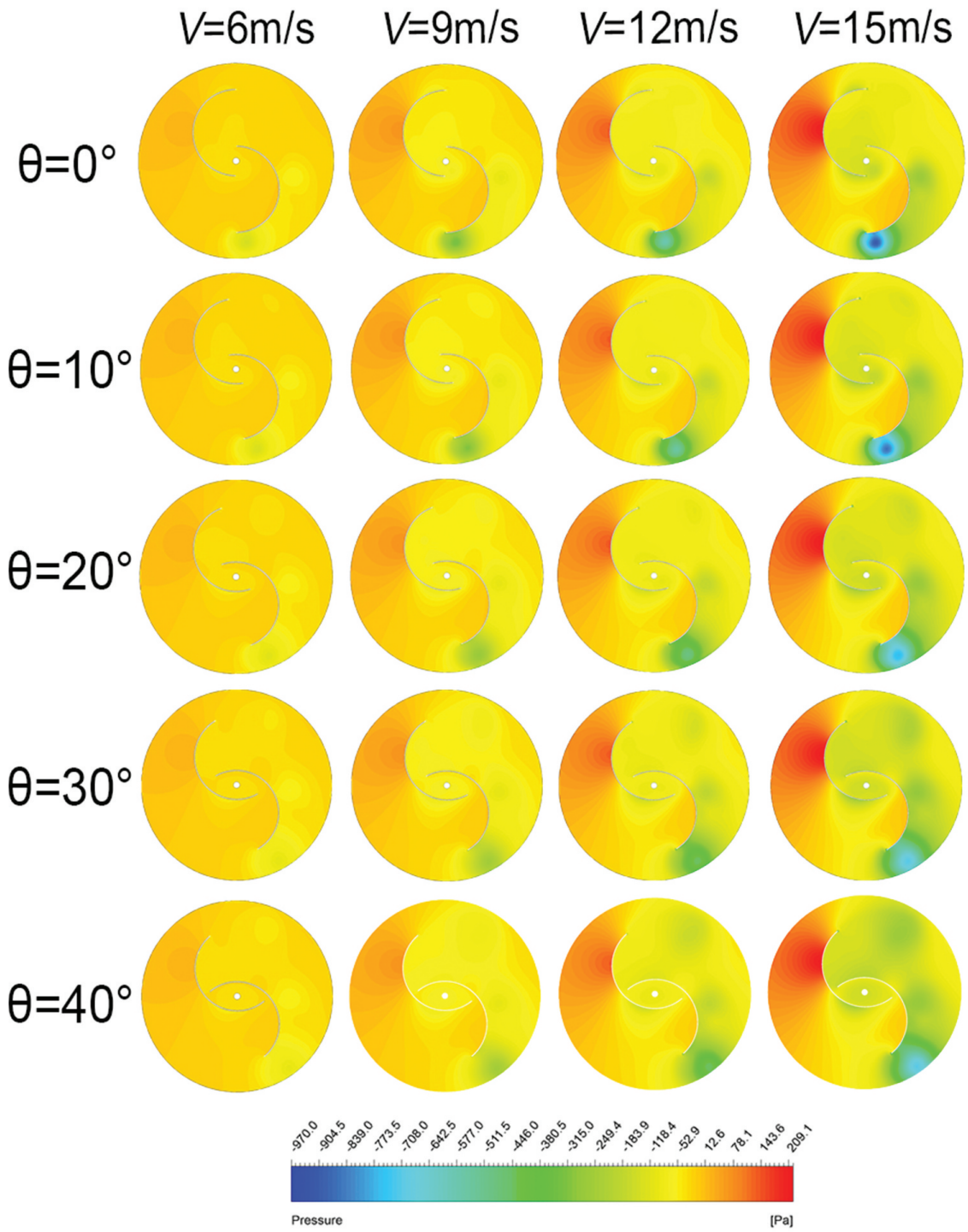


Figure 15. The total sectional pressure distribution around the turbine at $\lambda = 0.7$ for all blade positions (θ) and wind speed (V).

traditional Savonius turbine where blades are at θ = 0°. Also, it can be seen that for all blade positions, there is a continuous reduction in the values of C_p for λ higher than 0.7.

Figure 15 illustrates the pressure contours for all the five blade positions for different wind speeds at $\lambda = 0.7$. The distribution of pressure around the turbine blades is observed to be similar for all configurations and wind speeds. It is found that the absolute wind pressure is maximum at the advancing blade's tip, which reduces downstream along with the convex surface of the rotor blade. This variation in the wind pressure around the geometry of the turbine blade causes flow separation at the end of the approaching blade due to negative pressure gradients, which are noticed in all the cases for a specific blade angle.

Effect of wind speed

The reorienting Savonius turbine can position its blades according to prevailing wind conditions to harness maximum power from the wind using the results predicted in Figure 16. The general trend of C_p remains almost the same for each blade configuration in the studied wind speed range throughout all tip speed ratios. For each value of wind speed and λ, the values of C_p for $\theta = 10°$ and 20° are equal or more than that of standard Savonius turbine ($\theta = 0°$), which indicates the advantage of reorienting mechanism in the turbine. For lower λ such as 0.5 and 0.6, turbine blades at 20° position produce higher power coefficient, and beyond these values of λ, 10° blade position becomes more efficient as seen from Figure 16. It can also be found that for higher wind speed, the torque experienced by the turbine rotor at $\theta = 30°$ and 40° is lower as compared to other blade positions. Hence, these configurations are best suited to prevent the turbine from failing during extreme wind conditions.

It can be seen that the turbine can achieve a maximum value of C_p for $\theta = 10°$ in all the studied wind speeds due to the variations in the air mass flow around the advancing and returning blades. Figure 17 illustrates the velocity vectors for all the five blade positions at 15 m/s wind speed for $\lambda = 0.7$. The amount of wind hitting the concave surface of the approaching blade for all the orientations is higher than that of $\theta = 0°$, which leads to more positive torque. Moreover, for 10° and 20° blade positions, the separated flow in the overlapping region strikes the returning blade to add the overall positive torque, resulting in a higher value of power coefficient than at 0°. The reason for the least values of C_p for $\theta = 40°$ in all the wind conditions is the presence of a highly unstable zone behind the advancing blade, which increases the negative torque. This is the reason for obtaining the highest value of C_p at 10° blades position at $\lambda = 0.7$ for all wind speed as there is some turbulent region present behind the advancing blade in 20° case which lowers the positive torque to rotate the turbine. This is the reason for obtaining the highest value of C_p at 10° blades position at $\lambda = 0.7$ for all wind speed as the effect of

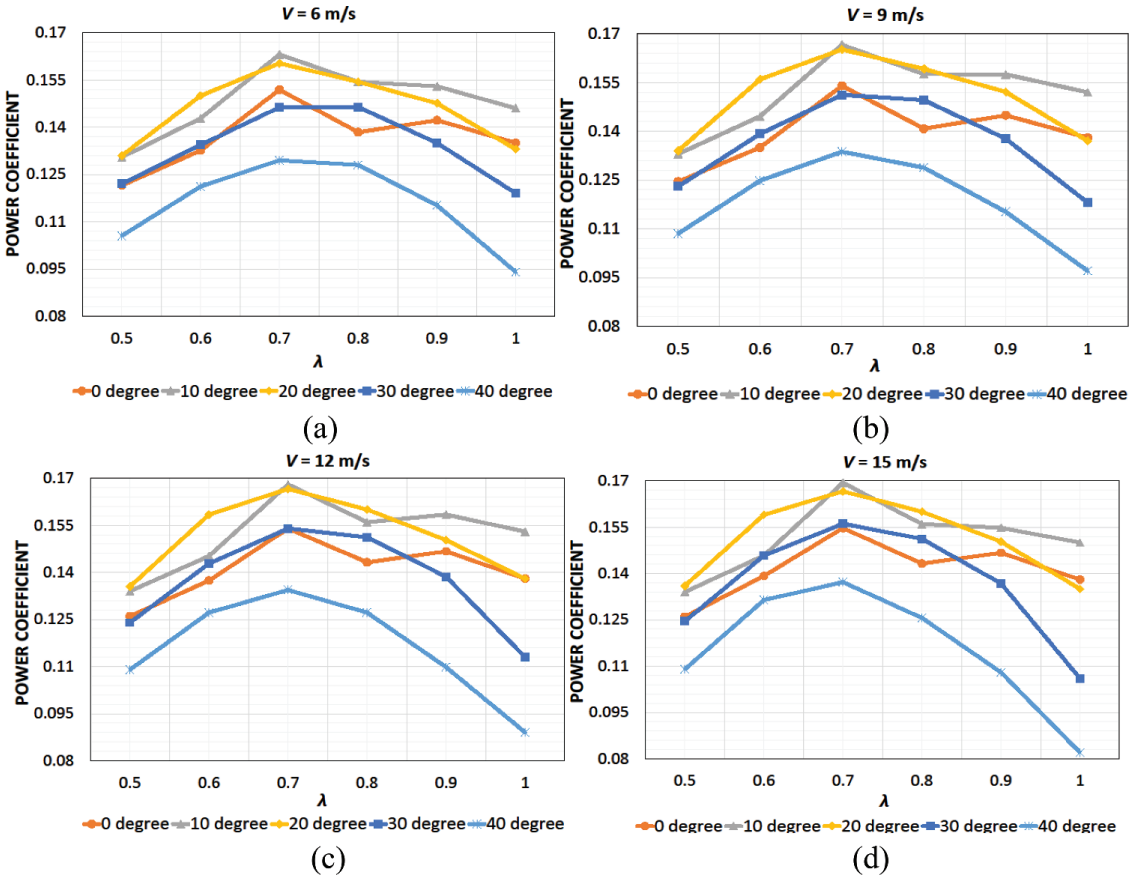


Figure 16. The variation of average power coefficient at different blade positions for wind speed (a) $V = 6$ m/s (b) $V = 9$ m/s (c) $V = 12$ m/s (d) $V = 15$ m/s.

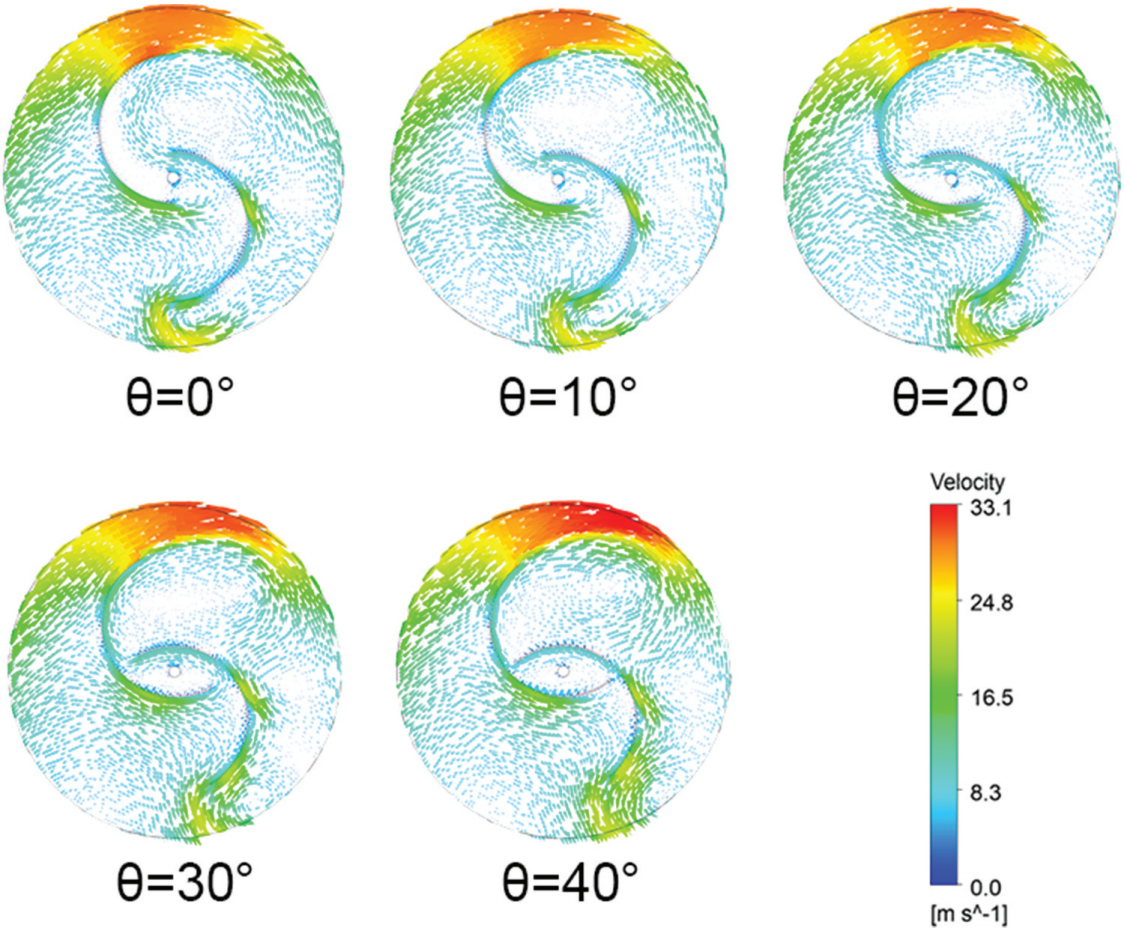


Figure 17. Sectional wind speed flow fields around the turbine at different blade configurations for V = 15 m/s and λ = 0.7.

the turbulent region present behind the advancing blade dominates for θ greater than 10° case, which lowers the positive torque to drive the turbine.

Conclusion

The effects on the performance of the reorienting Savonius turbine are explored by varying blade angle and wind speed. A two-dimensional incompressible URANS model is developed in conformation with a Realizable k–ε turbulence model with standard wall function to evaluate the effect of the suggested design on the performance of the wind turbine. The observations are summarized as follows:

(a) The Realizable k–ε turbulence model is better than other models such as standard k–ε, SST k-ω, and RSM, having the highest value of correlation coefficient (R^2) of 0.828 to existing experimental results.
(b) The value of C_p is found to be increasing with λ and reached the maximum at λ = 0.7, but for further increase in $\boldsymbol{\lambda}$, a decreasing trend is observed for all blade configurations.
(c) It is found that the power coefficient for blade angles 10° and 20° are higher than that of the standard Savonius turbine (θ = 0°), whereas C_p for 30° is comparable to the latter for all wind speed.
(d) A higher power coefficient is attained by the turbine at θ = 20° for lower λ and at θ = 10° for higher λ. The maximum value of C_p is 0.169 for θ = 10° for 15 m/s wind speed at λ = 0.7. The C_p values for θ = 10° and 20° are 9.5% and 7.6%, respectively, higher than that of θ = 0°.

(e) It should be noticed that the torque experienced by the turbine rotor at θ = 30° and θ = 40° is lower than other blade positions for higher wind speeds. Therefore, during extreme wind conditions, these configurations are ideally suited to prevent the turbine from failing.

The proposed design of the Savonius turbine provides researchers a new direction to harness wind energy by virtue of active aerodynamics using a reorienting mechanism by varying the blade positions. New studies can be performed to observe the effects of overlap ratio, the number of blades, the number of stages, etc., on the performance of this reorienting Savonius turbine.

NOMENCLATURE

A	Cross-sectional area of the rotor
a	Center shaft diameter
AR	Aspect ratio
b	Transmission shaft diameter
C_m	Moment coefficient
C_p	Power coefficient
D	Rotor diameter
d_o	Bucket diameter
H	Rotor height
HAWTs	Horizontal Axis Wind Turbines
k	Turbulent kinetic energy
m	Module
OL	Overlap ratio
R	Rotor radius
R^2	Correlation coefficient
RSM	Reynolds Stress Model
s	Overlap distance
SIMPLE	Semi-Implicit Method for Pressure-Linked Equations
SST	Shear Stress Transport
t	Blade thickness
t_c	Spur gear teeth
t_i	Internal spur gear teeth
URANS	Unsteady Reynolds-averaged Navier-Stokes
V	Wind speed
V_t	Rotor tip speed
VAWTs	Vertical Axis Wind Turbines
w	Face width
ε	Turbulent dissipation rate
θ	Blade angle
Φ	Pressure angle
λ	Tip speed ratio
ρ	Air density
ω	Angular speed

Notes on contributors

Prashant Chandra Pujari is working as a Research Scholar at the Centre for Advanced Studies and Research in Automotive Engineering, Dept. of Mechanical Engineering, Delhi Technological University. He did his B.Tech in Mechanical Engineering from Delhi Technological University, Delhi, India. His research interests include; renewable energy, fluid mechanics, aerodynamics, and computation fluid dynamics.

Amit Jain is working as a Research Scholar at the Centre for Advanced Studies and Research in Automotive Engineering, Dept. of Mechanical Engineering, Delhi Technological University. He did his B.Tech in Mechanical Engineering from Delhi Technological University, Delhi, India. His research interests include; renewable energy, finite element analysis, design, active aerodynamics, and computation fluid dynamics.

Devang S. Nath is working as a Research Scholar at the Centre for Advanced Studies and Research in Automotive Engineering, Dept. of Mechanical Engineering, Delhi Technological University. He did his B.Tech in Mechanical

Engineering from Delhi Technological University, Delhi, India. His research interests include; internal combustion engines, automobile engineering, aerodynamics, and computation fluid dynamics.

Naveen Kumar was born in August 1967 in Uttar Pradesh, India; Prof. Naveen Kumar did the B.E. from Dayal Bagh Educational Institute, Agra, the M.Tech from IIT Delhi, and Ph.D. from the University of Delhi. He is a Fellow of Institution of Mechanical Engineers (FIMechE), UK; Fellow of Institution of Engineers (FIE), India and Charted Engineer, Engineering Council, UK. His research interests include; alternative fuels with special emphasis on biofuels, decentralized energy systems, renewable energy, waste recycling, and sustainable development. Presently working as Professor of Mechanical Engineering at Delhi College of Engineering since 1996, Prof. Kumar possesses 27 years of experience in academics, industry, and research. He has published more than 100 research papers in the International Journals of repute and more than 100 in Indian Journals and Conferences.

ORCID

Prashant Chandra Pujari http://orcid.org/0000-0001-5632-5865
Amit Jain http://orcid.org/0000-0003-2047-795X
Devang S. Nath http://orcid.org/0000-0003-3887-2157
Naveen Kumar http://orcid.org/0000-0002-4408-9997

References

Akwa, J. V., H. A. Vielmo, and A. P. Petry. 2012. A Review on the Performance of Savonius Wind Turbines. *Renewable and Sustainable Energy Reviews* 16 (5):3054–64. doi:10.1016/j.rser.2012.02.056.

Battisti, L., L. Zanne, S. Dell'Anna, V. Dossena, G. Persico, and B. Paradiso. 2011. Aerodynamic Measurements on a Vertical Axis Wind Turbine in a Large Scale Wind Tunnel. *Journal of Energy Resources Technology, Transactions of the ASME* 133 (3). doi: 10.1115/1.4004360.

Bonalumi, D., and A. Giuffrida. 2016. Investigations of an Air-Blown Integrated Gasification Combined Cycle Fired with High-Sulphur Coal with Post-Combustion Carbon Capture by Aqueous Ammonia. *Energy* 117:439–49. doi:10.1016/j.energy.2016.04.025.

Chen, W. H., C. Y. Chen, C. Y. Huang, and C. J. Hwang. 2017. Power Output Analysis and Optimization of Two Straight-Bladed Vertical-Axis Wind Turbines. *Applied Energy* 185:223–32. doi:10.1016/j.apenergy.2016.10.076.

Edelenbosch, O. Y., K. Kermeli, W. Crijns-Graus, E. Worrell, R. Bibas, B. Fais, S. Fujimori, P. Kyle, F. Sano, and D. P. van Vuuren. 2017. Comparing Projections of Industrial Energy Demand and Greenhouse Gas Emissions in Long-Term Energy Models. *Energy* 122:701–10. doi:10.1016/j.energy.2017.01.017.

El-Askary, W. A., A. S. Saad, A. M. AbdelSalam, and I. M. Sakr. 2020. Experimental and Theoretical Studies for Improving the Performance of a Modified Shape Savonius Wind Turbine. *Journal of Energy Resources Technology, Transactions of the ASME* 142 (12). doi: 10.1115/1.4047326.

Ferrari, G., D. Federici, P. Schito, F. Inzoli, and R. Mereu. 2017. CFD Study of Savonius Wind Turbine: 3D Model Validation and Parametric Analysis. *Renewable Energy* 105:722–34. doi:10.1016/j.renene.2016.12.077.

Goodarzi, M., and R. Keimanesh. 2015. Numerical Analysis on Overall Performance of Savonius Turbines Adjacent to a Natural Draft Cooling Tower. *Energy Conversion and Management* 99:41–49. doi:10.1016/j.enconman.2015.04.027.

Hayashi, T., L. Yan, and Y. Hara. 2005. Wind Tunnel Tests on a Different Phase Three-Stage Savonius Rotor. *JSME International Journal, Series B: Fluids and Thermal Engineering* 48 (1):9–16. doi:10.1299/jsmeb.48.9.

Kumar, A., and R. P. Saini. 2017a. Performance Analysis of a Savonius Hydrokinetic Turbine Having Twisted Blades. *Renewable Energy*. doi:10.1016/j.renene.2017.03.006.

Kumar, A., and R. P. Saini. 2017b. Performance Analysis of a Single Stage Modified Savonius Hydrokinetic Turbine Having Twisted Blades. *Renewable Energy*. doi:10.1016/j.renene.2017.06.020.

Larin, P., M. Paraschivoiu, and C. Aygun. 2016. CFD Based Synergistic Analysis of Wind Turbines for Roof Mounted Integration. *Journal of Wind Engineering and Industrial Aerodynamics* 156:1–13. doi:10.1016/j.jweia.2016.06.007.

Laws, P., J. S. Saini, A. Kumar, and S. Mitra. 2020. Improvement in Savonius Wind Turbines Efficiency by Modification of Blade Designs - A Numerical Study. *Journal of Energy Resources Technology, Transactions of the ASME* 142 (6). doi: 10.1115/1.4045476.

Mahmoud, N. H., A. A. El-Haroun, E. Wahba, and M. H. Nasef. 2012. An Experimental Study on Improvement of Savonius Rotor Performance. *Alexandria Engineering Journal* 51 (1):19–25. doi:10.1016/j.aej.2012.07.003.

Mandal, A. K., K. B. Rana, and B. Tripathi. 2020. Experimental Study on Performance Improvement of a Savonius Turbine by Equipping with a Cylindrical Cowling. *Energy Sources, Part A: Recovery, Utilization and Environmental Effects* 1–19. doi:10.1080/15567036.2020.1772907.

Mari, M., M. Venturini, and A. Beyene. 2017. A Novel Geometry for Vertical Axis Wind Turbines Based on the Savonius Concept. *Journal of Energy Resources Technology, Transactions of the ASME* 139 (6). doi: 10.1115/1.4036964.

Marinić-Kragić, I., D. Vučina, and Z. Milas. 2020. Computational Analysis of Savonius Wind Turbine Modifications Including Novel Scooplet-Based Design Attained via Smart Numerical Optimization. *Journal of Cleaner Production* 262:121310. doi:10.1016/j.jclepro.2020.121310.

Mohamed, M. H., G. Janiga, E. Pap, and D. Thévenin. 2011. Optimal Blade Shape of a Modified Savonius Turbine Using an Obstacle Shielding the Returning Blade. *Energy Conversion and Management* 52 (1):236–42. doi:10.1016/j.enconman.2010.06.070.

Mohan Kumar, P. M., M. R. Surya, S. Narasimalu, and T. C. Lim. 2019. Experimental and Numerical Investigation of Novel Savonius Wind Turbine. *Wind Engineering* 43 (3):247–62. doi:10.1177/0309524X18780392.

Mosbahi, M., A. Ayadi, Y. Chouaibi, Z. Driss, and T. Tucciarelli. 2019. Performance Study of a Helical Savonius Hydrokinetic Turbine with a New Deflector System Design. *Energy Conversion and Management* 194 (February):55–74. doi:10.1016/j.enconman.2019.04.080.

Persico, G., V. Dossena, B. Paradiso, L. Battisti, A. Brighenti, and E. Benini. 2017. Time-Resolved Experimental Characterization of the Wakes Shed by H-Shaped and Troposkien Vertical Axis Wind Turbines. *Journal of Energy Resources Technology, Transactions of the ASME* 139 (3). doi: 10.1115/1.4035907.

Pourhoseinian, M., S. Sharifian, and N. Asasian-Kolur. 2021. Unsteady-State Numerical Analysis of Advanced Savonius Wind Turbine. *Energy Sources, Part A: Recovery, Utilization and Environmental Effects* 00:1–16. doi:10.1080/15567036.2020.1859011.

Rezaeiha, A., H. Montazeri, and B. Blocken. 2019. On the Accuracy of Turbulence Models for CFD Simulations of Vertical Axis Wind Turbines. *Energy* 180:838–57. doi:10.1016/j.energy.2019.05.053.

Rezaeiha, A., I. Kalkman, and B. Blocken. 2017. Effect of Pitch Angle on Power Performance and Aerodynamics of a Vertical Axis Wind Turbine. *Applied Energy* 197:132–50. doi:10.1016/j.apenergy.2017.03.128.

Roshan, A., A. Sagharichi, and M. J. Maghrebi. 2020. Nondimensional Parameters' Effects on Hybrid Darrieus–Savonius Wind Turbine Performance. *Journal of Energy Resources Technology, Transactions of the ASME* 142 (1). doi: 10.1115/1.4044517.

Saad, A. S., I. I. El-Sharkawy, S. Ookawara, and M. Ahmed. 2019, December. Performance Enhancement of Twisted-Bladed Savonius Vertical Axis Wind Turbines. ((2020)) *Energy Conversion and Management* 209:112673. doi: 10.1016/j.enconman.2020.112673.

Saeed, H., A. Hassan, A. M. Nagib Elmekawy, and S. Z. Kassab. 2019. Numerical Study of Improving Savonius Turbine Power Coefficient by Various Blade Shapes. *Alexandria Engineering Journal* 58 (2):429–41. doi:10.1016/j.aej.2019.03.005.

Saha, U. K., S. Thotla, and D. Maity. 2008. Optimum Design Configuration of Savonius Rotor through Wind Tunnel Experiments. *Journal of Wind Engineering and Industrial Aerodynamics* 96 (8–9):1359–75. doi:10.1016/j.jweia.2008.03.005.

Sharma, S., and R. K. Sharma. 2017. CFD Investigation to Quantify the Effect of Layered Multiple Miniature Blades on the Performance of Savonius Rotor. *Energy Conversion and Management* 144:275–85. doi:10.1016/j.enconman.2017.04.059.

Sheldahl, R. E., B. F. Blackwell, and L. V. Feltz. 1978. Wind Tunnel Performance Data for Two- and Three-Bucket Savonius Rotors. *J Energy* 2 (3):160–64. doi:10.2514/3.47966.

Tahani, M., A. Rabbani, A. Kasaeian, M. Mehrpooya, and M. Mirhosseini. 2017. Design and Numerical Investigation of Savonius Wind Turbine with Discharge Flow Directing Capability. *Energy* 130:327–38. doi:10.1016/j.energy.2017.04.125.

Ushiyama, I., H. Nagai, and J. Shinoda. 1986. Experimentally Determining the Optimum Design Configuration for Savonius Rotors. *Transactions of the Japan Society of Mechanical Engineers Series B* 52 (480):2973–82. doi:10.1299/kikaib.52.2973.

Wang, Yabo, Ling Huang, and Raymond Lau 2016. Conversion of Municipal Solid Waste Incineration Bottom Ash to Sorbent Material for Pollutants Removal from Water. *Journal of the Taiwan Institute of Chemical Engineers* 60:275–86. 10.1016/j.jtice.2015.10.013

Wood, D. 2011. *Small Wind Turbines, Green Energy and Technology*. London, UK: Springer. doi:10.1007/978-1-84996-175-2.

Yannopoulos, S. I., G. Lyberatos, N. Theodossiou, L. Wang, M. Valipour, A. Tamburrino, and A. N. Angelakis. 2015. Evolution of Water Lifting Devices (Pumps) over the Centuries Worldwide. *Water (Switzerland)* 7 (9):5031–60. doi:10.3390/w7095031.

Zhao, Z., Y. Zheng, X. Xiaoyun, W. Liu, and H. Guoxiang "Research on the Improvement of the Performance of Savonius Rotor Based on Numerical Study." 1st International Conference on Sustainable Power Generation and Supply, SUPERGEN, Nanjing, China, 2009, 1–6. 10.1109/SUPERGEN.2009.5348197

Energy Systems
https://doi.org/10.1007/s12667-021-00458-4

EDITORIAL

Preface—special issue "Energy Efficiency in Building using Intelligent computing for Smart Cities"

Anand Paul[1] · Anand Nayyar[2] · Akshi Kumar[3] · Jaffar Alzubi[4]

Received: 14 June 2021 / Accepted: 18 June 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Great strides have been made in various areas of building technologies to improve energy efficiency in buildings by applying innovative technologies to reduce energy consumption and enhance indoor environmental quality in buildings. Especially, these efforts are currently being considered together with the application of new energy systems harnessing natural energy and the possibility of taking advantage of today's digital technologies, exploiting big data and online data processing whenever it is possible.

Green building (also known as sustainable building, bioclimatic building, etc.) brings together an interdisciplinary combination of practices and techniques to reduce the impact of buildings on the environment and human health by considering the greenery of the site, soil water content, building energy conservation, recycling of building's rainwater and daily wastewater, sustainable building structure, green materials, etc. Energy efficient building is one of the green building issues and is closely related to other dimensions.

A smart city refers to a city equipped with the basic infrastructure to give a decent quality of life, a clean and sustainable environment to the citizens using smart technology-based solutions. It is a smarter way to deliver governance, providing city services to its residents and develop the infrastructure. Cities in many countries

✉ Anand Nayyar
anandnayyar@duytan.edu.vn

Anand Paul
paul.editor@gmail.com

Akshi Kumar
drkumarakshi@gmail.com

Jaffar Alzubi
j.zubi@bau.edu.jo

[1] Kyungpook National University, Daegu, South Korea

[2] Graduate School, Duy Tan University, Da Nang, Vietnam

[3] Delhi Technological University, New Delhi, India

[4] Al-Balqa Applied University, Amman, Jordan

Published online: 03 July 2021

globally are governed by multiple organizations and authorities. The different spatial entities with multiple boundaries deter effective planning and governance.

The special issue of the Journal titled "Energy Efficiency in Building using Intelligent Computing for Smart Cities" is an excellent collection of review and research articles in the field of Energy efficiency, smart building and smart cities. An open call for paper was issued for this special issue. The guest editors feel happy to announce this special issue of the most reputed journal of Springer.

From a wide range of interesting research papers on various aspects of Smart Buildings, Smart Cities cum related areas, the guest editors, after undergoing exhaustive peer-reviews from experienced and well-known reviewers, have carefully selected seven research papers and one review paper. The final decision for the inclusion of seven research papers and one review papers has been strictly based on the outcome of the rigorous peer-review process, shortlisting successful research papers by researchers as per reviewers' comments and guidelines.

A brief summary of the research papers included in this special issue is enlisted as follows:

The first article by Alawadi et al. [1] titled "A comparison of machine learning algorithms for forecasting indoor temperature in smart buildings" compares 36 Machine Learning algorithms that could be used to forecast indoor temperature in a smart building. More specifically, the researchers conducted experiments using real data to compare their accuracy in terms of R-coefficient and Root Mean Squared Error and their performance in terms of Friedman rank. The results reveal that the ExtraTrees regressor has obtained the highest average accuracy (0.97%) and performance (0,058 %) over all horizons.

The second article by Xijiang [2] titled "Green building design method based on system ecology. Energy Systems, 1–18." proposed a kind of green building design based on system ecology. The design method comprehensively reflect the connotation of the green building design, and construct the index groups into an organic whole at different levels. In addition, the whole process of the design method can be implemented by computer, which simplifies the cumbersome computation process and made the green building design approach more intelligent and model oriented. The researcher compared the proposed method with the traditional methods, and was observed that the new method is more feasible, effective and convenient, which has provided a new idea for the work of the green building design methods in the future.

The third article by Kumar and Jaiswal [3] titled "Scalable intelligent data-driven decision making for cognitive cities" proposes methodology to leverage deeper insights of application of intelligent computation techniques for sentiment analysis of cognitive cities using user-generated data to improve the urban ecosystem. The results are evaluated and analyzed on Twitter datasets that are constructed by the students of the Stanford University. It is observed that the alliance of big data and social media analytics using intelligent sentiment computation has helped in making citizens smart for taking smart decisions and eventually building a smart city.

The fourth article by Gadhavi and Bhavsar [4] titled "Adaptive cloud resource management through workload prediction" proposed a novel cloud framework model titled Auto-Regressive Integrated Moving Average-workload Prediction for

Efficient Resource Provisioning (ARIMA-PERP) to process data in automation with adaptive resource and workload management strategy. To serve the maximum number of user requests, performance metrics of the proposed approach are evaluated. It is observed that the proposed model has achieved an accurate prediction by 91.11%, which meets the efficient resource utilization for the demanded workload. As compared with the exiting approaches, the model achieved better performance by 0.11% for accurate prediction. The proposed architecture is intended to provide the resources dynamically and efficiently satisfying the demands of the user.

The fifth article by Sharma and Tayal [5] titled "Indian smart city ranking model using taxicab distance-based approach" proposes Taxicab Distance-Based Approach (TDBA) to rank the "Indian smart city". Researchers used the "Indian smart city model" define with eight dimensions matching to cities of India. The model takes a smart data-driven decision based on 80 indicators of the city. The "Indian smart cities" are ranking according to the calculated distance of optimal indicators values. The TDBA find the optimal solution on the bases of "Indian smart city" indicators multiple values. The approach calculates the optimal distance solution to find the best result. The result shows the ranking of "Indian smart cities" on the bases of a defined model using TDBA. The ranking gives a status of growth by comparing the cities' rank. The cities can change and update their planning according to cities rank. The result clarifies cities vision and makes a blueprint of the cities it wants to be in the future.

The sixth article by Mahapatra and Nayyar [6] titled "Home energy management system (HEMS): concept, architecture, infrastructure, challenges and energy management schemes" presents a comprehensive review covering the various technical and conceptual aspects of efficient power management at home front. A systematic review proposed by several researchers till date was considered. The study focuses on the concepts, technical background, architecture and infrastructure along with various schemes as well as goals including various issues and challenges faced with HEMS systems. In addition, the paper proposes a novel methodology for improvising the home design architecture by incorporating the concept of green building in order to reduce the energy consumption done by a resident at their home front.

The seventh article by Shrivastava and Kumar [7] titled "Intent and permission modeling for privacy leakage detection in android" explores Android intent and permission as an essential feature of malware detection and analysed their variation on different parameters. Experimental results show a high rate of malware detected by combining intents with permissions as well as sensitivity analysis have shown optimal and efficient results. Researchers also demonstrated the results by experimenting with the combination of Android intents and permissions, to show that these traits do not overlap and our optimal results have shown Android Intent is a plausible trait in malware detection.

The eight article by Alrabea et al. [8] titled "A task-based model for minimizing energy consumption in WSNs" provides a framework for preparing the tasks, in which each node decides to do next according to the observed portion of the request. Within this Framework, the application efficiency and the energy consumption is provided by a weighted reward function and results in better energy/ performance. The suggested approach was analyzed in a target monitoring program. And the

simulation experiments show that cooperative approaches for this type of application are superior to non-cooperative approaches.

1 Conclusion

The main objective of the special issue was to stimulate discussion on the design, use and evaluation of various technologies for smart sustainable cities with smarter applications, including, smart building, smart factories, Intelligent transportation, and smart data-driven decision making. Moreover, dynamic domains of applied artificial intelligence such as machine learning, data mining and semantic web, amongst others, provide novel and sophisticated techniques as scalable intelligent solutions.

We would like to thank the Editor-in-Chief of the journal, Professor Panos M. Pardalos and Jun Peifor their huge support for this issue. Our special thanks go to all editorial staff, for their valuable and prompt support throughout the preparation and publication of this special issue. We express our deep thanks to all authors for their novel contributions to this special issue. We also extend our thanks to all the reviewers for their time, devotion, hard work and on-time precision response to ensure the high-quality review of the accepted papers.

References

1. Alawadi, S., Mera, D., Fernández-Delgado, M., Alkhabbas, F., Olsson, C.M., Davidsson, P.: A comparison of machine learning algorithms for forecasting indoor temperature in smart buildings. Energy Syst. 1–17 (2020)
2. Xijiang, T.: Green building design method based on system ecology. Energy Syst. 1–18 (2020)
3. Kumar, A., Jaiswal, A.: Scalable intelligent data-driven decision making for cognitive cities. Energy Syst. 1–19 (2019)
4. Gadhavi, L.J., Bhavsar, M.D.: Adaptive cloud resource management through workload prediction. Energy Syst. 1–23 (2019)
5. Sharma, K., Tayal, S.: Indian smart city ranking model using taxicab distance-based approach. Energy Syst. 1–18 (2019)
6. Mahapatra, B., Nayyar, A.: Home energy management system (HEMS): concept, architecture, infrastructure, challenges and energy management schemes. Energy Syst. 1–27 (2019)
7. Shrivastava, G., Kumar, P.: Intent and permission modeling for privacy leakage detection in android. Energy Syst. 1–14 (2019)
8. Alrabea, A., Alzubi, O.A., Alzubi, J.A.: A task-based model for minimizing energy consumption in WSNs. Energy Syst. 1–18 (2019)

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

Proximity effects between two plus-plan shaped high-rise buildings on mean and RMS pressure coefficients

Suresh Kumar Nagar[1a], Ritu Raj[*2] and Nirendra Dev[3b]

[1]*Delhi Technological University (DTU), Delhi, India*
[2]*Faculty of Civil Engineering, DTU, Delhi, India*
[3]*Faculty of Civil Engineering, DTU, Delhi, India*

Author 1: Suresh Kumar Nagar, Ph.D Student, Dept. of Civil Engineering, Delhi Technological University, Delhi, Zip Code: 110042, Contact No.: +91-88751 82007.

Author 2: Dr. Ritu Raj (**Corresponding Author)** Assistant Professor, Dept. of Civil Engineering, Delhi Technological University, Delhi, Zip Code: 110042, Contact No.: +91-96500 59954.

Author 3: Prof. Nirendra Dev, Dept. of Civil Engineering, Delhi Technological University, Delhi, Zip Code: 110042, Contact No.: +91-99993 99375.

Abstracts. Wind-induced interference effects may result in a substantial increase or reduction in the wind forces on the buildings constructed in groups due to modifying wind flow around, according to the shape and relative position of the buildings. Experimental study of wind-generated proximity effects on mean and root mean square (RMS) pressure coefficients between two tall buildings are investigated in detail for various interference conditions in this study. The Interference effects on mean and RMS pressure coefficients are presented as interference factors, i.e., mean interference factor (*MIF*) and RMS interference factor (*RIF*). Results show that the interference effects on local wind pressure are significantly higher on the windward side's faces near the recessed corner. The full blockage condition generates suction on walls facing the gap. The half blockage and no blockage conditions create a more severe interference effect than the full blockage. The maximum value of MIF is 4, 9 and 13 in full, half, and no blockage condition, respectively. Interference effects result in reduced wind load on side faces and faces at the leeward side. Suction at side faces reduced approximately by 65% in full blockage condition. RIF's values less than unity are observed for all interference cases.

Keywords: Tall buildings, interference effect, wind load, wind tunnel, mean pressure coefficient, interference factor

1. Introduction

With the rapid urbanization and shortage of land in urban areas, modern high-rise buildings are often constructed in groups. When several high-rise structures are situated in the near vicinity, the wind flow patterns around the building are quite intricate than isolated buildings due to interaction effects. This may have a contrary wind effect in such areas depending on the shape of the building, as the external shape of tall buildings plays an influential role in the generation of wind load on high-rise buildings. This problem is likely to increase significantly in the future due to the increasingly dense arrangement of buildings in cities caused by the shortage of land. Guidelines for estimating the wind loads on tall buildings in current design codes and standards are only available for regular and symmetric shapes. Also, codes and standards offer little guidance regarding the proximity effect for

unconventional plan-shaped buildings. Since wind effects depend on the shape and size of buildings, their relative positions, wind direction, etc., are involved [1, 2], it is challenging to provide the compendious and generalized set of recommendations for modification in wind load due to the presence of adjacent buildings. Therefore, the proximity effect due to nearby structures should be appropriately investigated for the realistic design of buildings for wind [3].

Many researchers have studied proximity effects between tall buildings [4, 5, 6, 7, 8, 9, 10, 11, 12-19]. Behera et al. [20] studied the proximity effect between high-rise buildings with different plan ratios and interfering building positions. It was observed that interference effects were beneficial with respect to the maximum positive pressure, but minimum negative pressure significantly increased. Dongmei *et al.* [21] assessed the Interference factors of global aerodynamics & local wind pressure and lift spectra of a square building in the presence of a lower height building but of the same cross-section at different positions. It was observed that fluctuating aerodynamic forces on the principal building were increased significantly for the oblique upstream position of the interfering building. Kim et al. [22, 23] analyzed the effect of a structural link on the lateral wind response of two square-linked buildings. Results showed that at a small gap between buildings, channeling is strong, and the linked building system acts as a single bluff body. Kim et al. [24] studied the effect of gap distance on wind flow around linked buildings. The results showed that for a parallel arrangement of buildings, the single vortex street and biased flow is observed for a small gap between buildings at 0-degree wind incidence. Lam *et al.* [25] studied the proximity effects on closely spaced square buildings under different wind angles and gap distances between buildings. They concluded that strong channeling through the building gaps exists due to the close proximity. Xie and Gu [13] studied the mean proximity effects between a group of two and three tall buildings. It was concluded that shielding due to upstream building causes a reduction in mean wind forces on the principal building. Yan and Li [26] investigated the interference effects between a pair of aerodynamically modified super-tall buildings. Results showed that the dynamic response has substantially enlarged in a critical arrangement. Yu et al. [27] developed a relationship between interference factor and spacing through high precision regression equations for interference effects between two buildings with different arrangements. Results showed that the interference effect was beneficial for mean pressure, but the peak pressure at the lateral façade near the interfering building was amplified. Zhao and Lam [28] investigated the interference effects between five square tall buildings in the wind tunnel, arranging them in L- and T-shaped patterns. It was observed that a strong interference effect exists on all member buildings, which significantly modifies wind loads as compared to the isolated building. Zu and Lam [29] studied the shielding effect on building of square plan in presence of row of low rise or medium-rise building, and it was concluded that at normal incidence angle, the mean along-wind loads on the principal tall building is always reduced.

Hui *et al.* [30, 31] investigated the mutual interference effects on local peak pressure coefficients between two high-rise building models of different shapes. It was observed that proximity effects depend much on building shapes, configuration and wind angles. The smallest minimum peak pressure at the face increased up to 40%. Kim *et al.* [32] studied the Interference effects on local peak pressures for different wind incidence angles and height ratios of interfering building. The results showed that the peak suction enlarged with higher height ratio. Oblique arrangement generates more severe peak suction compare to tandem configuration. Michel et al. [33] studied aerodynamic interference between three cylindrical buildings present in a row and surrounded by small buildings. A substantial increase in the local pressure was observed through experimental results. Mara *et al.* [10] performed wind tunnel tests to assess the proximity effects among two buildings having similar heights and geometries associated with square plan building of interest. Aerodynamic interference factors greater than unity were observed for RMS cross-wind forces for the direct upstream location of the interfering building.

Flaga et al. [34] studied the effect of interference between closely spaced irregular shaped tall buildings on mean pressure distribution, global forces and pedestrian comfort for wind. It was discovered that the close proximity between buildings generates high negative peak pressure on the surfaces facing the gap. Kar and Dalui [35] investigated the wind interference effects in mean pressure coefficients on an irregular shape building due to a group of three square plan buildings of the equal heights at different locations in along-wind and across-wind directions. Li and LI [36] investigated the interference effect on the irregular shaped twin-tall tapered buildings with recessed corners. It was observed that the proximity effects between tapered buildings were considerably different from that of the square building. Nagar et al. [37] examined the proximity effects between closely spaced square and H plan shape twin tall buildings. It was observed that the interfering building generates suction at windward of the principal building when the interfering building present upstream. Zhang et al. [38] analyzed the interference effects between two linked “H” type

twin-tower structures. It was concluded that the surrounding building's channeling effect is the main source of the maximum cross-bridge displacement.

Chen et al. [39] and Quan et al. [40] studied the proximity effects on an existing target tall building with existence of a proposed super tall building and a pair of adjacent buildings of similar height. The results showed that the aerodynamic response of the target building was considerably increased when the proposed super-tall building existed upstream. Farhadi and Rehnama [41] studied the flow about a surface-attached cube through numerical simulation. Jing et al. [42] analyzed the proximity effects between two oil storage tanks on wind fields and dynamic response due to the wind effects. Results revealed that the pressure acting on the tank under interference is notably higher compared to the single one. Liang et al. [43] examined the interference effects on the wind pressure at the windward facade of the principal building through wind tunnel tests. It was observed that by varying the relative position of the secondary building, the bimodal distribution could be altered to a unimodal form. Sun et al. [44] conducted pressure measurement tests to study the proximity effects between two chimneys with various distances and wind angles. It was concluded that the interference effects were significantly higher on across-wind load compare to those of along-wind load. The maximum interference factor of 1.9 was observed for the extreme across-wind response. Tavakol and Yaghoubi [45] investigated the flow around a surface-mounted hemisphere. The study was done experimentally and numerically, and flow patterns were studied for different flow velocities at various sections. Xing and Qian [16] numerically studied the flow around a group of three circular cylinders arranged in equilateral-triangle arrangements. It was observed that the wake in the back of two parallel cylinders downstream of the three was asymmetrical at small spacing, which disappeared with an increase in spacing ratios.

Although numerous research work has been done to investigate the interference effects between building with square, rectangular or circular sections [1, 8, 9, 12, 16, 18, 31-32, 43, 46-53], the interference effects between unconventional plan tall buildings had been rarely studied [26, 35-39, 54]. Also, studies on proximity effects between closely spaced tall buildings are rare [23-25, 28, 34, 38]. Therefore, these findings need further investigations for a pair of high-rise buildings with unconventional plans. Large-sized recessed corners are generally provided in the twin residential tall buildings to provide amenities and to allow ventilation to washrooms and kitchens and for maximum views to all apartments [55]. The study of interference effects between closely spaced plus-plan shaped tall buildings with large-sized recessed corners has not yet been conducted. This paper centered on the wind-induced proximity effects between two identical plus-plan shaped high-rise buildings on mean and RMS pressure coefficients.

Three different arrangements of the interfering building that provides a full blockage, half blockage, and no blockage conditions for the principal building have been examined. The interference effects on mean and RMS wind pressure coefficients are presented as *MIF* and *RIF* for the three interference conditions. The contours of the interference factors greater than one at critical faces are presented.

2. Experimental Setup

2.1. Approach wind Characteristics

The tests have been performed in an open circuit boundary layer wind tunnel (BLWT) at the Civil Department of the Indian Institute of Technology Roorkee, Roorkee, India. The tunnel had a single fan to generate uninterrupted flow, which was operated by 125 HP motor. The working section of the tunnel was 15m long with cross-section of 2m size, and the total length was 38 m long (Fig. 1). Wind flow characteristics were simulated similar to the conditions of IS 875: Part 3 **(IS 875 Part 3, 2015)** by placing vortex generators, cubical blocks, and barrier wall at the inlet region.

The mean velocity (V) and turbulence intensity(T.I.) profile inside the tunnel are simulated with a power law index (α) = 0.22 for the current study. Models were placed at the center of the turntable, which was located at 12.21m from elliptical effuse, and which can be rotated to set the angle of wind incidence. A dyno drive attached to the diffuser or fan at the outlet of the tunnel was used to vary the wind speed in the tunnel with a maximum speed of 20 m/s. The velocity profile measured at the downstream end of tunnel and the variation of the turbulence intensity of flow is reflected in Fig. 2. The T.I. near the floor of the wind tunnel was found to be about 12 %, and wind velocity at building height was 9.87 m/sec. The wind speed in the wind tunnel was evaluated with the help of "TESTO-480". A probe was connected to this instrument to measure the wind velocity at a different height, which

had a length of 1 m. This instrument was connected to and operated through a computer. Wind pressure on the models was measured using a "Baratron Pressure Transducer", which was capable of measuring extremely low differential heads.

2.2. Details of Model

Experimental building models considered in this study comprised two buildings, out of which one is the rigid pressure model for the principal building under consideration, and the second is the wooden model of interfering building building. The principal building model was made of a transparent Perspex sheet of 4 mm thickness with stiff faces to ensure sufficient rigidity and strength of the model, whereas the interfering building model was a wooden model of the similar shape and size without pressure taping (Fig. 3). The length scale was set as 1:300 to study the model with Plan area 40,000 mm^2 and 600 mm in height in the wind tunnel. Pressure tapings were made of steel tubes with an internal diameter of 1 mm. and 15-20 mm in length. The total number of 196 pressure taping were installed on the wall of the pressure model, which were located at seven different heights levels of 10, 60, 180, 300, 420,540, and 590 mm (Level G-A) from the bottom as shown in Fig. 4(a) to get a proper distribution of wind pressure at all the surfaces. Total 28 measuring points were there at each level (Fig. 4 b). The plan showing different faces and detailed dimensions are shown in Fig.4 (c).

2.3. Pressure measurement

Total four sets of test arrangements were made for measurement of the pressure and interference effect between building models. In set-1, the principal building was tested in an isolated condition with face A as a windward face, as shown in Fig. 4(c). Set-2 to set-4 are shown in Fig. 5, which indicate the different positions of the interfering buildings with respect to the principal building. In set-2, the interfering building was positioned in line with the principal building. Distance (x) between the principal building and the interfering building was kept equal to 60 mm ($1/10^{th}$ of the model height) for all cases. In set-3 and set-4, the interfering building was positioned in oblique arrangement, creating a half blockage and no blockage conditions for the principal building, respectively. The value of the mean pressure coefficient ($C_{p,mean}$) at any pressure measuring point is calculated by normalizing the measuring pressure at the corresponding measuring point based on the following equation:

$$\bar{C}_P = \frac{\bar{P} - P_{static}}{P_{dyn}} \qquad (1)$$

Where $\bar{P}$ is mean pressure; P_{static} is the static pressure at reference height; P_{dyn} is pressure at reference point given by $\frac{1}{2}\rho_a U^2$, where U is the reference velocity at the reference height.

For the fluctuating pressure, root mean square (rms) pressure coefficient ($C_{p,rms}$) has been calculated using the following expression [3];

$$C_{p,rms} = \sqrt{\sum_{K=1}^{N} \left(C_{pk} - C_{p,mean}\right)^2 / (N-1)} \qquad (2)$$

Where C_{pk} is the time history of pressure measuring point; N is the total number of samples, and $C_{p,mean}$ is mean pressure coefficient.

3. Results and Analysis

3.1 Data validation

Before proceeding to study the distribution of pressure on the plus-plan shaped building, it is necessary to verify the calculation method used for wind loading measurements. The experimental parameters of a base square building model with aspect ratio 1:5 at 0-degree wind incidence have been compared with the experimental parameter [56, 57, 58] of classical model of Commonwealth Advisory Aeronautical Research Council (CAARC) standard tall building model tested at several research institutions, including the University of Bristol (BU), Monash University (MU), National Aeronautical Establishment (NAE) and National Physical Laboratory. Fig. 6 presents the comparison of $C_{p,mean}$ at 0.7H height of the square building model with the $C_{p,mean}$ of the CAARC standard tall building model at 2/3H. The mean wind pressure coefficients obtained from this study for the reference square building model are consistent with those of the CAARC model.

3.2 Mean and rms wind pressure coefficients

Prediction of wind load is a difficult task with interfering buildings present in the near vicinity. The relative location of these interfering buildings is an essential parameter that affects the characteristics of wind load on the principal building. A detailed study in isolated as well as in interference condition is employed to investigate the distribution of $C_{p,mean}$ and $C_{p,rms}$ on the surfaces of the principal building. Results are presented as contours of $C_{p,\,mean}$ and $C_{p,rms}$ on all faces.

3.2.1 Isolated Building Condition

The contour of $C_{p,mean}$ and $C_{p,rms}$ on different faces corresponding to the isolated condition are shown in Fig. 7 and Fig. 8, respectively. Herein only half of the surfaces are presented because due to symmetry in plan and wind flow about the axis in the direction of wind flow, similar pressure distribution occurs on symmetric surfaces. The distribution of $C_{p,mean}$ is symmetric about the vertical centerline at the central windward and leeward faces, i.e., at face A2 and C2, respectively. Pressure on the whole face A2 is positive. The maximum $C_{p,mean}$ is 0.58 with 0.34 face average value. The high pressure region is located on the central upper surface, while the lower pressure region is located on the bottom of side edges. Face C2 experiences suction with a small variation of pressure on face ranging from -0.44 to -0.54. High value zone is located at the top, while low value zone is located at central surface at H/3 from bottom. All the surfaces at side and leeward are under suction as expected due to side wash and vortices generated in the wake. Front side faces A1 and A3 have a maximum negative and average value of $C_{p,mean}$ -0.39, and -0.20 respectively, whereas leeward side faces C1 and C3 have an almost similar distribution of pressure throughout the face ranging between -0.47 to -0.58 with average $C_{p,mean}$ -0.54. Side front face B1 and D3 have suction on inside edges, which may be due to the formation of eddies in reentrant corners. Suction tends to reduce on the side faces B2 and D2 from windward edge to leeward edge, with maximum and face average value of $C_{p,mean}$ -0.86, and -0.69, respectively. Distribution of $C_{p,mean}$ varies between -0.48 to -0.62 with face average -0.56 on leeward front face B3 and D1. The RMS pressure coefficients vary from .04 to 0.19 for isolated building condition. The largest value of $C_{p,\,rms}$ is observed at side surface B2. Large values of fluctuating component is observed for side faces and faces which are adjacent to re-entrant corners due to increased turbulence at corners and edges of building in windward side.

3.2.2 Full Blockage Condition

The distribution of $C_{p,mean}$ corresponding to full blockage interference condition, is shown in Fig. 9. The distribution of $C_{p,mean}$ on all surfaces of the principal building in the presence of the interfering building, creating full blockage condition for the principal building is quite dissimilar from that of the isolated building. Pressure tends to increase with respect to height, like isolated building on the front face and reduces near the top edge. Strong interference effect is created on front faces because it immersed in the wake of interfering building, which causes suction on front face A2, unlike for isolated building and distribution of suction coefficient varies between -0.32 to -0.63. The high suction zone is located at a height between .75H to 0.85H. The absolute value is approximately 29% higher than those for isolated building. The mean suction coefficient for the wind-ward face (B1 and D3) at the side is significantly increased compared to an isolated building. The suction coefficient increased for windward side faces (A1 and A3) compared to the isolated building by 51% correspond to the average of the face; however suction decreases for leeward side faces (C1 and C3) with a small variation in suction coefficient between -0.21 to -0.24 on the surface. The considerable reduction (Approx. 65%) is noticed in suction on side faces (B2 and D2), and all leeward faces (B3, C1, C2, C3, and D1), and the percentage reduction is in the range of 55- 65 %. Suction coefficients on these side faces tend to decrease from windward to leeward edge, and maximum suction is at the middle height of the windward edge with $C_{p,mean}$. Distribution on windward face A2 and leeward face C2 is symmetric about the vertical centerline.

As shown in Fig. 10, RMS pressure coefficient on all faces of the principal building at full blockage condition is significantly decrease compared to isolated building and maximum decreased is for side faces from 0.13 to 0.04. Overall maximum face $C_{p,rms}$ is decreased compared to isolated building and is at faces attached with reentrant corners on the windward side.

3.2.3 Half Blockage Condition

Fig. 11 shows the distribution of Cp,mean at all faces of the principal building at half blockage interference condition. Interference building creates unsymmetrical wind flow around the principal building due to half blockage and thus creates different pressure

distribution on the symmetric faces. The interference effect on all faces depends on the face position in the wake area and orientation concerning the flow of wind. Active shielding is created by interfering building. Unlike isolated and full blockage interference conditions, the distribution of pressure is not symmetrical about the vertical centerline on front face A2. Both sides are under the opposite nature of pressure. The surfaces towards the interfering building, which come under the wake region of the interfering building, are under suction, whereas the opposite side has positive pressure. High pressure zone is located near the right edge from bottom to top. Suction on the side face A1 at the center on the windward side increased, and the average $C_{p,mean}$ of the face is -0.47, which is 57% higher than isolated building. Front side face A3 has positive pressure, which was in suction for isolated as well as full blockage conditions. Wind ward face B1 at the side also experienced positive pressure, whereas face D3 is under increased suction compare to the isolated building and full blockage case. The effect of interference on all other faces is small, and the distribution of $C_{p,mean}$ is likely similar to that of isolated building case.

The fluctuating pressure component, as shown in Fig. 12, again decreased significantly at half blockage interference position compared to the isolated building for all faces. But maximum of $C_{p,rms}$ for face A1 is increased by 58% compared to isolated building and doubled with respect to full blockage building condition. For all other faces, the maximum of $C_{p,rms}$ is decreased compared to an isolated building.

3.2.4 No Blockage Condition

Contour plot of $C_{p,mean}$ and $C_{p,rms}$ at all faces of principal building for the no blockage interference condition are shown in Fig. 13 and 14, respectively. Interference effects for this location of interfering building are critical for faces in the left side of the centerline along the windward direction due to the fact that faces on this side are also affected by the wake generated due to the existence of the nearby building. Suction at the front side face A1 significantly increased and was maximum among isolated and all three interference conditions. The distribution of $C_{p,mean}$ varies from -0.68 to -0.86 with an average at the face is -0.76, which is 60% higher than half blockage condition. Pressure distribution on front face A2 is similar to that for half blockage interference condition. Maximum and face average values of $C_{p,mean}$ for front face D3 at the side toward the interfering building are also increased crucially as compared to isolated, and all other interference conditions and are -0.89 and -0.65, respectively. The face average value for face D3 is increased by 45% from half blockage condition. The interference effect on all other faces is favorable i.e. reduced.

RMS pressure coefficient $C_{p,rms}$ is either similar for most of the surfaces or reduced to that of isolated building case. Maximum value increased by 40% to that of an isolated building for face A1 and for all other faces, it has reduced.

A detailed study of mean and rms pressure coefficients is employed along the surface at each level of the principal building in isolated as well as with an interfering building at all three interference conditions. Fig. 15 renders the distribution of $C_{p,\ mean}$ along measuring points at different levels for isolated building and three different blockage conditions of interference. The fluctuation in $C_{p,mean}$ is significantly large for measuring points which lie on faces at the front(i.e., at points 1 to 9, 27 and 28). The effect of re-entrant corners and edges can be seen from the figure clearly. Fluctuation in $C_{p,mean}$ at measuring points on faces at side and back, is significantly small and absolute $C_{p,\ mean}$ is maximum for isolated building. It seems that the largest absolute value of $C_{p,mean}$ occurs for no blockage condition at the bottom most level at 10 mm from the bottom.

The distribution of rms pressure coefficients is shown in Fig. 16 at different levels of the principal building. $C_{p,rms}$ is increased toward the central height of the building and significant at middle levels compare to the top and bottom levels except face A1. $C_{p,rms}$ is reduced for leeward building side i.e. between measuring points 11 and 25 at each level for all conditions. For half blockage, the fluctuating component of pressure coefficient is significant at the windward edge of face A1 at level A; however, at level B, it is significant for no blockage condition near the corner on face A1. At the center height of building $C_{p,rms}$ is greater and variation along the surface is also significant for isolated building. For bottom level G, distribution is similar to that of top level A except for face A1 for half blockage.

3.3 Interference Factor

The interference effects corresponding to different interference conditions on pressure coefficients (C_p) related to all measurement points are of composite nature and challenging to present for all. To clarify the intricacy and to scrutinize the interference effects on C_p in detail, interference factors (I.F.) for an average of $C_{p,mean}$ and $C_{p,rms}$ at each face are proposed as given by ***Khanduri et al.*****[1]**, to indicate the severity of interference effects on $C_{p,mean}$ and $C_{p,rms}$ as follows:

$$(MIF) = \frac{C_{p,mean}\ with\ interfering\ building}{C_{p,mean}\ isolated\ building}$$

$$(RIF) = \frac{C_{p,rms}\ with\ interfering\ building}{C_{p,rms}\ isolated\ building}$$

Where MIF and RIF are the interference factors for mean and RMS pressure coefficients, respectively. Fig. 17 renders the variation of MIF and RIF along the different faces of the principal building starting from the front side face A1 to side front face D3 anticlockwise for all three interference conditions. From Fig. 17(a), interference effects are significant for face A1, B1, and D3 only. Interference effects for full blockage interference condition are critical at Face B1 and D3 only, for which MIF greater than 4. For half blockage condition, MIF is greater than 2, whereas for face B1, absolute MIF is greater than 7 with a negative sign, the significance of which is that the nature of pressure is changed due to interference. MIF for face D3 is greater than 9, which is maximum among all faces and signifies that effects of interference for full blockage condition are maximum for face D3, which is a side front face toward the interfering building and falls in the wake zone of interfering building.

Interference effects for full blockage condition are more severe for D3 among three critical faces. MIF for face D3 is greater than 13, which shows a large interference-effect. Interference effect on this face is most severe among all faces and all interference conditions, which can be explained by the fact that this face immerged in the wake region of interfering building corresponding to no blockage condition and the velocity of flow increased after separating from the upwind interfering building due to which suction is increased by a significant amount and results in a high value of interference factor. Form Fig. 17(b), the maximum value of RIF for surface average $C_{p,rms}$ is 1.03, which shows that the fluctuating component of pressure coefficient is not much affected due to the existence of an interfering building at three positions. Average of surface $C_{p,rms}$ for all surfaces are reduced due to the presence of the interfering building, but peak values at the top level for some surfaces have increased, which may be due to vortex shedding for half blockage and no blockage conditions. From Fig. 17(a) it is clear that interference effects are significant for three critical faces, namely A1, B1 and D3 for which I.Fs are very high, hence the distribution of I.Fs herein only deals with the results for these three faces of the principal building. Fig. 18 render the contour of *MIFs* on these faces for three interference conditions, created due to different relative positions of the interfering building. In order to highlight the unfavorable positions of measuring points on the face, only those absolute I.F ≥ 1 are retained. For full blockage interference condition, on face A1 I.Fs distributed evenly on the surface from 0.51 to 2.31. Distribution on face B1 and D3 are similar due to symmetry in position and wind flow around. Outer edges are significantly affected by interference, and from Fig. 18 (a) it is quite clear that positions above the middle of the inner edge are favorable, and regions of unfavorable locations concentrate on outer edges at 1/3 height from the bottom. MIF for half blockage condition on face A1 distributed on bottom 2/3 height with small variation, whereas variations at top 1/3 height are significant, and MIF varies from 1.67 to 4.27. For face B1 and D3 distribution is similar, but values are slightly different. MIF tends to increase near outer edges. The positions of favorable and unfavorable regions are similar as in the case of full blockage condition. For no blockage interference condition, the distribution of MIFs is quite different from previous conditions. A significantly large interference effect is noticed at face A1and D3. MIF on face A1 is distributed evenly throughout the face, and varies from 2.0 to 5.33, whereas at face D3 MIF varies between a large range from 0.39 to 11.83, and regions of unfavorable positions concentrate on the bottom half. Unfavorable measuring points at face B1 concentrates on the outer edge between a quarter to the half-height.

The distribution of RIFs is shown in Fig. 19 for critical faces and unfavorable positions are highlighted by retaining IF ≥ 1 only at surfaces. From the figure, it is evident that for full blockage regions of unfavorable are concentrate at the corner part of the building surfaces. The maximum value is 1.13. For half blockage distribution at central surface A2 is quite different than other because the boundary of the flow field at the back of interfering building lies at the center of this face. For no blockage interfering condition, windward faces A1 and A2 and leeward face D1 are the most critical surfaces for which larger area is covered by unfavorable positions.

4. Conclusions

Wind-induced proximity effects on $C_{p,mean}$ and $C_{p,rms}$ between two identical plus-plan models have been investigated in this study, using detailed wind tunnel experiments. Three different interference conditions as full blockage, half blockage, and no blockage were considered, which were created by arranging two buildings in tandem and oblique configurations. The distribution

of $C_{p,mean}$ at the surfaces of the principal building are presented as contour plots for isolated building and three interference conditions. Interference effects are presented by interference factors for C_p at the face as *MIF*, and *RIF* and contour of *MIF* and *RIF* were plotted for critical faces, where interference effects were significantly large. The following conclusion can be drawn from this study:

- Pressure distribution is symmetric for isolated building condition and full blockage interference condition but does not remain symmetrical for half blockage and no blockage interference conditions.
- For isolated building condition, side faces B2 and D2 have maximum pressure (suction), and the absolute value of average $C_{p,mean}$ at the face is 0.69. The maximum face average value of $C_{p,mean}$ equal for full blockage condition, is observed at front face A2. For half blockage and no blockage, maximum values are at face B2 and A1, respectively.
- Interference effects are significant at faces on the windward side only at which interference factors are greater than unity. Interference factors at other faces are less than unity, which shows that pressure coefficients are reduced compare to isolated building condition.
- The maximum mean interference factor (MIF) for full blockage condition is greater than 4 and observed at side faces B2 and D2. For half blockage condition, MIFs are significantly large at face B1 and D3. Values are higher than 7 and 9, respectively. Whereas for no blockage interference condition, the largest value of MIF greater than 13 is observed at face D3, which is the largest among all interference conditions. RIF is more significant for no blockage condition at face A1.
- Fluctuation in values of $C_{p,mean}$ at any level is large for measuring points at the windward side, which immersed in the wake region of upwind interfering building.
- Interference effects for individual measuring points on the face are more critical for face A1, B1, and D3, at which MIFs are greater than 2. For face A1 MIF are distributed evenly, whereas for face B1 and D3 regions of unfavorable locations concentrate on outer edges between ¼ to ½ height from bottom except for no blockage condition. Unfavorable regions corresponding to RIF have concentrated at corners parts of the building faces, except central windward face A2.
- The interference effect between closely spaced plus-plan twin tall buildings still needs more in-depth study and investigations. The study of the effect of wind direction is also required in the future, which is not presented here because of the large content of the paper.

Acknowledgments

The authors gratefully acknowledge the Indian Institute of Technology, Roorkee (IITR), India, for providing the required infrastructure and resources to perform this research study.

References

1. Khanduri, A.C., Stathopoulos, T. and Bédard, C. “Wind-induced interference effects on buildings – a review of the state-of-the-art”, Eng Struct, 20, pp. 617-630 (1998).
2. Bhattacharyya, B. and Dalui S.K. “Experimental and numerical study of wind-pressure distribution on irregular-plan-shaped building.” Journal of Structural Engineering 146(7) pp. 04020137 (2020).
3. Li, Yi, Li, Q.S. and Ju., K. L. “Experimental investigation of the wind pressure distribution and wind interference effects on a typical tall building.” Advanced Materials Research. 639 pp. 444-451 Trans Tech Publications Ltd, (2013).
4. Amin, J.A. and Ahuja, A.K. “Mean interference effects between two buildings: effects of close proximity”, Struct Design Tall Spec Build, 20, pp. 832–852 (2011).
5. Amin, J.A. and Ahuja, A.K. “Wind-induced mean interference effects between two closed spaced buildings”, KSCE Journal of Civil Engineering, 16, pp. 119-131 (2012).
6. Baily, P.A. and Kwok, K.C.S. “Interference excitation of twin tall buildings”, J Wind Eng Ind aerodyne, 21, pp. 323-338 (1985).
7. Cho, K., Hong, S. and Hwang, K.S. “Effects of neighboring building on wind loads”, In: CTBUH 2004 Seoul Conference, pp. 516-523 (2004).
8. Khanduri, A.C., Stathopoulos, T. and Bédard, C. “Generalization of wind-induced interference effects for two buildings”, Wind Struct, 3, pp. 255-266 (2000).
9. Kim, W., Tamura, Y. and Yoshida, A. “Interference effects on aerodynamic wind forces between two buildings”, J Wind Eng Ind aerodyne, 147, 186–201 (2015).

10. Mara, T.G., Terry, B.K., Ho, T.C.E., et al. “Aerodynamic and peak response interference factors for an upstream square building of identical height”, J Wind Eng Ind aerodyne, 133, pp. 200-210 (2014).
11. Tse, K.T., Wang, D.Y. and Zhou, Y. “Wind pressure characteristics for a double tower high-rise structure in a group of buildings”, Wind Struct, 5, pp. 491-515 (2013).
12. Wu, G., Du, X. and Wang, Y. “LES of flow around two staggered circular cylinders at a high subcritical Reynolds number of 1.4× 105”, J Wind Eng Ind aerodyne, 196, pp. 104044 (2020).
13. Xie, Z.N. and Gu, M. “Mean interference effects among tall buildings”, Eng Struct, 26, pp. 1173-1183 (2004).
14. Xie, Z.N. and Gu, M. “A correlation-based analysis on wind-induced interference effects between two tall buildings”, Wind Struct, 3, pp. 163-178 (2005).
15. Xie, Z.N. and Gu, M. “Simplified formulas for evaluation of wind-induced interference effects among three tall buildings”, J Wind Eng Ind aerodyne, 95, pp. 31-52 (2007).
16. Xing, Q., and Qian, J. “CFD Analysis of Wind Interference Effects of Three High-rise Buildings”, Journal of Asian Architecture and Building Engineering, 17, pp. 487-494 (2018).
17. Yahyai, M., Kumar, K., Krishna, P., et al. “Aerodynamic interference in tall rectangular buildings”, J Wind Eng Ind aerodyne, 41(1-3), pp. 859-866 (1992).
18. Zu, G.B. and Lam, K.M. “Across-wind excitation mechanism for interference of twin tall buildings in tandem arrangement”, Wind Struct, 6, pp. 397-413 (2018a).
19. Kim, W., Tamura, Y., and Yoshida, A. “Simultaneous measurement of wind pressures and flow patterns for buildings with interference effect.” Advances in Structural Engineering, 16(2), pp. 287-305 (2013).
20. Behera, S., Ghosh, D., Mittal, A.K., et al. “The effect of plan ratios on wind interference of two tall buildings”, Struct Design Tall Spec Build, 29(1), pp. e1680 (2020).
21. Dongmei, H., Xue, Z., Shiqing, H., et al., “Characteristics of the aerodynamic interference between two high-rise buildings of different height and identical square cross-section”, Wind Struct, 24, pp. 501-528 (2017).
22. Kim, B., Tse, K. T., and Tamura, Y. “POD analysis for aerodynamic characteristics of tall linked buildings”, J Wind Eng Ind aerodyne, 181, pp. 126-140 (2018).
23. Kim, B., and Tse, K.T. “POD analysis of aerodynamic correlations and wind-induced responses of two tall linked buildings”, Engineering Structures, 176, pp. 369-384 (2018).
24. Kim, B., Tse, K.T., Yoshida, A., et al. “Investigation of flow visualization around linked tall buildings with circular sections”, Building and Environment, 153, pp. 60-76 (2019).
25. Lam, K.M., Leung, M.Y.H. and Zhao, J.G. “Interference effects on wind loading of a row of closely spaced tall buildings”, J Wind Eng Ind aerodyne, 96, pp. 562-583 (2008).
26. Yan, B. and Li, Q.S. “Wind tunnel study of interference effects between twin super-tall buildings with aerodynamic modifications”, J Wind Eng Ind aerodyne, 156, pp. 129-145 (2016).
27. Yu, X.F., Xie, Z.N., Zhu, J.B., et al. “Interference effects on wind pressure distribution between two high-rise buildings”, J Wind Eng Ind aerodyne, 142, pp. 188-197 (2015).
28. Zhao, J.G. and Lam, K.M. “Interference effects in a group of tall buildings closely arranged in an L- or T-shaped pattern”, Wind Struct, 11, pp. 1-18 (2008).
29. Zu, G.B and Lam, K.M. “Shielding effects on a tall building from a row of and medium rise buildings”, Wind Struct, 6, pp. 439-449 (2018b).
30. Hui, Y., Tamura, Y. and Yoshida, A. “Mutual interference effects between two high-rise building models with different shapes on local peak pressure coefficients”, J Wind Eng Ind aerodyne, 104–106, pp. 98–108 (2012).
31. Hui, Y., Yoshida, A. and Tamura, Y., “Interference effects between two rectangular-section high-rise buildings on local peak pressure coefficients”, Journal of fluids and structures, 37, pp. 120-133 (2013).
32. Kim, W., Tamura, Y. and Yoshida, A. “Interference effects on local peak pressures between two buildings”, J Wind Eng Ind aerodyne, 99, pp. 584–600 (2011).
33. Macháček, M., Urushadze, S., Pospíšil, S., et al. “Aerodynamic interference of wind flow around three cylindrical bodies with surface roughness”, In MATEC Web of Conferences, 313, pp. 00051. EDP Sciences, 2020.
34. Flaga, A., Kocoń, A., Kłaput, R., et al. “The environmental effects of aerodynamic interference between two closely positioned irregular high buildings”, J Wind Eng Ind aerodyne, 180, pp. 276-287 (2018).

35. Kar, R., Dalui, S.K., "Wind interference effect on an octagonal plan shaped tall building due to square plan shaped tall buildings", Int J Adv Struct Eng, 8, pp. 73–86 (2016).
36. Li, X., and Li, Q.S. "Wind- induced interference effects between twin tapered skyscrapers", Struct Design Tall Spec Build, 28(6), pp. e1594 (2019).
37. Nagar, S.K., Raj, R. and Dev, N. "Experimental study of wind-induced pressures on tall buildings of different shapes", Wind and Structures, 31(5), pp. 431-443 (2020).
38. Zhang, L., Cheng, W. and Xie, Z. "Wind effect of a twin- tower super high- rise building with weak connection", Struct Design Tall Spec Build, 27(15), pp. e1503 (2018).
39. Chen, J., Quan, Y., and Gu, M. "Aerodynamic interference effects of a proposed super high-rise building on the aerodynamic forces and responses of an existing building", J Wind Eng Ind aerodyne, 206, pp. 104312 (2020).
40. Quan, Y., Chen, J. and Gu, M. "Aerodynamic interference effects of a proposed taller high- rise building on wind pressures on existing tall buildings", Struct Design Tall Spec Build, 29(4), pp. e1703 (2020).
41. Farhadi, M. and Rahnama, M. "Large eddy simulation of separated flow over a wall-mounted cube", Scientia Iranica, 13(2), pp. 124-133 (2006).
42. Jing, W., Wang, J. and Cheng, X. "Dynamic responses of oil storage tank considering wind interference effect", Engineering Failure Analysis, 104, pp. 1053-1063 (2019).
43. Liang, Q. S., Fu, J. Y., Li, Z., et al. "Bimodal distribution of wind pressure on windward facades of high-rise buildings induced by interference effects", J Wind Eng Ind aerodyne, 200, pp. 104156 (2020).
44. Sun, Y., Li, Z., Sun, X., et al. "Interference effects between two tall chimneys on wind loads and dynamic responses", J Wind Eng Ind aerodyne, 206, pp. 104227 (2020).
45. Tavakol, M.M., and Yaghoubi, M. "Experimental and Numerical Analysis of Turbulent Air Flow Around a Surface Mounted Hemisphere", Scientia Iranica, 17(6), pp. 480-491 (2010).
46. Gu, M. and Xie, Z.N. "Interference effects of two and three super-tall buildings under wind action", Acta Mech Sin, 27, pp. 687-696 (2011).
47. Lim, J. and Bienkiewicz, B. "Wind tunnel investigation of correlation and coherence of wind loading on generic tall twin buildings in close proximity", Wind Struct, 4, pp. 443-456 (2014).
48. Sy, L.D., Yamada, H. and Katsuchi, H. "Interference effects of wind-over-top flow on high-rise buildings", J Wind Eng Ind aerodyne, 187, pp. 85-96 (2019).
49. Taniike, Y. "Interference mechanism for enhanced wind forces on neighboring tall buildings", J Wind Eng Ind aerodyne, 42, pp. 1073-1083 (1992).
50. Yu, X.F., Xie, Z.N., Wang, X., et al. "Interference effects between two high-rise buildings on wind-induced torsion", J Wind Eng Ind aerodyne, 159, pp. 123-133 (2016).
51. Yu, X., Xie, Z. and Gu, M. "Interference effects between two tall buildings with different section sizes on wind-induced acceleration", J Wind Eng Ind aerodyne, 182, pp. 16-26 (2018).
52. Zu, G.B., and Lam K.M. "Across-wind excitation mechanism for interference of twin tall buildings in staggered arrangement." J Wind Eng Ind aerodyne, 177, pp. 167-185 (2018).
53. Germi, M.S. and Kalehsar, H.E. "Numerical investigation of interference effects on the critical wind velocity of tall buildings." Structures, 30, pp. 239-252, Elsevier, 2021.
54. Blessmann, J. "Neighbouring wind effects on two tall buildings", J Wind Eng Ind aerodyne, 41-44, pp. 1041-1052 (1992).
55. Wong, S.Y., and Lam, K.M. "Effect of recessed cavities on wind-induced loading and dynamic responses of a tall building." J Wind Eng Ind aerodyne, 114, pp. 72-82 (2013).
56. Meng, F.Q., He, B.J., Zhu, J., et al. "Sensitivity analysis of wind pressure coefficients on CAARC standard tall buildings in CFD simulations", Journal of Building Engineering, 16, pp.146-158 (2018).
57. Zhao, D.X. and He, B.J. "Effects of architectural shapes on surface wind pressure distribution: case studies of oval-shaped tall buildings." Journal of Building Engineering, 12, pp. 219-228 (2017).
58. Li, S.L., Liu, L.L., Wu, H., et al. "New test method of wind pressure coefficient based on CAARC standard model determined using vehicle driving wind." Experimental Techniques 43, no. 6, pp. 707-717 (2019). https://doi.org/10.1007/s40799-019-00330-2

List of Figures

Fig. **1** Wind tunnel layout
Fig. 2 Mean Velocity and turbulence intensity(TI) profile
Fig. 3 Rigid Pressure Models in the wind tunnel
Fig. 4 (a) Pressure tapping levels (unit: mm); (b) pressure tap distribution at each layer; (c) plan dimension and building faces (unit: mm)
Fig. 5 interference conditions (a) full blockage (b) half blockage (c) no blockage
Fig. 6 Comparison of $C_{p,\,mean}$ of reference square model with CAARC standard models
Fig. 7 Distribution of mean pressure coefficients (Isolated building)
Fig. 8 Distribution of rms pressure coefficients (Isolated building)
Fig. 9 Distribution of mean pressure coefficients (full blockage)
Fig. 10 Distribution of rms pressure coefficients (full blockage)
Fig. 11 Distribution of $C_{p,mean}$ on windward faces (half blockage)
Fig. 12 Distribution of $C_{p,rms}$ on windward faces (half blockage)
Fig. 13 Distribution of $C_{p,mean}$ on windward faces (no blockage)
Fig. 14 Distribution of $C_{p,rms}$ on windward faces (no blockage)
Fig.15 Mean pressure coefficients at different levels of principal buildings
Fig.16 RMS pressure coefficients at different levels of principal buildings
Fig. 17 Interference Factor on Face (a) MIF; (b) RIF
Fig. 18 MIF (I.F ≥ 1) contour (Left to right face A1, B1 and D3) (a) Full (b) Half (c) No blockage
Fig. 19 RIF (IF $\geq$ 1) contour (a) Full blockage (b) Half blockage (c) no blockage

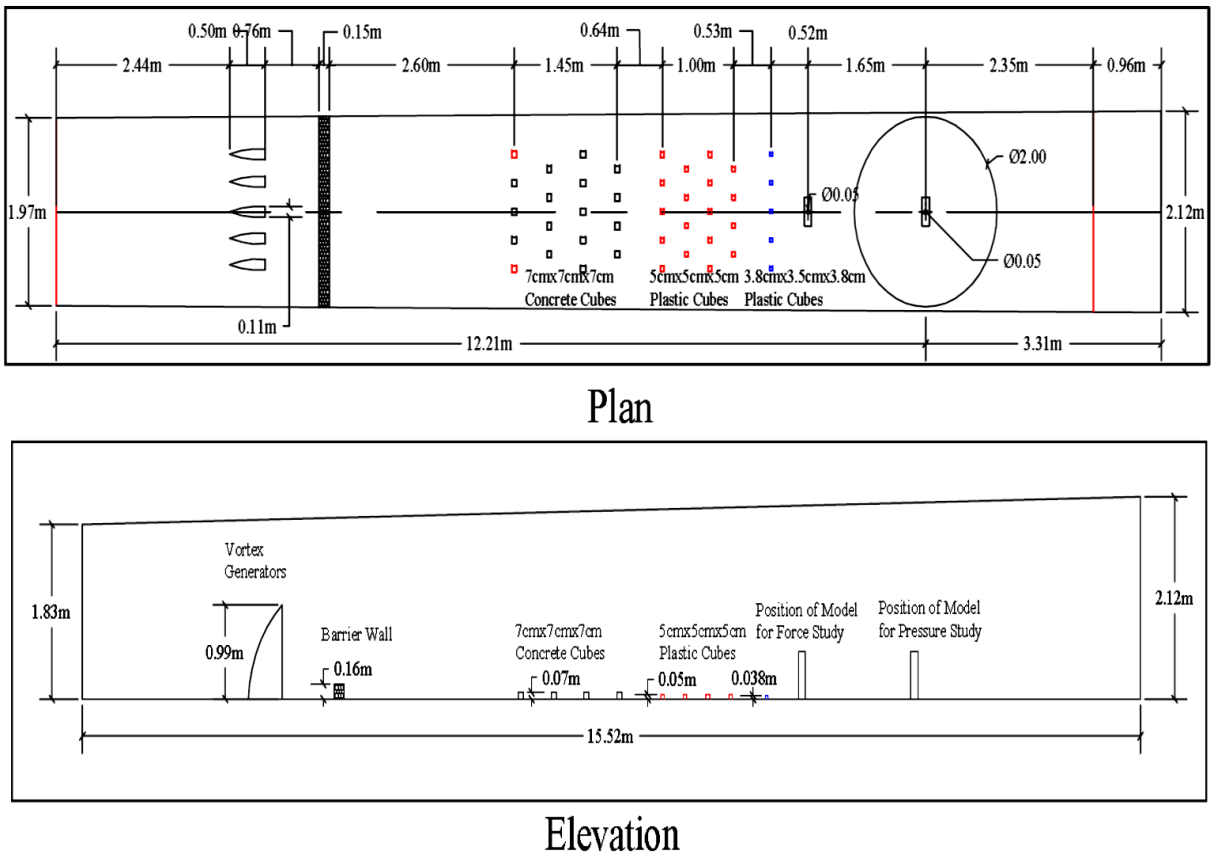


Fig. 1 Wind-Tunnel layout

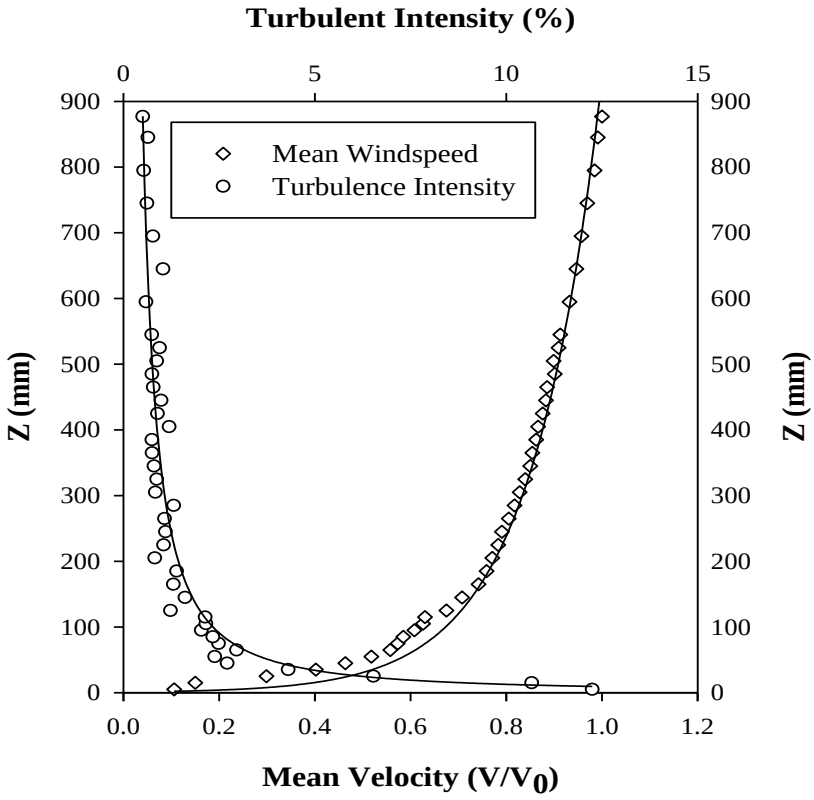
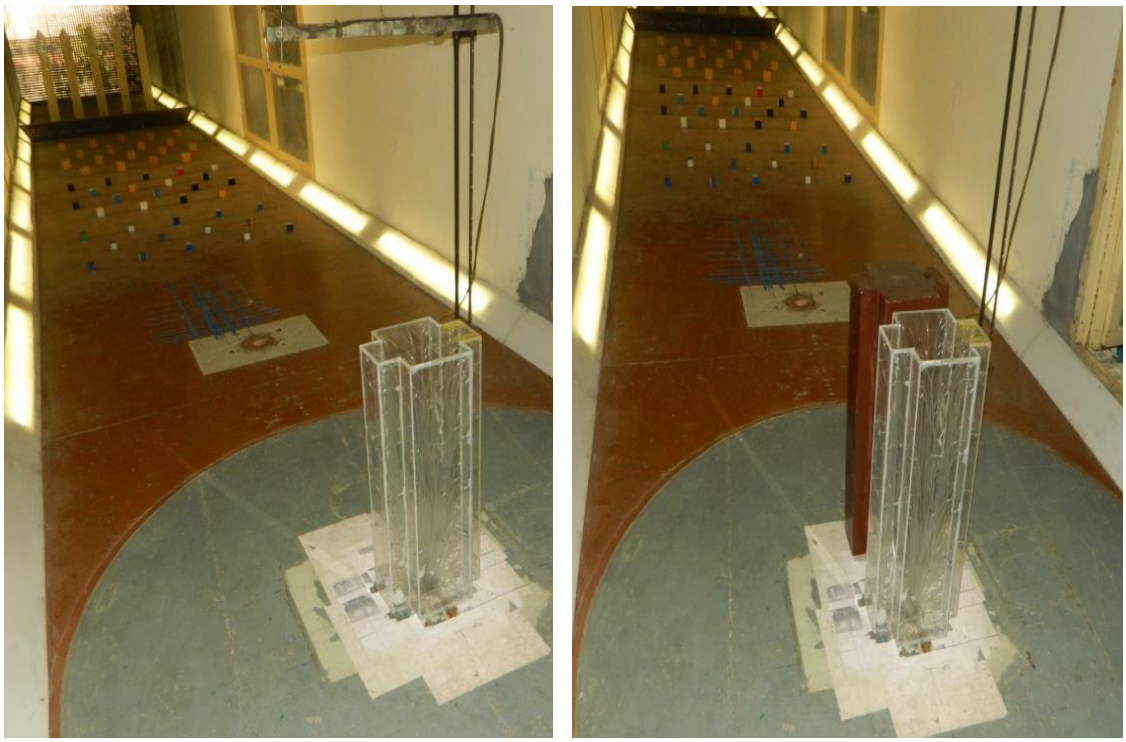


Fig. 2 Mean Velocity and turbulence intensity(TI) profile



(a) Isolate (b) Interference

Fig. 3 Rigid Pressure Models in the wind tunnel

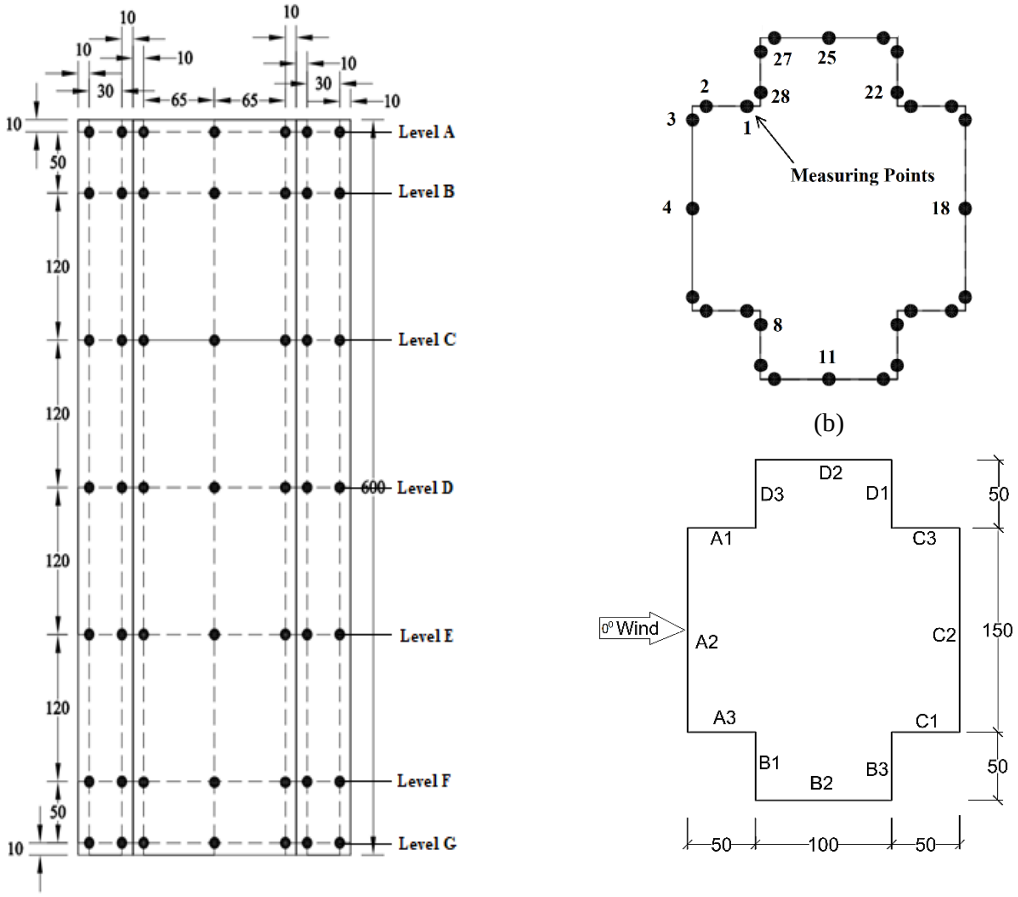
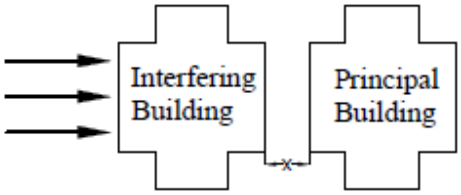
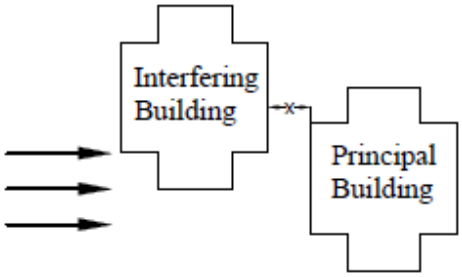


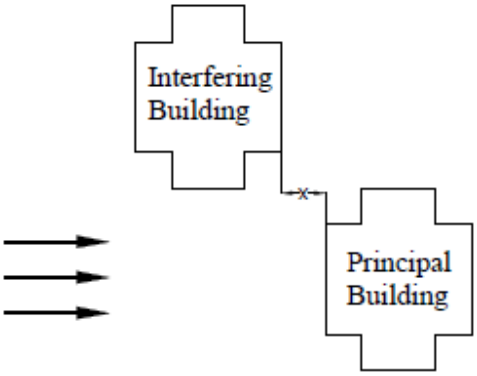
Fig. 4 (a) Pressure tapping levels (unit: mm); (b) pressure tap distribution at each layer; (c) plan dimension and building faces (unit: mm)



(a)



(b)



(c)

Fig. 5 interference conditions (a) full blockage (b) half blockage (c) no blockage

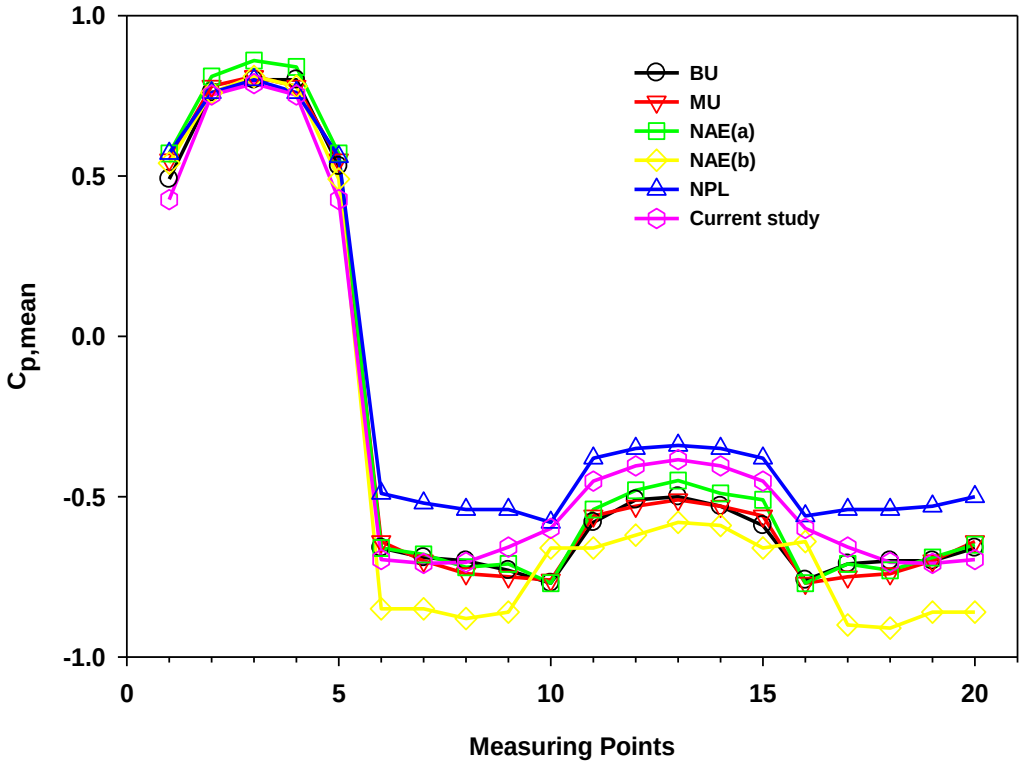


Fig. 6 Comparison of $C_{p,\,mean}$ of reference square model with CAARC standard models

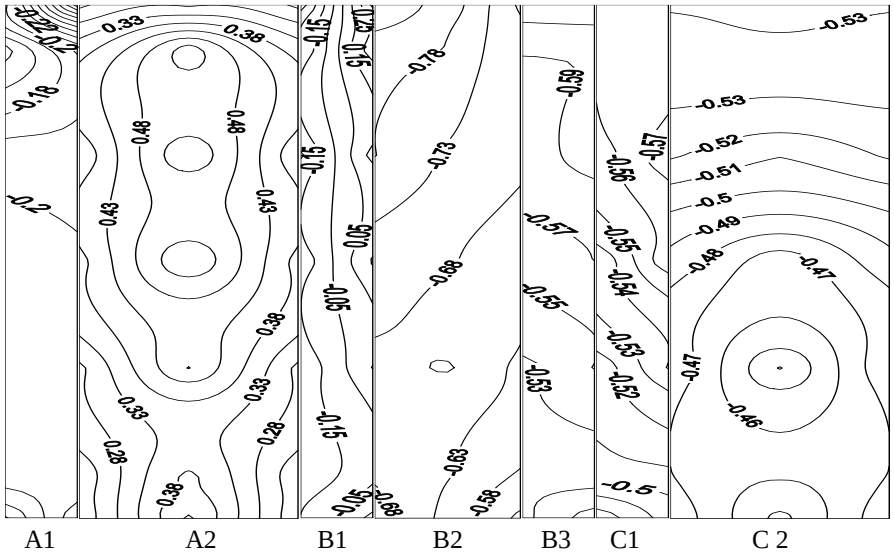


Fig. 7 Distribution of mean pressure coefficients (Isolated building)

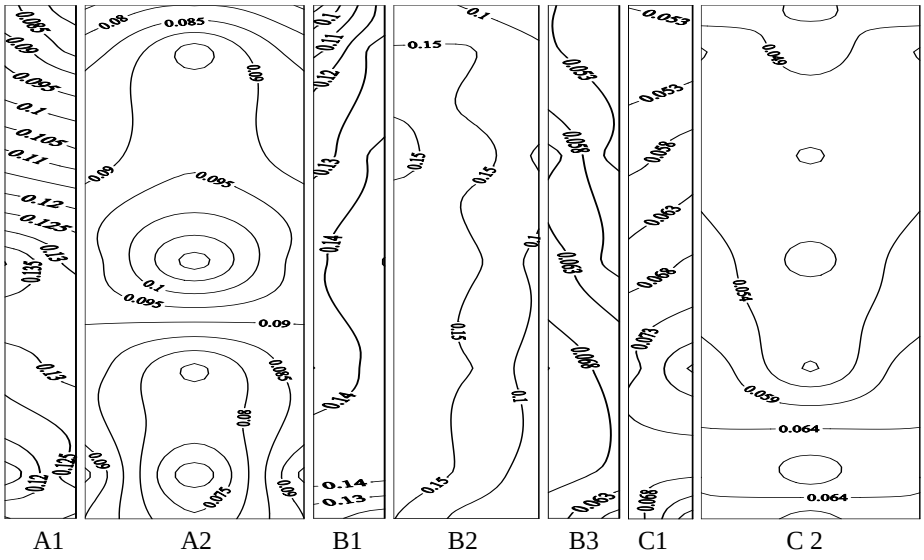


Fig. 8 Distribution of rms pressure coefficients (Isolated building)

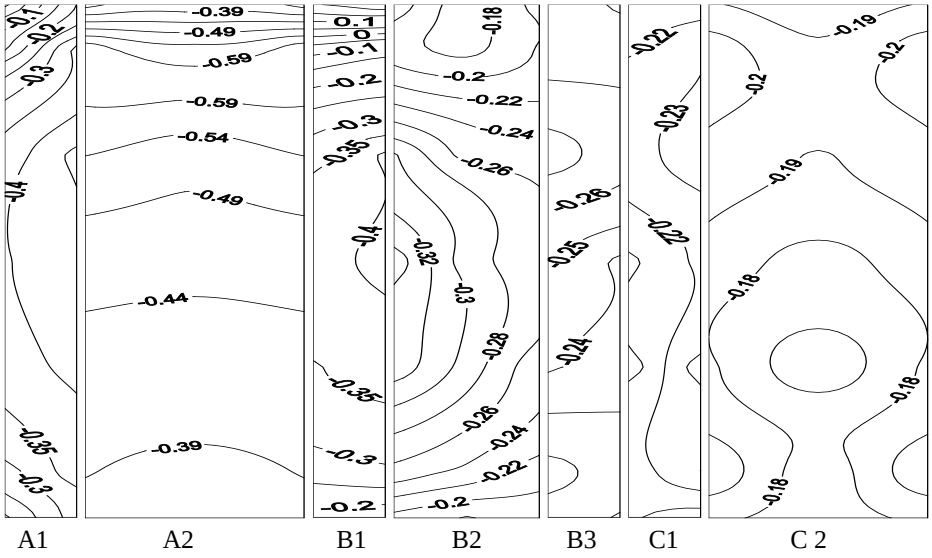


Fig. 9 Distribution of mean pressure coefficients (full blockage)

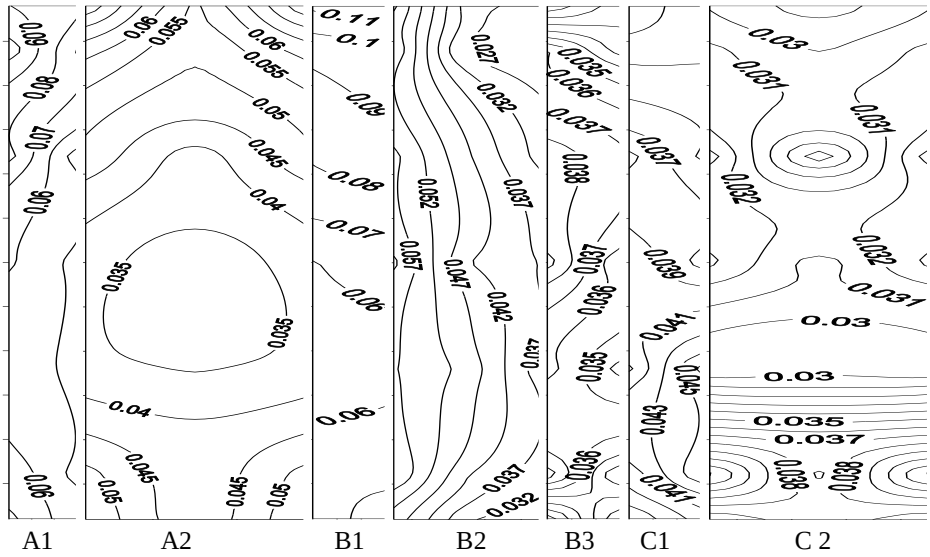


Fig. 10 Distribution of rms pressure coefficients (full blockage)

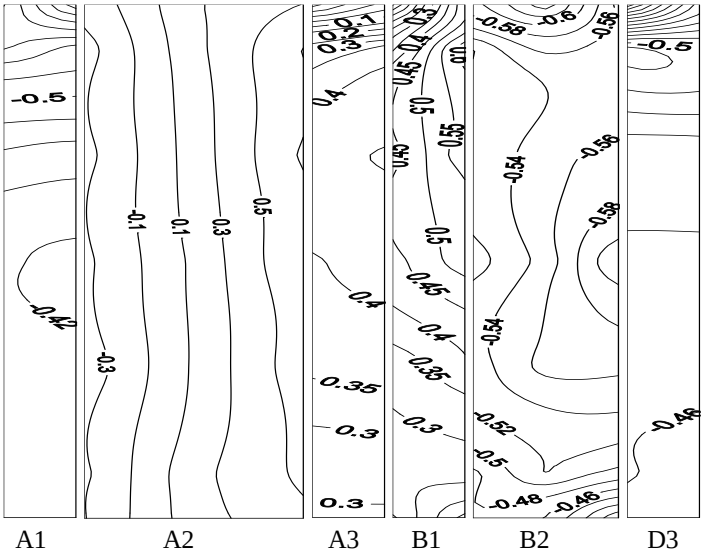


Fig. 11 Distribution of $C_{p,mean}$ on windward faces (half blockage)

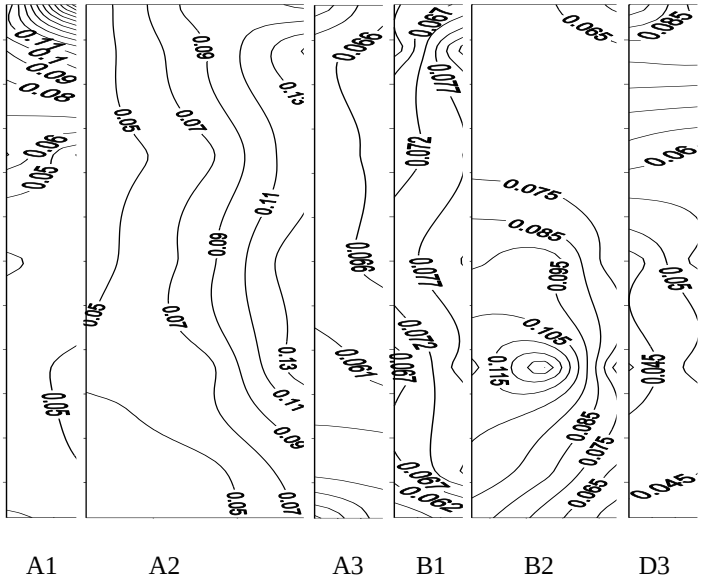
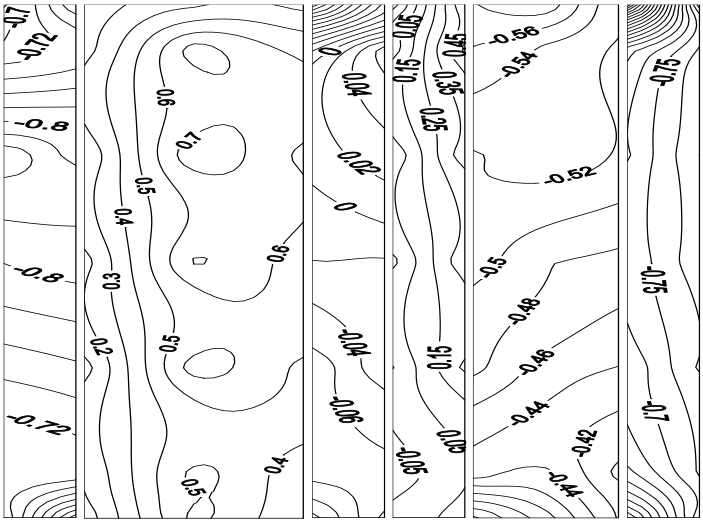
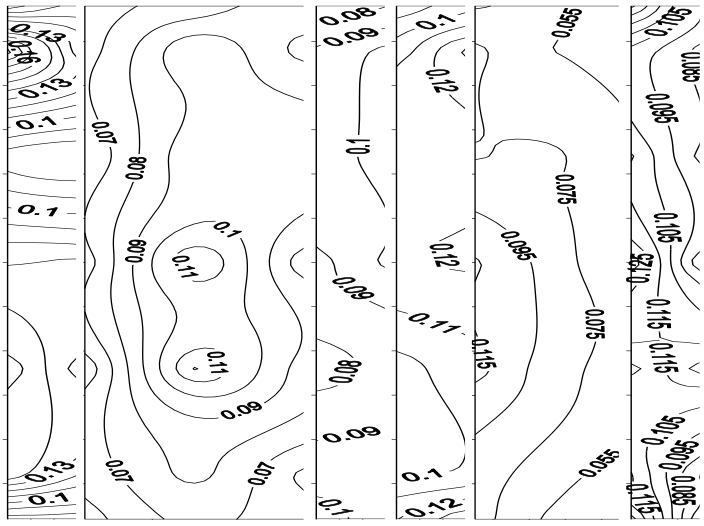


Fig. 12 Distribution of $C_{p,rms}$ on windward faces (half blockage)



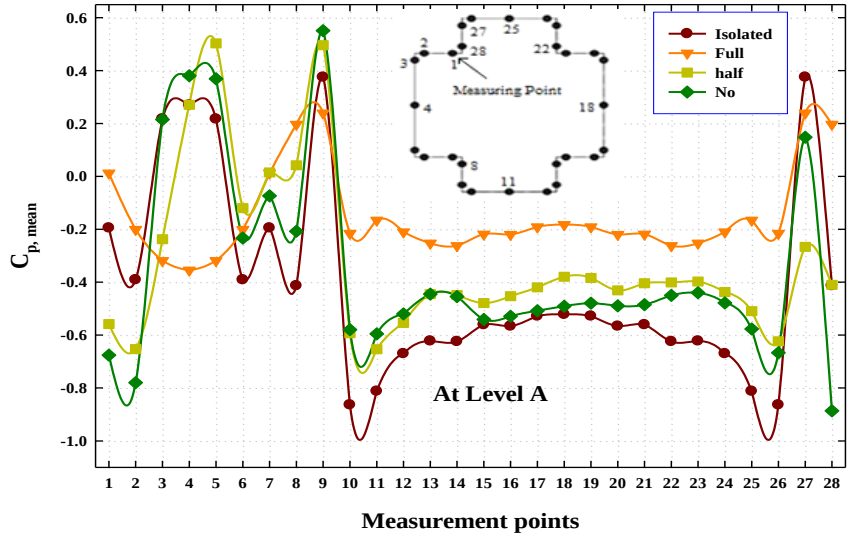
A1 A2 A3 B1 B2 D3

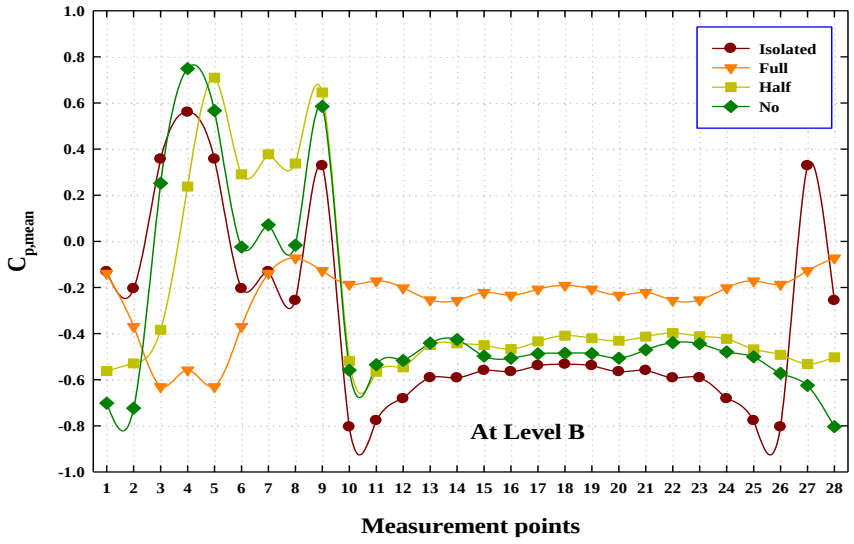
Fig. 13 Distribution of $C_{p,mean}$ on windward faces (no blockage)

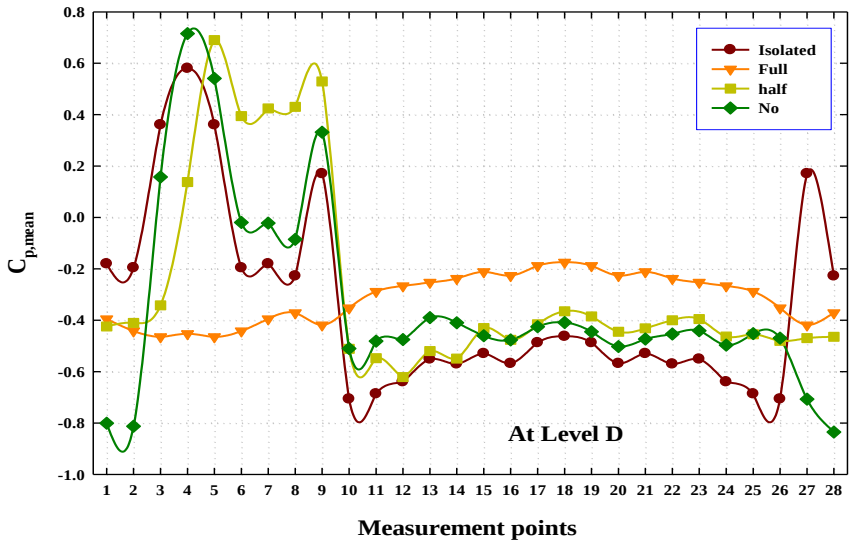


A1 A2 A3 B1 B2 D3

Fig. 14 Distribution of $C_{p,rms}$ on windward faces (no blockage)


Isolated
Full
half
No
Measuring Point
At Level A
$C_{p,mean}$
Measurement points


Isolated
Full
Half
No
At Level B
$C_{p,mean}$
Measurement points


Isolated
Full
half
No
At Level D
$C_{p,mean}$
Measurement points

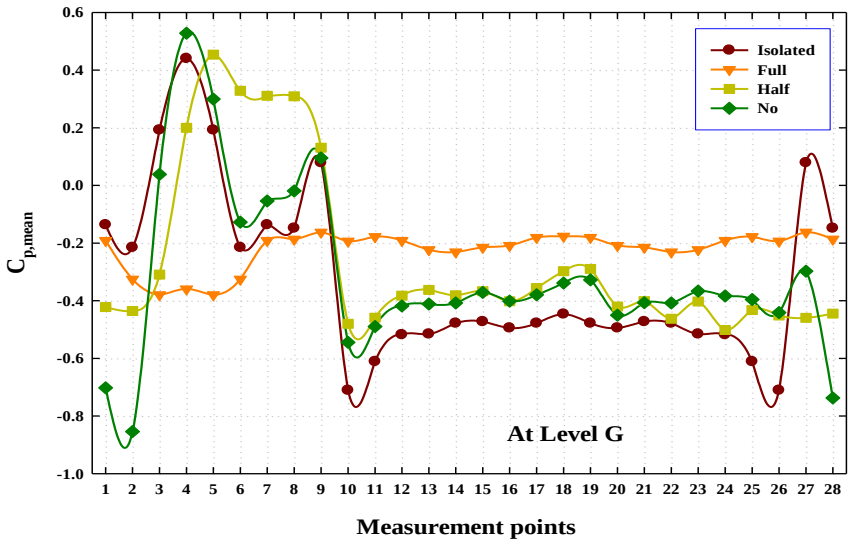
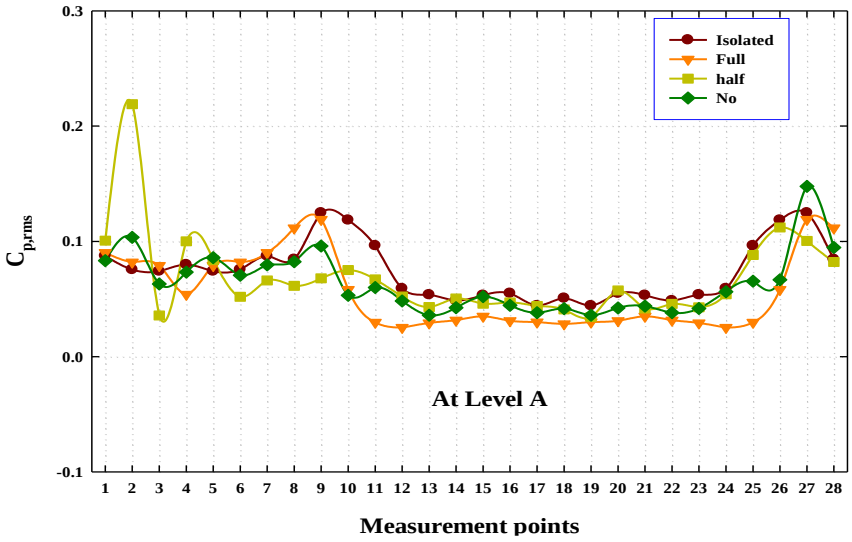
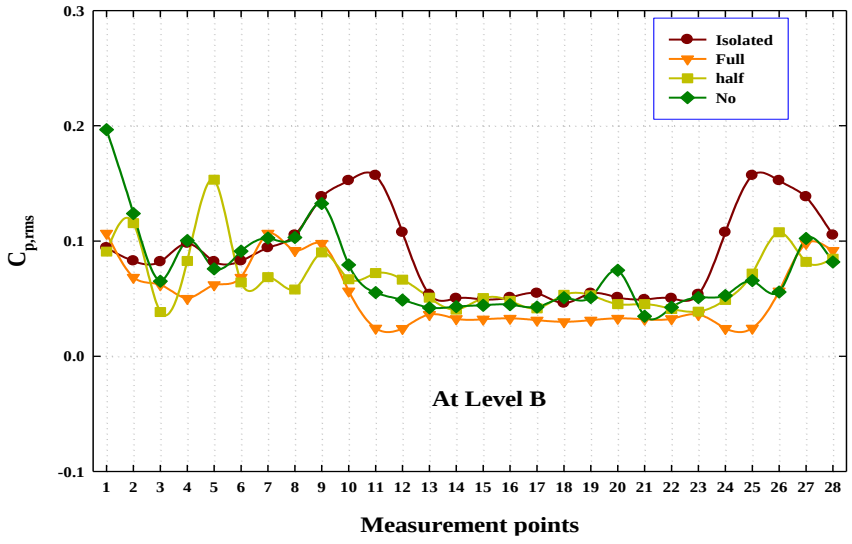
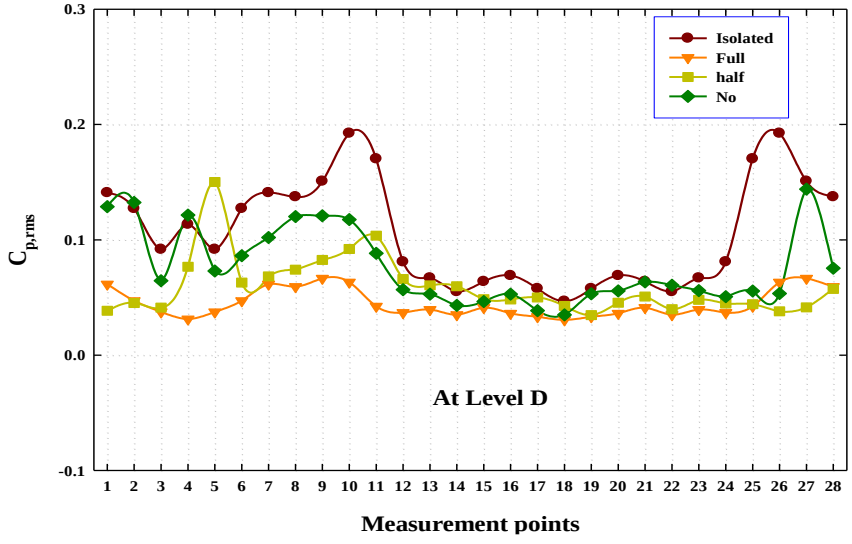


Fig.15 Mean pressure coefficients at different levels of principal buildings


Isolated
Full
half
No
$C_{p,rms}$
At Level A
Measurement points


Isolated
Full
half
No
$C_{p,rms}$
At Level B
Measurement points


Isolated
Full
half
No
$C_{p,rms}$
At Level D
Measurement points

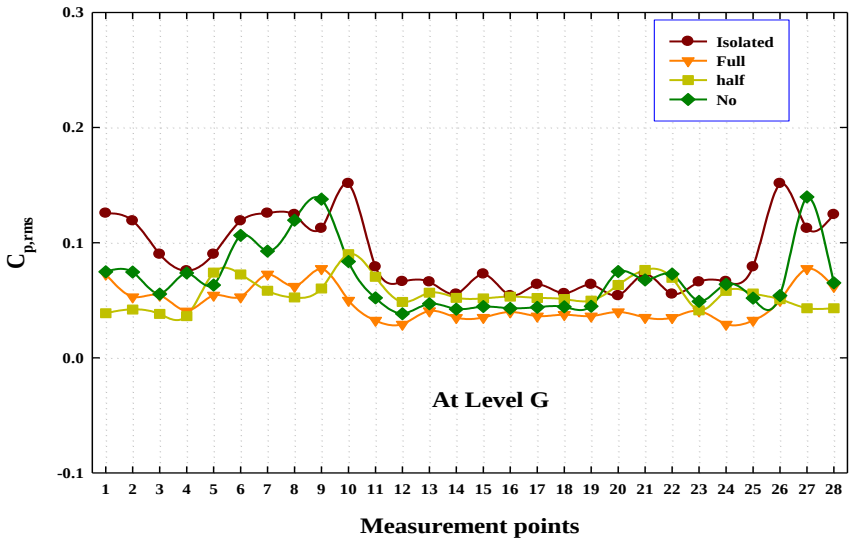
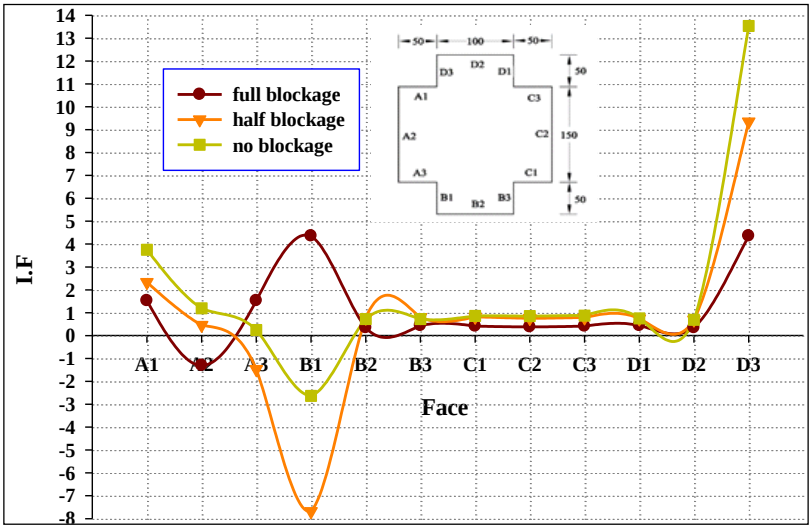
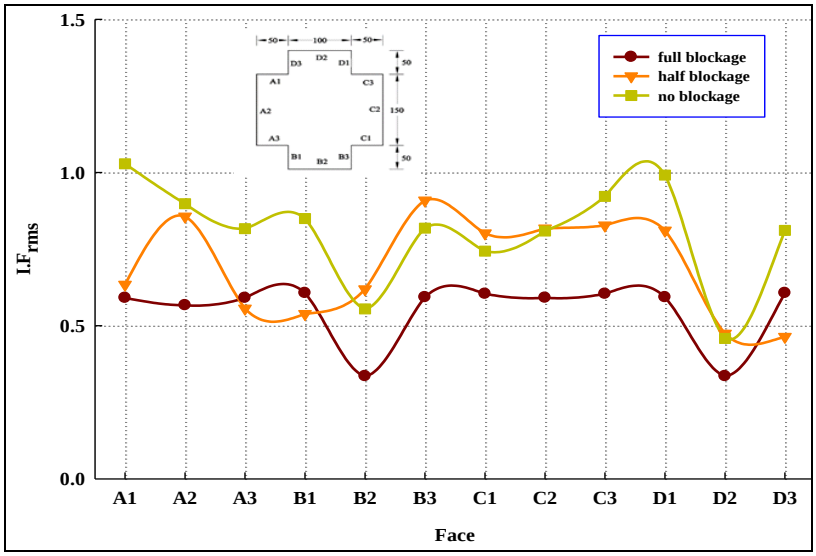


Fig.16 RMS pressure coefficients at different levels of principal buildings

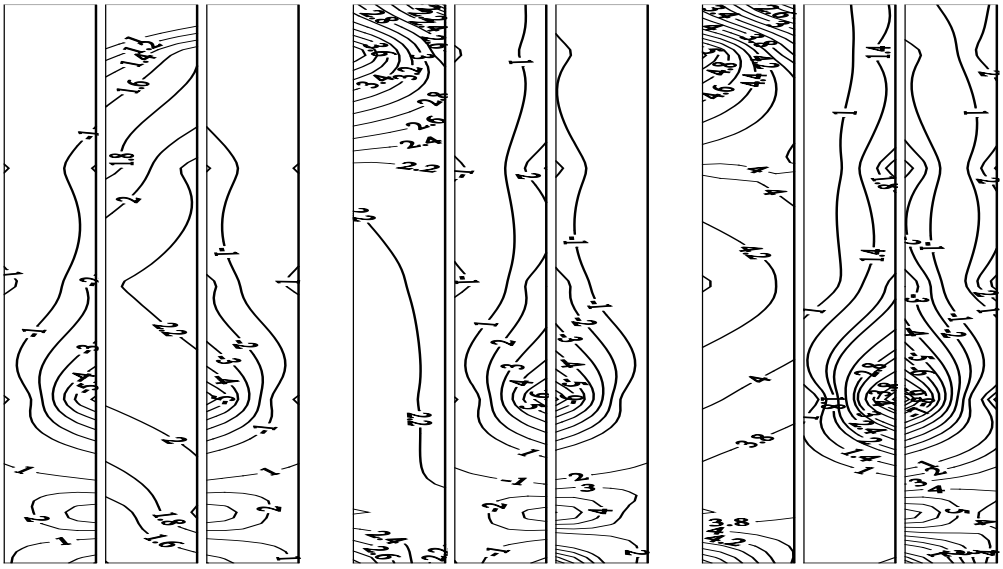


(a)



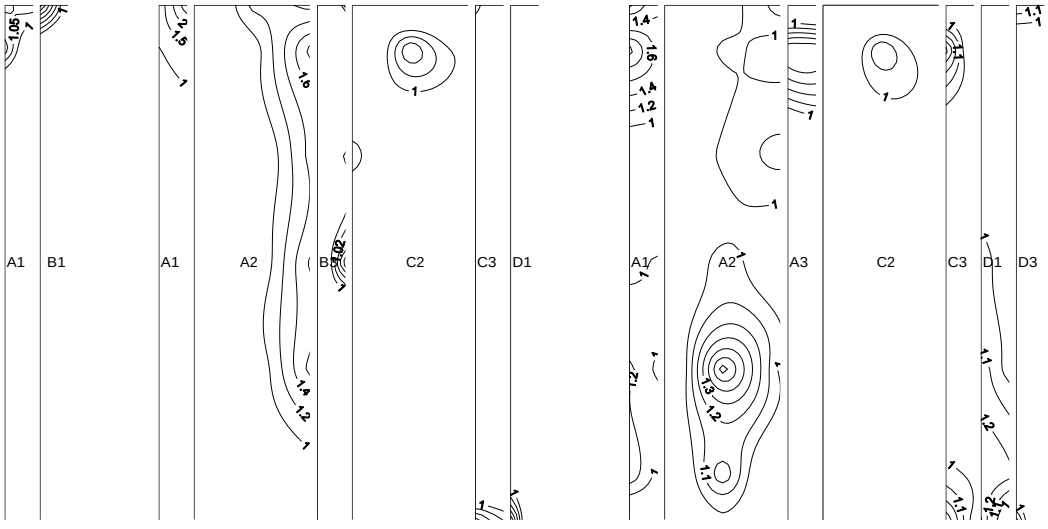
(b)

Fig. 17 Interference Factor on Faces (a) MIF; (b) RIF



(a) (b) (c)

Fig. 18 MIF (I.F ≥ 1) contour (Left to right face A1, B1 and D3) (a) Full (b) Half (c) No blockage



(a) (b) (c)

Fig. 19 RIF (IF ≥ 1) contour (a) Full blockage (b) Half blockage (c) no blockage

Biography of all contributing author.

Suresh Kumar Nagar is currently a currently a PhD student at the Department of Civil Engineering, Delhi Technological University, Delhi, India and Assistant Professor at Department of Civil Engineering, Rajasthan Technical University, Kota, Rajasthan India. He received his B.E from the M.B.M Engineering College, Jai Narain Vyas University, Jodhpur, India in the field of Civil Engineering in 2009 and his M.E in the field of Structural Engineering from Delhi University in 2011. His research interests include structural analysis, Wind effects on tall buildings, CFD. He published 20 research paper in journals and conferences.

Dr. Ritu Raj is Assistant Professor in Civil Engineering Department, DTU. He has published more than 50 technical papers in reputed journal and International conferences. He has supervised no. of Students at the level of graduates, postgraduates, and doctorates. His research interest includes behavior of wind load on the response of tall building, Aerodynamics Study, aero-elastic model study inside BLWT, Behavior of high strength concrete with and without fiber, earthquake engineering. He has more than 7.5 years of graduate and post-graduate teaching experience. Teaching Areas: Structure Engineering, Strength of Materials, Steel Structures, Design of Tall Buildings.

Prof. Nirendra Dev is Professor in Civil Engineering Department, DTU. He has more than 35 years of academic and educational administration experience. He has published numerous research paper in reputed journal and international conferences. His research interest includes Behavior of high strength concrete with and without fiber, earthquake engineering and concrete technology. He has supervised more than 45 students in gradates, post-graduates and doctorates level.

2021 2nd International Conference for Emerging Technology (INCET)
Belgaum, India. May 21-23, 2021

QnA System on Educational Textbooks: Digital Library Doubt Support System

Rahul Gupta
Department of Computer Engineering
Delhi Technological University
NCT of Delhi, India
rahulgupta100689@gmail.com

Gaurav Dabas
Department of Computer Engineering
Delhi Technological University
NCT of Delhi, India
dabasgaurav05@gmail.com

Himanshu Yadav
Department of Computer Engineering
Delhi Technological University
NCT of Delhi, India
hemanshuyadav1999@gmail.com

Nijdat Hasnain
Department of Computer Engineering
Delhi Technological University
NCT of Delhi, India
nijdathasnain1234@gmail.com

2021 2nd International Conference for Emerging Technology (INCET) | 978-1-7281-7029-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/INCET51464.2021.9456357

***Abstract* — The domain of content and knowledge on any particular topic of someone's interest is ever - growing. This broadening of the knowledge base of a subject combined with the easy accessibility to the Internet through various devices has resulted in easy access to the loads of content when a search related to the topic is made. But this gigantic collection of data also possesses a challenge. One needs to make some efforts to find the desired information. For this purpose, an individual might be required to go through a number of documents until he/she finds the answer that they craved for.**

This paper centers around a model to deal with this issue and helps in finding the relevant and precise answer to a query raised. To fulfill this purpose the model uses Natural Language Processing methods combined with knowledge graphs. The paper comes to a conclusion by presenting the performance of the model built.

Keywords — Natural Language Processing, Entity Extraction, Knowledge Graph, Parsing, Question - Answer

I. INTRODUCTION

In today's world, students are surrounded by massive chunks of information present in the form of full text documents on the Internet. But, generally a student might be more interested in knowing a direct and precise answer to his/her question rather than going through a full document to find the correct answer. All modern day search engines like Google , Microsoft Bing etc. have great capabilities but they typically display a list of documents or web pages related to the query raised by an individual. So in order to find the answer one needs to go through a complete document. Sometimes, it might also happen that a document displayed has a lot of information about the question searched but it does not include the specific information which the user is seeking.

For example,

For the following question :

"Which vitamins are present in carrots?"

Direct answers such as Vitamin A, C and K would be of greater interest to users instead of a large document which he/she needs to read to find the desired answer.

This paper facilitates finding direct answers to such questions provided the relevant text or information by using Natural Language Processing (NLP) techniques and knowledge graphs (the term made popular by Google in 2012). Since then the size of Google's knowledge graph (KG) has increased exponentially. The extent of growth has been such that Google's contains around 500 billion facts related to 5 billion entities as of May,2020. Various other organisations have also developed their knowledge graphs such as DBPedia by Wikipedia, WordNet etc. KG's have various other domain specific applications. They are used in the field of Medical Sciences, Financial Industry, Media industry and numerous other industries.

The rest of the paper is organised as follows. Section II is about related work. Section III describes the methodology used. Section IV presents our results. In section V we discuss the future scopes and limitations. Section VI lists the references.

II. RELATED WORK

A decent amount of work has been done in the past related to this field. Some of the works undertaken previously in this field are described below.

In [1] , a community QA system is deployed which when provided with multiple statements (comments) related to a question ranks them in order on the basis of good, bad and potential answers. To achieve this , the model uses convolution neural networks (CNN) with convolution filters 3,4,5. Further, long term short memory (LSTM) is used with 100 smart neurons as LSTM. Finally, the most suitable comment is provided as the answer for the question asked. This model shows an accuracy of 68.8 % with only CNN, 66.7 % with only LSTM and 72.5% with both CNN and LSTM.

Author in paper [2] has proposed a closed domain question answering system which works on education acts. Data related to education acts is collected and pre – processed using stemming and stop word removal. The keywords thus extracted are stored in an index term dictionary that contains the keyword as its key and a list of document numbers that contain the keyword corresponding to it. For a query, the keywords of the query are matched with those in the dictionary and the intersection of document numbers of all the matched keywords are retrieved and are then ranked using the jaccard similarity function. Then part of speech (POS) tagging is applied to the documents and finally the document in which the keywords have the same grammatical sense as the keywords of the query is returned as the final answer. As mentioned, this method is restricted and only returns the whole

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:28:06 UTC from IEEE Xplore. Restrictions apply.

document as an answer instead of the direct answer related to the query.

In [3] the author proposes a QA system based on DeepLSTM and versatile global T-max pooling. The model thus proposed, transforms the words in a text document to a word vector using word embedding algorithms such as word2vec. In the subsequent step the answer is predicted using a combination of deepLSTM and versatile global T-max pooling. For the purpose of ranking it uses a deep factorization classifier. Four different sources have been used for data collection such as Wikipedia. The model thus built shows an accuracy of 85.3 % on the Wikipedia dataset.

In paper [4], a method of classifying the questions according to their type has been discussed. The data set of questions used for this purpose is related to medical sciences collected from doctors and patients. It involves pre – processing the questions by stemming and removing stop words from them. Then it finds the keywords in the question and then determines the expected type of answer for that question. For example, in a question which contains the primary keyword, the answer that is expected is a "DESCRIPTION" and not a "LOCATION" or a "NUMBER". But, this model only categorises the questions and does not describe any method to predict answers for the questions.

Author in [5] proposed an intelligent QA system based on data related to basketball collected from various sources such as the NBA. Then Neo4j tool in python is used to construct a knowledge graph(KG). When a question is entered, then its similarity with the existing question in question bank is measured and if found similar, then the corresponding answer is directly returned. Otherwise, the question's entity is identified and is matched with corresponding entity in KG and all related attributes of the entity are obtained. Further, a logistic regression model is used to map the question with the attributes obtained and finally the target answer is retrieved. This system is limited only to a single entity and is not capable of finding associations in case of multiple entities.

III. METHODOLOGY

In this section we will present the overview for the data collection and its transformation which will form the basis of our structured dataset for question answering system and knowledge graph. Further we will present the algorithm for query retrieval with high precision.

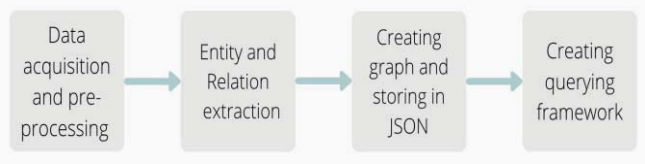


Fig. 1. Process Flow

A. Data acquisition

Even though our proposed system works on most kinds of texts but we chose our data to be textbooks of subjects like geography, history etc. in order to explain a good practical use of it.

The main source of our data is kaggle, from which we acquired some educational textbooks of the subjects of our interests (mainly theoretical subjects). Around 500 pages of each subject were considered. The data is in unstructured text format initially and not in the best form for our direct use. Thus the next step is to preprocess the data and make it suitable for our use.

B. Data Transformation

The first task is to preprocess the data to maintain uniformity in the text information. Preprocessing includes resolution of problems such as that of uppercase and lowercase in the data, conversion of pronouns into their corresponding proper nouns (also known as coherence property) to make the data more knowledgeable for the system.

The unstructured data further needs to be converted to a structured format i.e. entities and relations in order to build a knowledge graph from it and further query it too; such extracted structured data is then integrated into our knowledge graph.

C. Entity and relation extraction

We perform named entity recognition on the given data using spacy library in python. We accomplish the extraction of entities using some predefined grammar rules.

The basic idea to identify the entities is to find the subject and the object in the sentence. Similarly for relation, the verb of the sentence needs to be identified. Using parts of speech tagging we can directly find the nominal subject, direct object and the root verb in the sentence but the problem with this method is that mere nominal subject, direct object and root verb won't be enough for us to create an intelligent system. For creation of an efficient system we need to consider the whole subject, object and verb.

For example if we have a sentence, 'Indian cricketer Virat kohli is married to actress Anushka Sharma'. The nominal subject is 'Kohli', the direct object is 'Sharma' and the root verb is 'is', which is not capable of depicting any semantic knowledge. Thus, we need to consider the whole term 'Indian cricketer Virat kohli' as the subject, 'actress Anushka Sharma' as the subject and 'is married to' as the verb in order to be able to derive semantic information from these entities later. For this, we use dependency parsing along with a set of english grammar rules.

For example, to find the object of the sentence, we run a loop on the words of the sentence and check for the nsubj (nominal subject) or the nsubjpass (passive nominal subject) . Further we take in account all the compounds, nmod (nominal modifier), amod (adjectival modifier), poss (possessive nominal modifier), case (case marking) and nummods (numeric modifier) . Combining all of them in their original order will give us our required complete object. Similarly we obtain the object and the relation and we have our triplet.

We also need to take in account if any sentence has any time or place in it as it will help us to answer the questions having 'where' and 'when'. Thus to find this, for each sentence, we run a scan on the labels of each word of the sentence. If it is a date or gpe, we store it in its corresponding triplet making it a quintet. If the sentence does not have any place or time, we leave the fourth and fifth position of the quintet blank.

Quintet - [subject, predicate, object, time, place]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:28:06 UTC from IEEE Xplore. Restrictions apply.

Now that we have a quintet for each sentence in the form of a dataframe we need to store them.

D. Storing entities and relations in JSON format

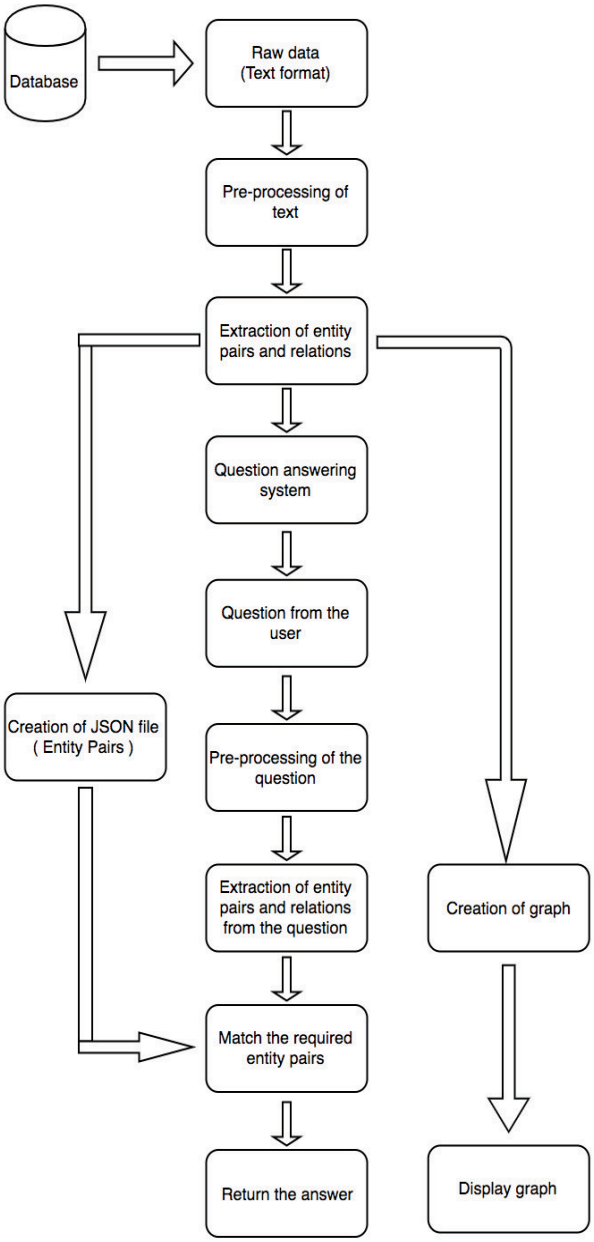


Fig. 2. Workflow chart

The derived quintets need to be stored in some structured format. Out of many format options like csv, xml, json etc. we have chosen json for the storage purpose. These quintets will be used at the time of creation of the knowledge graph and while querying the dataset.

E. Creation of knowledge graph

Knowledge graph is a method to display information by integrating information into an ontology and it helps the system to derive new knowledge from the graph.

It is used as it is a very flexible data structure and we can easily map our json file with the graph dynamically.

NetworkX library in python is used to display the graph. The subject, object and predicate is used from each quintet to form an edge of the graph.

F. Query retrieval system

The last step is to input a question from the user and answer it using our dataset. We mainly catered four types of simple questions in this system i.e. questions starting with what, who, when, where.

After preprocessing the user's question we make use of our self-defined algorithm shown below :

For questions containing **'WHO'**

```
for all entity pairs :
    if Object(Ques) == Object(entity_pair) and
        Relation(Ques) == Relation(entity_pair) :
            return Subject(entity_pair)
```

For questions containing **'WHAT'**

```
for all entity pairs :
    if Subject(Ques) == Subject(entity_pair) and
        Relation(Ques) == Relation(entity_pair) :
            return Object(entity_pair)
```

For question containing **'WHERE'**

```
for all entity pairs :
    if Subject(Ques) == Subject(entity_pair) and
        Relation(Ques) == Relation(entity_pair) and
        Object(Ques) == Object(entity_pair) :
            return Place(entity_pair)
```

For questions containing **'WHEN'**

```
for all entity pairs :
    if Subject(Ques) == Subject(entity_pair) and
        Relation(Ques) == Relation(entity_pair) and
        Object(Ques) == Object(entity_pair) :
            return Time(entity_pair)
```

Fig. 3. Pseudocode for answer extraction.

Using these rules and our previous entity and relation extraction algorithm, we derive the entity and relation from the question and return the answer if present.

IV. Result and Conclusion

For evaluation purposes we have subjected our model to a total of 500 questions out of which 450 questions have answers in the text provided and 50 do not have answers in the data. The criteria used for performance evaluation are accuracy, precision, recall and F1 score.

For calculations following attributes are used :

TP(True Positives) : total number of correct answers. Our model predicted a total of 416 correct answers out of total 450 questions whose answers are present.

FP(False Positives) : total number of incorrect answers predicted. The number of such answers in our model was 34.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:28:06 UTC from IEEE Xplore. Restrictions apply.

TN(True Negatives) : number of questions whose answers are not present in the data and hence no answer was predicted by the model. Such questions accounted to a total of 29.

FN(False Negatives) : number of questions whose answers are not present in text but the model predicted some answer. A total of 21 questions lie in this category.

Calculations are done as follows :

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

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

$$F1\ Score = \frac{2 * Precision * Recall}{Precison + Recall}$$

Table 1 displays the values of the above mentioned measures.

TABLE I. PERFORMANCE MEASURES FOR OUR MODEL.

MEASURE	Value
Precision	92.4 %
Recall	95.19 %
Accuracy	89 %
F1 Score	93.77 %

As shown in table 1 our model shows an accuracy of 89% , precision of 92.4% and recall of 95.19 % . The corresponding F1 score is 93.77 %

V. LIMITATIONS AND FUTURE SCOPE

Some of the limitations of the proposed system are :

- The rules are specified for limited sentences only, i.e. only 4 kinds of questions and have a non-complex structure.
- Redundant entities will be generated by sentences with similar meanings but distinct structures.
- extraction of entities takes into account the dependency tree of the sentence as the concept is highly dependent on parsing of the dependency tree.

Some of the future scopes of the proposed system are:

- Model can be modified to include voice recognition for the purpose of taking questions and returning answers which can be beneficial for visually impaired students.
- The algorithm of solving complex questions can be implemented by creating their own individual rules.
- The data accumulated can be trained to create neural networks for improving the QnA system.
- To boost accuracy, user reviews can be taken for accurate responses.

REFERENCES

[1] M. Wakchaure and P. Kulkarni, "A Scheme of Answer Selection In Community Question Answering Using Machine Learning Techniques," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 2019, pp. 879-883, doi: 10.1109/ICCS45141.2019.9065834.

[2] S. P. Lende and M. M. Raghuwanshi, "Question answering system on education acts using NLP techniques," *2016 World Conference on Futuristic Trends in Research & Innovation for Social Welfare (Startup Conclave)*, Coimbatore, 2016, pp. 1-6, doi: 10.1109/STARTUP.2016.7583963.

[3] D. V. Vekariya and N. R. Limbasiya, "A Novel Approach for Semantic Similarity Measurement for High Quality Answer Selection in Question Answering using Deep Learning Methods," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 518-522, doi: 10.1109/ICACCS48705.2020.9074471.

[4] T. Dodiya and S. Jain, "Question classification for medical domain Question Answering system," 2016 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), Pune, 2016, pp. 204-207, doi: 10.1109/WIECON-ECE.2016.8009118.

[5] Y. Li, J. Cao and Y. Wang, "Implementation of Intelligent Question Answering System Based on Basketball Knowledge Graph," 2019 IEEE 4th Advanced Info Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, 2019, pp. 2601-2604, doi: 10.1109/IAEAC47372.2019.8997747.

[6] D. Elalfy, W. Gad and R. Ismail, "A hybrid model to predict best answers in question answering communities", *Egypt. Inform. J.*, vol. 19, no. 1, pp. 21-31, 2018.

[7] D. Song et al., "Building and Querying an Enterprise Knowledge Graph," in IEEE Transactions on Services Computing, vol. 12, no. 3, pp. 356-369, 1 May-June 2019, doi: 10.1109/TSC.2017.2711600.

[8] Q. Shuai and C. Zhang, "Question Answering system based on Knowledge Graph of Film Culture," 2020 International Conference on Culture-oriented Science & Technology (ICCST), Beijing, China, 2020, pp. 150-153, doi: 10.1109/ICCST50977.2020.00035.

[9] Rani, K. Mehla and A. Jangra, "Parsers and parsing approaches: Classification and state of the art," 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), Noida, 2015, pp. 34-38, doi: 10.1109/ABLAZE.2015.7154963.

[10] T. Wang, Y. Wang and C. Tan, "Construction and Application of Knowledge Graph System in Computer Science," 2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), Jinan, China, 2018, pp. 169-172, doi: 10.1109/SPAC46244.2018.8965547.

[11] Xiaohan Zou 2020 J. Phys.: Conf. Ser. 1487012016

[12] Z. Chen, Y. Wang, B. Zhao, J. Cheng, X. Zhao and Z. Duan, "Knowledge Graph Completion: A Review," in IEEE Access, vol. 8, pp. 192435-192456, 2020, doi: 10.1109/ACCESS.2020.3030076.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:28:06 UTC from IEEE Xplore. Restrictions apply.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/352993288

Random Forest Classifier Based Network Intrusion Detection System

Article *in* Engineering, Technology and Applied Science Research · June 2021

DOI: 10.22214/ijraset.2021.35406

CITATIONS
0

READS
15

1 author:

Aadhar Dutta
Delhi Technological University
1 PUBLICATION **0** CITATIONS

SEE PROFILE

All content following this page was uploaded by Aadhar Dutta on 05 July 2021.

The user has requested enhancement of the downloaded file.

RANDOM FOREST CLASSIFIER BASED NETWORK INTRUSION DETECTION SYSTEM

Aadhar Dutta
Department Information Technology, Delhi Technological University, Delhi, India-110042
aadhardutta_2k18it001@dtu.ac.in

***Abstract*— In today's digital world, we all use the Internet and connect to a network, but all the data we send or receive, is safe? Some kind of attack is present in network packets that might access the computer's private information to the hacker. We cannot see and tell whether a network is safe to connect with or not, so we made a Network Intrusion Detection Model predict whether these network packets are secure or some attack is there on the package. We use Random Forest Classifier to obtain the maximum accuracy. To test our model in real-time, we have created a packet sniffer that would sniff out network packets, convert them into required features, and then try it in our model to predict the legitimacy of the network packet.**

Keywords*— *Intrusion Detection System, Network packets, packet sniffer, Random Forest Classifier, Anomaly Detection.

I. INTRODUCTION

Due to the web's general uses, electronic attacks on organizations and information systems of the monetary associations, military, and energy areas are expanding. Different intruders and hackers assault enormous sites of any association. The data of government and private associations might be spilled or harmed by unapproved clients. Intruders can have numerous structures, for example, viruses, spyware, worms, malicious logins, spam ware.

Data security is a significant aspect of securing critical information about any association. The associations need security applications that adequately shield their organizations from malicious assaults and abuse. The intrusion detection system can identify the intrusions and protect the data framework from security infringement. The intrusion detection system distinguishes intruders and makes a move against the intruder. It is utilized to recover the data by fixing the association's harm brought about by an unapproved client and recognizing the malicious utilization of P.C. and P.C. organization. It identifies admittance to an un-approved client, the infringement of security, and finds illegal clients.

There are two sorts of intrusion detection strategies, viz., misuse identification and anomaly. Misuse recognition is information or example-based, though anomaly identification is behaviour-based. Misuse location is dependable for identifying known assaults with low false positives. The misuse recognition method cannot distinguish the new assault. An anomaly identification procedure can recognize new assault with a high positive error. Existing intrusion detection systems have a high detection rate, though they experience high false alerts ill effects. The undertaking of decreasing false positives is very vital for the intrusion detection system.

Since the system has the upside of finding helpful information from datasets, different Machine learning approaches are implemented. These methodologies can diminish false positives. Bayes principle, Artificial Neural Network, Hidden Markov Model, Bayesian Belief Network, Genetic Algorithm, and Association of rules and Bunching strategies for machine learning are generally used to execute an intrusion detection system. The mix of various base Machine learning calculations is called an ensemble technique.

In writing a study, it is discovered that an Ensemble technique for Machine learning assists with diminishing false-positive rates. There are three principal strategies to join fundamental Machine learning classifiers viz., Bagging, Boosting, and Stacking. In this paper, the Bagging group technique for machine learning is proposed to execute an intrusion detection system. NSL_KDD dataset and Defense Advanced Research Projects Agency (DARPA) datasets are broadly utilized as preparing and testing datasets for the intrusion detection system. The NSL_KDD dataset gives 42 features dataset to train and test the intrusion detection system. However, every 42 elements in the dataset are not essential and needed for training and testing reasons. If all components are utilized to prepare intrusion detection systems, model structure time is expanded. The critical feature selection is a whole cycle in intrusion detection systems. In light of this examination, significant features of the NSL_KDD dataset are physically chosen, which improves the order exactness.

II. RELATED WORK

Lee et al.[1] in his paper has explained mature Data mining techniques, which could help significantly finding out the best way to detect Intrusion systematically; thus, we understand his report and further use his research to find much better results in our Random Forest Classifier technique.

Idan et al.[2] helped us interpret the proper understanding of Machine learning in the field of Cyber scrutiny and how we should be processing our datasets in a much better and informative way.

Aggarwal et al.[3] This paper is an essential paper to figure out how to change over TCP dump data to the 42 features in the KDD data set; thus, this paper explains the variable disparities in a very defined manner.

Wu et al.[4] In this research paper, we were able to comprehend how we can work towards building an maximum attack comprehension and detection while maintaining a minute low false alarming rate when worked out against ML concepts.

III. DATASET

The experimental arrangement is isolated into two stages. In the principal stage, the NSL_KDD dataset is pre-processed. In the pre-processing step, valuable features are chosen.

Feature selection is an essential information pre-preparing step to reduce the component of the dataset. A decrease in the measurement of the dataset prompts a better reasonable model. The NSL-KDD dataset has proposed 41 highlights to implement an intrusion detection system. Each of the 41 highlights is not needed to implement an intrusion detection system.

This paper has included 16 features for experimental work that would give high accuracy with low false-positive rates. These features are: -

TABLE I. Different features that give high accuracy with low false-positives.

Src-bytes	Number of data bytes transferred from source to destination in single connection.
count	Number of connections to the same destination host as the current connection in the past two seconds.
service	Destination network service used.
srv_count	Number of connections to the same service (port number) as the current connection in the past two seconds.
protocol_type	Protocol used in the connection.
diff_srv_rate	The percentage of connections that were to different services, among the connections aggregated in count.
same_srv_rate	The percentage of connections that were to the same service, among the links aggregated in count.
flag	Status of the connection – Normal or Error.

dst_bytes	Number of data bytes transferred from destination to source in single connection.
srv_serror_rate	The percentage of connections that have activated the flag s0, s1, s2 or s3, among the connections aggregated in srv_count.
logged_in	Login Status: 1 if successfully logged in; 0 otherwise.
duration	Length of time duration of the connection.
lnum_compromised	The number of "compromised" conditions.
wrong_fragment	The total number of wrong fragments in this connection.
is_guest_login	1 if the login is a "guest" login; 0 otherwise.

IV. Methodology & Model

The Basic Pipeline of our model is as follows: -

1. Feature Selection – We used WEKA to obtain the most correlated features. We selected a total of 16 most correlated features.

2. Tcpdump - Tcpdump application of Linux is used to dump all the network data transmission on all ports. Wifi is set as the default interface for the data dump, and it output the data in a cap file.

3. Feature Extractor- In the pcap file, Pyshark is used to analyze the data. This script basically creates all connection records in the pcap file, and then using pyshark, it scrapes all the connection and creates records according to our intrusion detection model.

4. Intrusion Detection - The finalized ML model is then used to predict whether the packet is regular or attack type.

5. Model – We divided the KDD dataset into a training and testing set of 65:35, and then we used the Random Forest Classifier of ensemble learning to train the model. We saved the model using a pickle, which we used to predict the packets in testing mode directly.

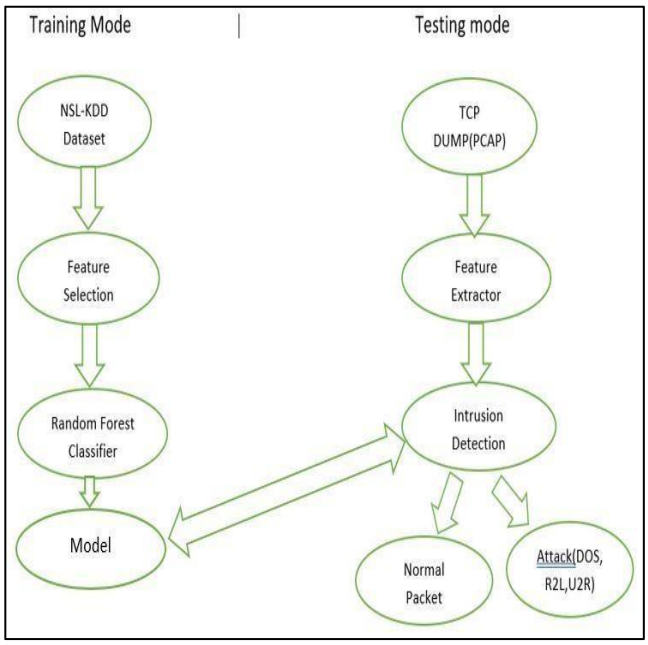


Fig. 1: Flowchart depicting the relationship between both Training Mode & Testing Mode.

The attacks are broadly being classified into four types: -

1. DOS – In this attack, the intruder seeks to make a machine or network resource unavailable to its intended users by Temporarily or indefinitely disrupting services of a host connected to the Internet.

2. R2L – This type of attack is launched by an attacker to gain unauthorized access to a victim machine in the entire network.

3. U2R – This type of attack is launched illegally to obtain the root's privileges when legally accessing a local machine.

4. Probe – A probe is an attack that is deliberately crafted so that its target detects and reports it with a recognizable "fingerprint" in the report. The attacker then uses the collaborative infrastructure to learn the detector's location and defensive capabilities from this report.

If the packet does not belong to these four types of attack, then it is classified as a regular or safe packet.

V. RESULTS

The features were extracted from WEKA, so these features were given to our model to train with the 65% of the KDD dataset; the Random Forest Classifier was used to train our dataset, we used other methods like K-NN and Decision Trees also to train our model, but thRandom Forest Classifier gave the best results Rest 35% of the dataset was tested on the trained model. It gave the best accuracy of 99% and had a shallow rate of false positives.

For real-time analysis of the model, we created a packet sniffer which would sniff real-time packets from the network, and then our feature extractor script decodes it to obtain the required features and then stores the data obtained in the CSV, and then our trained model tests this real-time data and classifies the packets into the five categories which are [Normal, DOS, U2R, R2L, Probe].

The best method to measure the accuracy is to calculate precision, Recall, and f1 score, which are given below: -

TABLE II. Classes to measure the Weighted Accuracy

Class	value
NORMAL	
True Positive	25479
True Negative	9291
False Positive	29
False Negative	17
DOS	
True Positive	8927
True Negative	26503
False Positive	11
False Negative	5
PROBE	
True Positive	229
True Negative	34566
False Positive	9
False Negative	11
R2L	
True Positive	119
True Negative	34674
False Positive	5
False Negative	18
U2R	
True Positive	6
True Negative	34804
False Positive	1
False Negative	5

The formula to find the value of Precision we can use the formula given below,

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

In the above formula Precision, the True Positive & False Positive are achieved using Table II.

The weighted accuracy acquired from training the above Precision Model, we get,

TABLE III. Weighted accuracy achieved from the Precision Model

ATTACK MODEL	**WEIGHTED ACCURACY**
NORMAL	~0.99
DOS	~0.99
PROBE	0.95
R2L	0.96
U2r	0.86

The formula to find the value of Recall we can use the procedure given below,

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

In the above formula for Recall, the True Positive & False Positive are achieved using Table II.

The weighted accuracy acquired from training the above Recall Model, we get,

TABLE IV. Weighted accuracy achieved from the Precision Model

ATTACK MODEL	WEIGHTED ACCURACY
NORMAL	~0.99
DOS	~0.99
PROBE	0.96
R2L	0.88
U2r	0.55

After achieving the above formulas for Precision and Recall, we can find out the F1 score by using the below procedure,

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

In the above formula for the F1 Score, the True Positive & False Positive are achieved using TABLE III & IV.

The weighted accuracy acquired from training the above F1 Score, we get,

TABLE V. Weighted accuracy achieved from the Precision Model

ATTACK MODEL	WEIGHTED ACCURACY
NORMAL	~0.99
DOS	~0.99
PROBE	0.96
R2L	0.92
U2r	0.67

VI. CONCLUSION

Today we live in a digital world, and every day, we tend to use the Internet and establish the connection on the network with the ports and services inside the network, so there are high risks of being attacked by a hacker who uses these kinds of attack to obtain access to monitor, so our network intrusion detector model can be used here to predict the legitimacy of the packets as it has an accuracy of 99% and could be used in an organization also for security purposes.

REFERENCES

[1] L. W and S. S.j, "Data mining approaches for intrusion detection. In: Proceedings of the 7th," *USENIX Secur. Symp. (SECURITY-98*, pp. p79--94, 1998, [Online]. Available: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwjBho6oo9vdAhVExoUKHdnYDFoQFjAAegQICRAC&url=https%3A%2F%2Fpdfs.semanticscholar.org%2F1a76%2Ff0539a9badf317b0b35ea92f734c62467138.pdf&usg=AOvVaw2lfvokfrG-_FA7LzIoDPHj.

[2] I. Amit, J. Matherly, W. Hewlett, Z. Xu, Y. Meshi, and Y. Weinberger, "Machine Learning in Cyber-Security - Problems, Challenges and Data Sets," 2018, [Online]. Available: http://arxiv.org/abs/1812.07858.

[3] P. Aggarwal and S. K. Sharma, "Analysis of KDD Dataset Attributes - Class wise for Intrusion Detection," *Procedia Comput. Sci.*, vol. 57, pp. 842–851, 2015, doi: 10.1016/j.procs.2015.07.490.

[4] P. Wu, H. Guo, and N. Moustafa, "Pelican : A Deep Residual Network for Network Intrusion Detection," no. Ml.

View publication stats

Global Journal of Flexible Systems Management (September 2021) 22(3):233–249
https://doi.org/10.1007/s40171-021-00276-8

ORIGINAL RESEARCH

Ranking the Factors Influencing e-Trading Usage in Agricultural Marketing

Sanjay Chaudhary[1] · **P. K. Suri**[2]

Received: 28 December 2020 / Accepted: 4 June 2021 / Published online: 1 July 2021
© Global Institute of Flexible Systems Management 2021

Abstract *This research article attempts to rank the factors influencing continued e-Trading usage in Indian agriculture marketing. The ranking process is the first of its kind application in agricultural e-trading. A review of published literature has helped in identifying key factors influencing e-trading usage in agricultural marketing. The factors are ranked using the efficient Interpretive Ranking Process (IRP) methodology adopted in the context of agricultural marketing. Sixteen expert members, grouped into three different expert panels, have helped generate ranks and validation besides giving suggestions for improving e-trading usage in the context of the National Agriculture Market project (eNAM) in India. It has been found that the top factors influencing e-trading usage are 'Trust', 'Cost', 'Perceived Ease of Use', and 'Facilitating Conditions', respectively. These factors need to be supported with adequate resources to strengthen eNAM in terms of improved usage among the beneficiaries. It is further revealed that immediate attention is needed on aspects such as transparency, quick information dissemination, adequate quality assurance, uniformity in taxes and market fee, improvement in marketing infrastructure, inter-market trade logistics, conflict resolution, mobility and training. Further, the e-trading system's flexibility needs to be enhanced by incorporating modular design options, configurable new features, open-source innovation, cloud computing, and progressive artificial intelligence application.*

Keywords Agriculture · Digitalization · E-commerce · E-trading · Flexibility · IRP · Marketing · Supply chain

Introduction

The Indian agricultural supply chain is fragmented and characterized by several non-value-add intermediaries. The information asymmetry in terms of demand and supply patterns is also high. The main reasons for this are the geographic dispersion of the agricultural trade markets and the lack of adequate agricultural marketing infrastructure. These markets have little coordination and bear the negative consequences of trader cartels.

One of the most challenging areas in the procurement/sourcing stage in the agricultural supply chain is when the buyers of crops (trader/agent/agribusinesses) interact with farmers. The digital solutions in the procurement/sourcing stage (Fig. 1) are vital to improving the supply chain. Digitalization helps the buyer in terms of transparency, easy monitoring of operations, and making transactions efficient. The farmer benefits from better access to markets, information, and services that help him adopt the recommended agricultural practices and transparent trading (Global System for Mobile Communications Association [GSMA], 2020).

The e-trading platform is viewed as a game-changer and a valuable instrument to address the supply chain improvement by offering an alternative to the rigid

✉ Sanjay Chaudhary
one.sanjay@gmail.com

P. K. Suri
pksuri@dtu.ac.in

[1] Delhi School of Business, Vivekananda Institute of Professional Studies, AU Block, Outer Ring Road, Pitampura, Delhi, India

[2] Delhi School of Management, Delhi Technological University, Shahabad Daulatpur, Main Bawana Road, Delhi, India

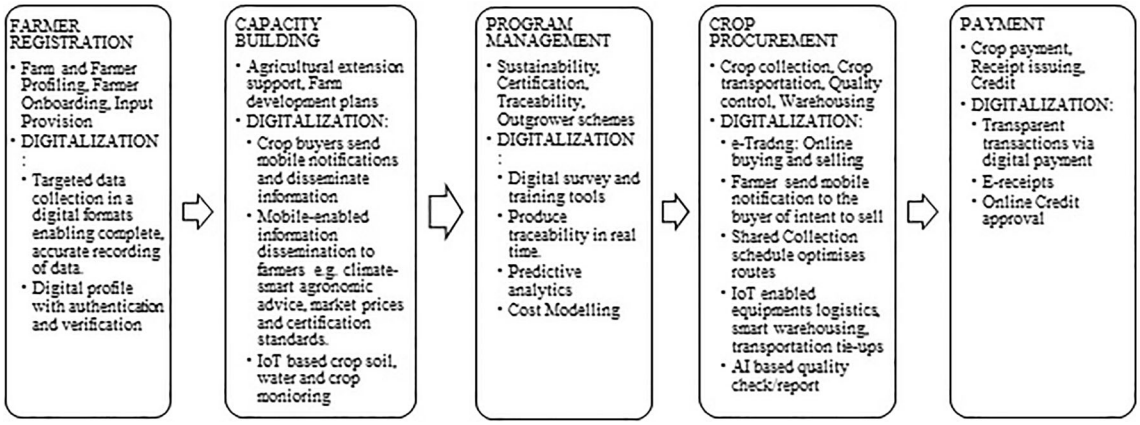


Fig. 1 Digitalization in Procurement (Source: GSMA, 2020)

Agricultural Produce Market Committee (APMC) controlled markets. It spans geographical boundaries, is economical, and reduces transaction costs besides improving price discovery (Smart & Harrison, 2003; Shirzad & Bell, 2013).

The electronic National Agriculture Market (eNAM) is a unique national e-trading platform. Within four years, it has been launched in 1000 wholesale markets—reaching 14 percent of its Year 2022 target of 7500 markets. As a result, it has emerged as the most effective and government-supported e-trading platform for agricultural produce in India. In comparison with eNAM, other corporate initiatives are relatively limited in scope and scale.

So far, just 14 percent of India's farmers have registered for eNAM. Out of these, only about 50 percent have started using eNAM. Thus, transaction growth and continued regular usage are relatively low. The domain experts have identified the need to improve the usage of eNAM among the participants (farmers, traders, corporate agents) to deliver the intended benefits (Hindu, 2019; Naik, 2019; Sajwan, 2020).

Empirically, it is shown that the high user registration and use are critical for deriving value in a B2B e-commerce system. The critical mass of farmers and traders with many usage transactions contributes to B2B e-marketplaces, e. g. eNAM. The high number of transactions also enhances market efficiency (Chircu & Kauffman, 2000; Subramaniam & Shaw, 2002; Li & Li, 2005; Engström & Salehi-Sangari, 2007; Chordia, Roll, & Subrahmanyam, 2008).

Thus, to ensure the success of an e-trading initiative in agricultural marketing, the factors responsible for increasing e-trading use need to be identified. The identified factors would help policymakers and market officials in the prioritization of the requisite incentives to strengthen eNAM in terms of these factors.

So, keeping the facts described above in mind, the research study aims to:

1. Rank the factors influencing the continued use of e-trading in agricultural marketing.

2. Bring out suggestions to improve upon the top-ranked factors for the continued use of e-trading in agricultural marketing in the Indian context.

The research article is organized as follows: The literature review section presents an understanding of factors influencing e-trading use in agricultural marketing and the national agriculture market project. Then, the research methodology section illustrates the Interpretive Ranking Process (IRP) methodology used to achieve the research objectives. Next, the analysis and findings section explains the data analysis undertaken, which is followed by research findings. Next, the discussion and limitations section brings out a few suggestions while mentioning the limitations of the research work. Finally, the article concludes with recommendations.

Literature Review

A brief literature review is conducted under the subsections: traditional agricultural produce wholesale marketing system in India, e-trading in eNAM, flexibility, and decision making. Literature is also reviewed to identify factors in the context of e-trading in agricultural marketing.

Traditional Agricultural Produce Wholesale Marketing System in India

Traditionally, transactions in wholesale agricultural markets are either in the form of an open auction, closed auction, or mutual agreement mode. Thus, only the registered farmers, registered agents, and licensed traders participate in wholesale trade.

A farmer brings his produce to the commission agent (CA) shop in the APMC area, who grades the product's quality. At open auctions (generally used for perishable produce), traders gather at the CA shop and announce their bid for the lot as per quality. The highest bidder gets the produce. In closed auctions (most popular auctions, generally used for grains), all bidders (buyers) write their bids on slips during the permitted time (2 h to 1 day)—for each lot. The APMC official confidentially selects the highest bid. The APMC official announces the highest bidder (buyer), who collects the produced lot from the CA shop after making the payment. Under mutual negotiation (generally used when there are very few, or a single buyer or the produce is bought by processor/mill), the lot price is mutually decided between the farmer and the trader/agent. The mutually decided price is subsequently reported to the APMC market office (Aggarwal et al., 2016; Suri, 2018).

As per the State APMC Act, the first sale of the notified agricultural commodities in a particular market area (such as wheat, maize, pulses, oilseeds, vegetables, and other food grains) must be made in the APMC controlled market through a licensed agent or a trader, after paying off due fees and tax.

The Indian agricultural supply chain is full of non-value-add intermediaries. Because of many go-betweens, the end consumer price escalation in the supply chain is more than sixty percent. Depending on the nature of produce, the original producer receives between twenty-eight and seventy-eight percent of the end consumer price. Thus, many intermediaries add more costs than the value (Patnaik, 2011; Kaur, 2015; National Horticulture Research & Development Foundation [NHRDF], 2018).

E-Trading in eNAM

As part of India's initiatives for reforming the agricultural marketing sector, India's Government launched the 'Agmarknet' project in the year 2000 (www.agmarknet.gov.in). This ongoing e-governance project's objective is limited to empowering farmers with the latest market information (Suri, 2005). Indian Government is now pro gressively deploying an e-trading platform, the 'National Agriculture Market' (eNAM), in all 2477 regulated whole sale agricultural markets and the linked 4,843 sub-market yards. It has been implemented in 1000 wholesale markets by May 2020, with the registration of 16.6 million farmers, 131 thousand traders, 73 thousand commission agents, and one thousand farmer producer organizations (Suri, 2018; Ministry of Agriculture and Farmers' Welfare [MOAFW], 2020; Press Information Bureau [PIB], 2020).

The eNAM platform has a virtual electronic trading portal on the front with a physical market ('Mandi') infrastructure at the back end. The end-to-end e-trading process activities (Fig. 2) include registration of farmers/traders/buyers/agents, lot entry at the gate, quantity and quality checks, trading, and payments held online in digital form. At the same time, the actual material movement happens in the physical market. In any market covered under eNAM, the select agricultural commodities are mandatorily traded online on e-NAM.

The e-trading transaction process (Fig. 3) is outlined below:

- Online bidding is held on the portal.
- The initial invoice is generated automatically on intra-market or inter-market trade confirmation (by e-NAM software) and shown to traders. The winning bidder also gets an email / SMS.
- The winning bidder deposits the amount online/offline (as per the sale deed, including market and agent's charges, labour/packaging charges).
- The system sends a confirmation message to the farmer/trader/agent on receiving the funds.
- The delivery happens as per the terms and conditions. It is either on the spot market or through a logistics service provider (arranged by the supplier/buyer) listed on the portal.
- As soon as the buyer or representative accepts the delivery, payment is made to the farmer/trader/agent in online mode through registration. The timing is T+1 business day routed through the e-NAM bank account.

Need for a Flexible E-Trading System

Due consideration must be given to the eNAM e-trading system flexibility. Here, e-trading flexibility is viewed as the capability to add new features as per changing user requirements. Flexibility is vital in all the three-stage of eNAM adoption, implementation, and post-implementation of the system. During the post-implementation stage of eNAM, it is vital to constantly undertake process redesign based on market functionaries' feedback to establish a versatile e-trading system that is flexible enough to accommodate the changing needs effectively. (Chen et al., 2009; Suri & Sushil, 2011; Kar & Rakshit, 2015; Evans & Bahrami, 2020).

The unified platform approach adopted by the eNAM e-trading system provides the necessary foundation for achieving increased flexibility in collaboration among farmers, traders, corporate and market officials. As a result, flexibility may help extend the eNAM life cycle and improve its effectiveness, thereby justifying the considerable investment made in the system (Aulkemeier et al., 2019).

In this research, e-trading system flexibility is viewed as 'perceived ease-of-use'. Prior research has shown that

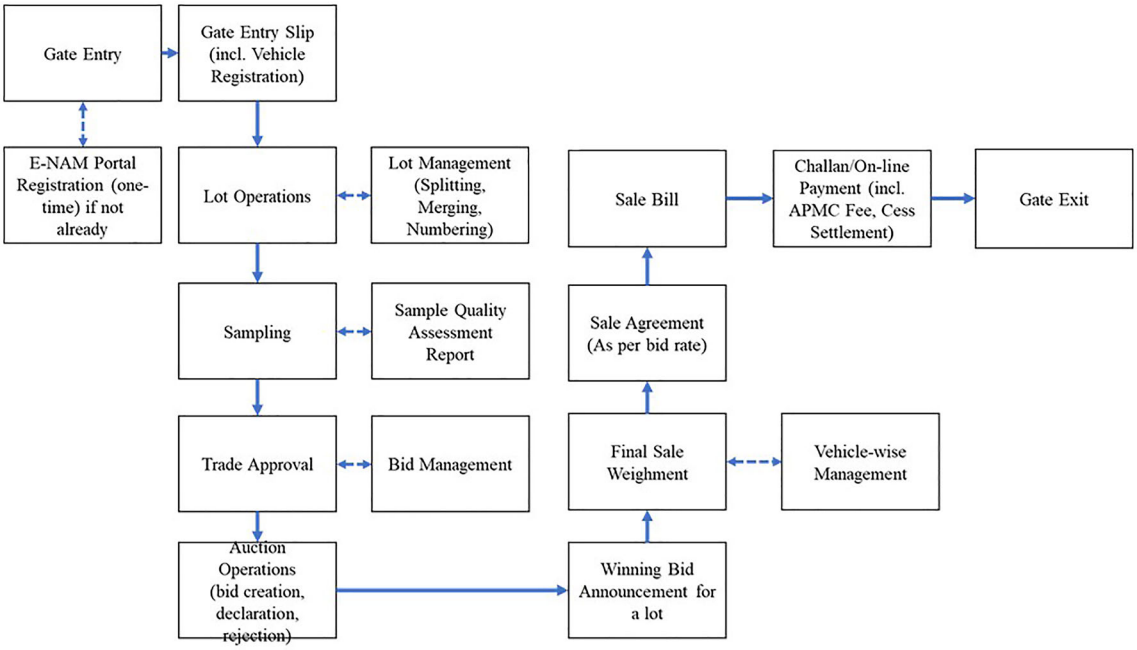
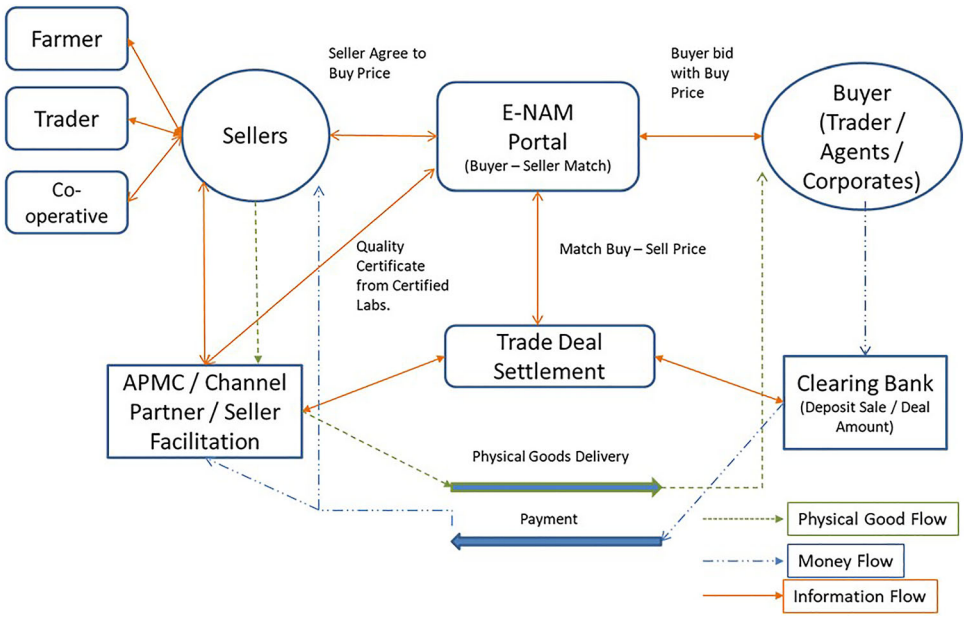


Fig. 2 Complete E-Trading Process (Source: MOAFW, 2020)

Fig. 3 eNAM Transaction Flow (Source: MOAFW, 2020)



perceived flexibility significantly directly affects user-perceived ease-of-use (Goodwin, 1987; Davis, 1989; Goeke & Faley, 2007).

The customer care feature 'Facilitating Conditions', and the quality linked pricing, with an option to do intra-market/inter-market trade under 'Perceived Usefulness' also address the flexibility aspect of the e-trading system operations (Bichler et al., 2002; Gorod et al., 2008; Zhang et al., 2010; Sushil, 2012; Singh et al., 2019).

Factors in the Context of E-Trading Use in Agricultural Marketing

The six factors are identified based on a review of the scholarly articles, reports, and news articles. The details are mentioned in Table 1

Table 1 Factors in e-trading use (Source: Adapted from Chaudhary & Suri, 2021)

S. No	Variable	Explanation in Study Context	References
1	Perceived Usefulness	The eNAM extends benefits to users in terms of better price and speed of the market transaction	(Tyndell et al., 1998), (Kilpatrick & Factor, 2000), (Venkatesh et al., 2003, 2012), (Dejen & Solomon, 2015), (Nirmal, 2017, 2019), (Pavithra et al., 2018), (Michels et al., 2019), (Salimi et al., 2020)
2	Perceived Ease of Use	The eNAM is simple in operation and learnt without difficulty	(Auramo et al., 2005), (Engotoit et al., 2016), (Michels et al., 2019)
3	Social Influence	Close friends, leading farmers, associated traders, and community leaders encourage or promote the eNAM	(Ajzen, 1991), (Venkatesh et al., 2003, 2012), (Dejen & Solomon, 2015), (Engotoit et al., 2016), (Dwivedi et al., 2019)
4	Trust	The user's confidence in the information and trade based on e-trading. It also signifies a belief that the eNAM is reliable and that the management will act in the interest of farmers and other stakeholders	(Moorman et al., 1992), (Killpatrick & Factor, 2000), (Ridings et al., 2002), (Kuttainen, 2005), (Casalo et al., 2011), (Ramesh et al., 2012), (Bisen & Kumar, 2018), (Jayashanker et al., 2018), (Wiyada et al., 2018)
5	Cost	The transaction costs in the eNAM platform	(Dodd et al., 1991), (Garicano & Kaplan, 2001), (Clasen & Mueller, 2006), (Solaymani et al., 2012), (Seifbarghy & Esfandiari, 2013), (Dey, 2016), (Nabhani, 2016), (Mustaqquim, 2017), (Wiyada et al., 2018)
6	Facilitating Conditions	The infrastructure, including quality testing laboratories, bidding halls, logistics support, IT/product training, and customer care, complements the use of eNAM	(Ajzen, 1991), (Bender, 2000), (Killpatrick & Factor, 2000), (Venkatesh et al., 2003, 2012), (Dejen & Solomon, 2015), (Tomar et al., 2016), (Dwivedi et al., 2019)

Decision-Making Process and Interpretive Ranking Process (IRP)

Decision making is a problem-solving activity ending with an optional or satisfactory solution. The decision-making process may be intuitive or rational choice-based (Kahneman, 2011). In the intuitive process, the information is gained through associated learning for storage in long-term memory. During decision making, we access it unconsciously to form the basis of a decision, e.g. matchmaker for marriage, car colour or model selection, meal selection while dining out, selecting the book from the best-seller list, choosing a dress from a wardrobe. In the rational choice process: we use a precise, analytical approach in a fact-based decision, e.g. the best solution to increase shop profits.

A large part of the decision-making process involves ranking the alternatives or finding the best option. Especially when multiple criteria (maybe conflicting) are present at the same time, e.g. investment in various portfolio stocks in a mutual fund (here we seek higher return but low risk), customer satisfaction versus the cost of service (Madhurika & Hemakumara, 2015).

Such problems can also be solved using multiple-criteria decision making (MCDM). The multiple criteria (which may involve various stakeholders) can be weighed implicitly based on intuition or explicitly evaluated while structuring the problem. These alternatives may also be known in advance or unknown (they may be generated using mathematical models). Then, the performance is checked under multiple criteria.

The various competing MCDM methods are the Analytical Hierarchy Process (AHP), the Analytical Network Process (ANP), the Data Envelopment Analysis (DEA), and the Interpretive Ranking Process (IRP) (Govindan et al., 2015).

Initially, within the supply chain domain, the AHP method was preferred as a reliable and flexible tool (Yahya & Kingsman, 1999; Muralidharan et al., 2002; Kar & Pani, 2014). However, during the last decade, the IRP method has been preferred by several researchers. The reason is the IRP method's capability to extract the expert knowledge and reasoning used for the rating during its pairwise comparison. Expert reasoning lacks expert judgements used during AHP and ANP (Triantaphyllou, 2000; Sushil, 2009; Ho et al., 2010).

The IRP methodology mixes the analytical logic (rational choice process) with an intuitive process. The IRP is essentially a knowledge-intensive MCDM method. It may be used in decision making, a process for ranking alternatives/actions/actors concerning criteria/performance/process.

The IRP's strengths include easily setting the interaction boundaries; the interaction details help in comparison; and quick decision concerning the relative dominance of one interaction with the other. The paired comparison reduces cognitive overload, reduces reliance on the factor or the variable weight, applies to any assemblage of variables, and different influencers with varied interests can be included in the evaluation to prevent preconceived opinions. The IRP is not software or calculation intensive. The knowledge generated can be re-used with added information for future decision making. The limitations of IRP include: the approach is judgemental and interpretive and is subject to biases, given that all criteria are considered as equal, it neglects their relativistic importance, objective validation tests are difficult to be administered, and interpretation of a matrix of size beyond ten by ten is very complex given the exponential increase in paired comparison (Warfield, 1974; Sushil, 2009).

The IRP method considers the subjectivity in prioritizing various perspectives and the consensus-building ranking through a systematic procedure, an essential dimension of group decision-making flexibility (Kar & Pani, 2014).

So far, it is used in several research areas, including the public value of e-governance projects, third-party logistics provider, mapping of IS failure factors, a ranking of flexibility initiatives, energy security and sustainability, lean implementation manufacturing sector, risks in the business analytics practices adoption, integrated supplier selection, world-class manufacturing, flexible supplier selection, and risk management in the supply chain (Gangotra & Shankar, 2016; Soni et al., 2016; Narkhede et al., 2017; Sushil, 2017a; Gupta & Suri, 2018; Kumar & Anbanandam, 2020).

Research Methodology

A systematic assessment of the scholarly articles, reports, and news articles relevant to the research has helped in identifying the factors influencing the continued e-trading usage in India's agricultural marketing. A semi-structured questionnaire for the experts was developed after reviewing recent academic articles.

The factors are ranked using the efficient Interpretive Ranking Process (IRP). The method is in sync with the similar use of an integrated evaluation tool using a fuzzy matrix in the agriculture sector (Girdziute, 2012). The IRP method involved developing binary and interpretive matrices, followed by a detailed interpretive logic knowledge base (Sushil, 2009). Then, a dominance matrix is prepared for ranking the factors concerning the three reference factors (criterion). The final Interpretive Ranking Model is based on a dominance index derived from the factors adjusted net dominance. Finally, the recommendations to support the top-ranked factors are made using learning from the case study of the eNAM project in four Indian cities, e.g. Meerut, Aligarh, Bharatpur, and Kashipur.

This study's trustworthiness is established using Lincoln and Guba's Evaluative Criteria of credibility, transferability, dependability, and conformability (Lincoln & Guba, 1985).

The credibility of qualitative research is established using referential adequacy. Then, a team of six experts from academia and industry is formed to apply the IRP method. After this, no new information is elicited by adding more experts, so a detailed questionnaire response was restricted to six experts in the first panel (Morse, 2000; Berg, 2001; Astalin, 2013).

The transferability is established using a detailed description (interpretive logic, knowledge base, and context for the recommendation) obtained through a National Agriculture Market (eNAM) case study.

The qualitative research's dependability and confirmability are established using the validity check of IRP findings by the five experts in the second panel.

The third panel of five experts deliberated on suggestions to improve factors leading to the high usage of the e-trading platform.

In all three expert panels, members are drawn from a mix of Academia, Industry, and Users.

Overall, the total number of 16 expert opinions obtained through questionnaire and semi-structured interviews conducted during the year is within the suggested range of 15 to 30 in the information systems research and 5 to 25 for the phenomenological studies (Creswell, 2007; Marshall et al., 2013). The outlined research process is suited to understanding and clarifying current issues and exploring new issues (Bryman, 2008).

Analysis and Findings

The Interpretive Ranking Process (IRP) has been applied to influencing e-trading usage in Indian agricultural marketing. The eight steps in the methodological process are outlined next.

Step 1: Identification of ranking factors concerning the reference factors.

The six ranking factors (Table 2) are identified based on a review of the scholarly articles, reports, and news articles relevant to the research, as detailed in Table 1. Three reference factors (Table 3), as suggested by experts, are also included.

Step 2: The relationship between the ranking and the reference factors in the context.

Table 2 Ranking factors

Code	Variable
F1	Perceived Usefulness
F2	Perceived Ease of Use
F3	Social Influence
F4	Trust
F5	Cost
F6	Facilitating Conditions

Table 3 Reference factors

Code	Reference factors
C1	Start actual transactions
C2	Frequency of transactions
C3	Total volume of transactions

If the factor 'Fx' enhances/influences (in terms of dominance) the factor 'Fy', the factor 'Fx' dominates the factor 'Fy'. If it is otherwise, then the factor 'Fx' is dominated by 'Fy'.

Step 3: Development of a cross-interaction matrix (CIM).

The CIM listing the contextual relationship between the ranking and reference factor is shown in Table 4. The cell value '1' indicates a contextual relationship, that the corresponding ranking factor 'Fx' 'enhances/influences' the reference factor 'Fy', and '0' shows no contextual relationship.

An interpretive matrix (Table 5) is developed by explaining the '1' values in the cross-interactions' matrix (Table 4) (Sushil, 2005).

Step 4: Dominating interaction matrix.

The pairwise interaction of ranking factors concerning the reference factor (s) listed in Table 4 is shown in Table 6. For each reference factor, the cell value for the pairwise dominance is '1' if the ranking factor 'Fx' dominates the ranking factor 'Fy'. If there is no dominance, then the cell value is '0'.

The overall reference factor-dominated interaction matrices for the three reference factors (C1, C2, C3) are collated (Table 7).

Step 5: Development of ranking and interpretation.

The number of dominating interactions in Table 6 is aggregated in a dominance matrix (Table 8). The weight of all three reference variables is '1': thus, initial column cell values in Table 8 equals '1*instance values in Table 6'. The total 'D' is the weighted sum of the number of instances where the ranking factor(s) dominate other ranking factors. The column 'B' is the weighted sum of the number of instances in which other ranking factors dominate a ranking factor. The difference between 'D' and 'B' is termed as a ranking factor's net dominance. The net dominance is adjusted to the scale by adding the positive value equivalent to the maximum negative net dominance value to each net-dominance value. Then, the index value is determined as the percentage of the total. The ranking factor with maximum dominance index value is ranked first, and so on.

The percentage of each type of paired comparison of ranking factors for a particular reference factor in Table 4, leading to the cell value entries in Table 8, is shown in Table 9.

In implicit dominance, if the ranking factor 'Fxi' has '1' entry and 'Fxj' has '0' entry for a reference factor, then cell entry is considered as '1' in Table 6. Suppose the ranking factor 'Fxi' has '0' entry and 'Fxj' has '1' entry for a reference factor. In that case, cell entry is considered as '0' in Table 6. For implicit non-dominance comparison, both ranking factors 'Fxi' and 'Fxj' have '0' entry for a reference factor, leading to '0' cell entry.

Table 4 Cross-interaction matrix

Alternatives (ranking factor)	Code	Criteria (reference factor)		
		Start actual transactions C1	Frequency of transactions C2	Total volume of transactions C3
Perceived Usefulness	F1	1	1	1
Perceived Ease of Use	F2	1	1	0
Social Influence	F3	1	0	0
Trust	F4	1	1	1
Cost	F5	0	1	1
Facilitating Conditions	F6	1	1	0

Table 5 Interpretive matrix

Alternatives (ranking factor)	Code	Criteria (reference factor)		
		Start actual transactions C1	Frequency of transactions C2	Total volume of transactions C3
Perceived Usefulness	F1	A trial of a useful (possibility of higher price, quick transaction cycle) system	Higher price/return than the offline market average due to quality linked pricing and reduction in trade to payment cycle time	Why not use e-trading for more significant transactions? Even aggregation of produce and e-trade is attempted for small farmers
Perceived Ease of Use	F2	A simple procedure, intuitive website, and the logical graphical user interface of the application. APMC support staff is available to help	The user gains confidence in handling transactions due to familiarity with the website/mobile application. The intuitive and logical graphical user interface in the local language also makes the use easy	
Social Influence	F3	Many farmers adopt due to uncertainty reduction and encouragement from influencers as a follower approach		
Trust	F4	Knowledge acquisition/Awareness Camps build trust in the Government promoted platform	Using knowledge, transparency, and experience for further gains from e-trade	Taking advantage of application expertise, process transparency, and e-trade experience for higher benefits
Cost	F5		After comparing with the offline market, the user makes decisions, and lower transaction cost matters for traders and farmers	An incentive of better margins on higher volume
Facilitating Conditions	F6	Understanding of stakeholders, physical infrastructure commitments, quality laboratories, training, and customer care staff to answer queries and get going on the platform	Increased realization of benefits. The quality linked pricing and inter-market trade possibility both increase trade frequency	

Table 6 Dominance interaction matrix for each reference factor

FOR C1 (Weight of reference factor = 1)							FOR C2 (Weight of reference factor = 1)							FOR C3 (Weight of reference factor = 1)						
Code	F1	F2	F3	F4	F5	F6	Code	F1	F2	F3	F4	F5	F6	Code	F1	F2	F3	F4	F5	F6
F1	0	0	0	0	1	1	F1	0	0	1	0	0	0	F1	0	1	1	0	0	1
F2	1	0	1	1	1	0	F2	1	0	1	0	0	0	F2	0	0	0	0	0	0
F3	1	0	0	0	1	0	F3	0	0	0	0	0	0	F3	0	0	0	0	0	0
F4	1	0	0	0	1	0	F4	1	1	1	0	1	0	F4	1	1	1	0	0	1
F5	0	0	0	0	0	0	F5	1	1	1	0	0	0	F5	1	1	1	1	0	1
F6	0	0	0	0	1	0	F6	0	1	1	1	0	0	F6	0	0	0	0	0	0

If both 'Fxi' and 'Fxj' cells have '1' entries in Table 4, we refer to the interpretive matrix (Table 5). If the interpretation is the same, then the corresponding entry is '0' in Table 6, with implicit non-dominance. However, if the interpretations are different, then the expert opinion is taken. The opinions lead to the interpretive dominance comparison. If 'Fxi' dominates 'Fxj', then the corresponding cell entry in Table 6 is '1', and vice versa. Sometimes, the cell values' Fxi', 'Fxj', and 'Fxk', are all '1' in Table 4. Then, if 'Fxi' dominates 'Fxj', and 'Fxj' dominates 'Fxk', 'Fxi' dominates 'Fxk' with corresponding paired dominance comparison cell entry '1' in Table 6 and marking it as the transitive dominance comparison in Table 9.

The IRP advancements are used as per the efficient IRP method due to the implicit or transitive dominance relationships, weights of reference variables, and computation of dominance index. The experts' interpretive dominance

Table 7 Overall reference factor wise dominant interactions matrix

Ranking factor Code	F1	F2	F3	F4	F5	F6
F1		C3	C2, C3		C1	C1, C3
F2	C1*, C2		C1, C2	C1	C1	
F3	C1				C1	
F4	C1, C2*, C3	C2*, C3	C2, C3		C1, C2	C3
F5	C2*, C3*	C2, C3	C2, C3	C3		C3
F6		C2*	C2	C2	C1	

Table 8 Dominance matrix

Ranking factor	Code	F1	F2	F3	F4	F5	F6	Number Dominating (D)	Net Dominance (D-B)	Adjusted Net Dominance	Dominance Index (%)	Rank Dominating
Perceived Usefulness	F1		1	2	0	1	2	6	− 2	5	12	4
Perceived Ease of Use	F2	2		2	1	1	0	6	0	7	17	3
Social Influence	F3	1	0		0	1	0	2	− 7	0	0	5
Trust	F4	3	2	2		2	1	10	7	14	33	1
Cost	F5	2	2	2	1		1	8	2	9	21	2
Facilitating Conditions	F6	0	1	1	1	1		4	0	7	17	3
Number Being Dominated (B)		8	6	9	3	6	4	36	0			

Table 9 Different types of dominance comparisons

Criteria (reference factor)	Implicit dominance comparisons	Implicit non-dominance comparisons	Transitive dominance comparisons	Interpretive dominance comparisons	Total comparisons	% Interpretive comparisons
C1	5	0	1	5	11	45
C2	5	0	4	4	13	31
C3	8	0	1	3	12	25
Total	18	0	6	12	36	
Percentage	50%	0%	17%	33%	36	

comparisons are limited to a small set of 33 percent (Table 9) or 12 comparisons, which are further detailed for suggestions after discussion with the second set of experts (Sushil, 2017b, 2020; Parmeshwar et al., 2020).

Step 6: Validation of ranks derived.

The ranking of factors derived from the dominance matrix (Table 8) is validated for the confidence-building in the ranking, which is interpretive. The multiple validations (Sushil, 2018) done are as follows:

- A structured walk-through for the cross-interaction matrix was done in the expert panel (second) workshop. All relevant ranking and reference factors were included. Though the list may not be exhaustive, a reference factor may be missed out.
- The interpretations and their wording were corrected based on a structured walk-through in an expert panel discussion.
- For the correct assessment of dominance relationships, the system graphs are drawn for the reference factors. The arrows in the digraph are expected to be unidirectional with no feedback loop/cycle. The feedback loop indicates unclear dominance relationships. The dominance interactions of various ranking factors concerning reference factors C1, C2, and C3 are unidirectional, as shown in Figure 4, thereby passing the internal validation test for the paired comparisons and its assessment.
- The cross-validation test is passed as the net dominance values' sum is zero in the dominance matrix (Table 8).

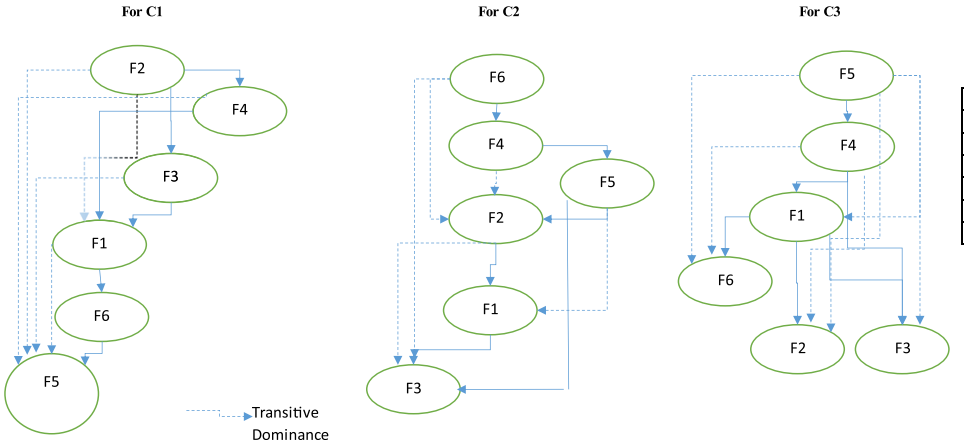


Fig. 4 Diagraph For Validity Check

- The sensitivity analysis gives different ordinal weights to the reference factors (Sushil, 2020). The results are summarized below (Table 10). The sensitivity analysis of ranking is not highly sensitive as the original weights assigned to reference factors changes, and evident modifications are minor. Thus, the efficient IRP ranking is quite robust.
- The rankings obtained were cross-checked in a five-member expert panel discussion.
- The real-life implications of ranking and related suggestions were discussed with a third expert panel. This includes prioritizing one factor over another. For example, suppose the usage of e-trading must be improved. In that case, the highest-ranking factor, 'Cost', may be given more management attention, supported with more financial and organizational resources, plus 'lower transaction cost' benefit may be promoted in public interactions.

Step 7: An 'Interpretive Ranking Model' diagram.

The ranks of six factors concerning their influence on the usage of e-trading in the Indian agriculture marketing sector are diagrammatically represented in an 'Interpretive Ranking Model" (IRM), shown in Fig. 5.

The arrows in the IRM diagram show a reference factor whose ranking factor dominates the other ranking factor. The interpretation of how a specific factor is influencing various reference factors is also provided. It can also be read in conjunction with the interpretive matrix (Table 5).

Step 8: Ranking decision and knowledge base.

The findings section details the IRP model output (Fig. 5). Based on the IRP model's ranking, one may prioritize the higher-ranking factors (Trust, Cost, Facilitating Conditions, Perceived Ease of Use). Suggestions for action are provided in the conclusion section of this research study.

The interpretive logic–knowledge base (Table 5) generated in this research study is the starting point for the conclusion section's suggestions. In the future, improvements may be made by adding more factors and additional learning about relationships.

Table 10 Comparison of ranks using sensitivity analysis

Code	Alternative (ranking factor)	Base Case Criteria (reference factor) weights: C1 = 1, C2 = 1, C3 = 1	Case 1 Criteria (reference factor) weights: C1 = 1, C2 = 2, C3 = 3	Case 2 Criteria (reference factor) weights: C1 = 3, C2 = 2, C3 = 1	Case 3 Criteria (reference factor) weights: C1 = 2, C2 = 3, C3 = 1	Case 4 Criteria (reference factor) weights: C1 = 2, C2 = 1, C3 = 3
F1	Perceived Usefulness	5	3	4	5	4
F2	Perceived Ease of Use	3	5	2	3	3
F3	Social Influence	6	6	6	6	6
F4	Trust	**1**	**1**	**1**	**1**	**1**
F5	Cost	2	2	4	4	2
F6	Facilitating Conditions	3	4	3	2	5

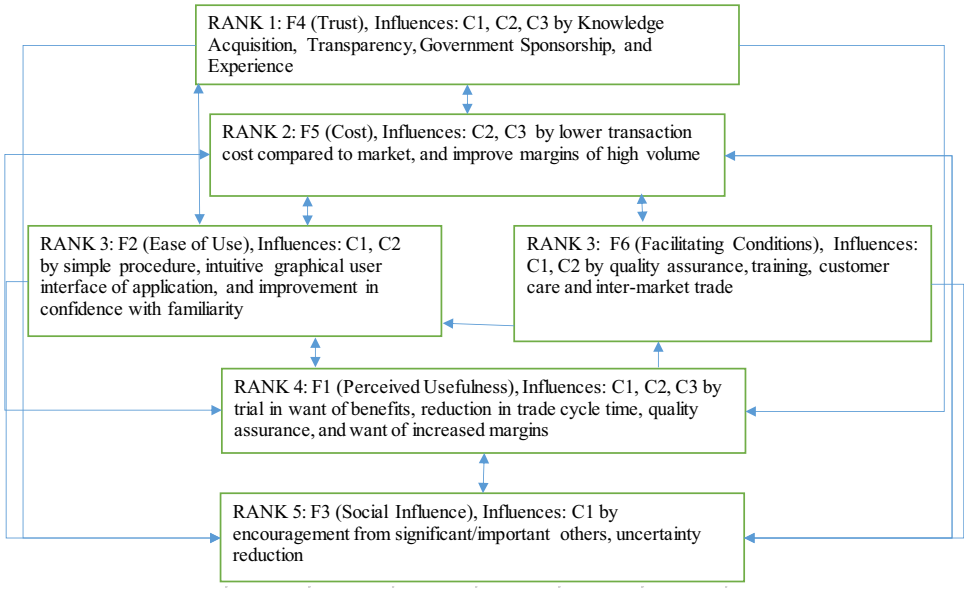


Fig. 5 Interpretive Ranking Model

Discussion and Limitations

The research findings are aligned with the research objectives, and based on the expert panel discussion, a few suggestions are brought out for the strengthening of the e-trading factors influencing the continued use of e-trading in agricultural marketing:

Theoretical Contributions

As per the first research objective of ranking the factors influencing the continued use of e-trading, the IRP model of six factors ranked based on three criteria is shown in Fig. 5. The rank values are taken from the dominance matrix (Table 8).

The most important (top) factor is 'Trust' in the hierarchy. It dominated ten factors-criteria situations and dominated in three, with an adjusted net dominance value of fourteen (highest). It influences criteria 'C1' (start actual transactions) via knowledge acquisition, criteria 'C2' (frequency of transactions) through an attempt to further gains from knowledge, transparency, and experience, and criteria 'C3' (total volume of transactions) in terms of attempts to get more benefits from the market knowledge, openness of process and application experience gained.

The second-ranking factor is 'Cost' as it dominated eight factor-criteria situations and got dominated in three. The adjusted net dominance value is nine. The 'Cost' (transaction costs) consists of search cost, coordination cost, motivation cost, and commitment cost. It influences criteria 'C2' (frequency of transactions) via cost comparison with the open market to get lower transaction cost and criteria 'C3' (total volume of transactions) regarding a user seeking better margins on higher volume.

The factors 'Facilitating Conditions' and 'Perceived Ease of Use' is at the third position. They dominate four factor-criteria situations and are dominated in four factor-criteria cases, thus having an adjusted net dominance value of seven. The factor 'Facilitating Conditions' influences the criteria 'C1' (doing transactions) by understanding stakeholders & infrastructure (IT system, bidding room, mobile application, quality laboratories, training, and support staff) that help to break inertia and try out the system. It influences criteria 'C2' (frequency of transactions) via trust-building, quality-linked pricing, higher bids in the inter-market trade, and the increasing realization of benefits leading to the system's frequent use. The factor 'Perceived Ease of Use' influences criteria 'C1' (doing transactions) through the easy graphical interface of the website and mobile application, plus support from the staff, and 'C2' (frequency of transactions) via in terms of gain in confidence post-usage and information availability in the local language.

The factor 'Perceived Usefulness' is in the fourth position with an adjusted net dominance value of five. The criteria 'C1' (doing transactions) is influenced in terms of an intent to get a higher price. The criteria 'C2' (frequency of transactions) is influenced through the possible realization of higher price due to quality linked pricing and payment realization within a day. The criteria 'C3' (total volume of transactions) is influenced due to an inherent desire to better the margins or benefits. The factor 'Social Influence takes the fifth spot'. The initial registration and system trial are affected by the influencer farmer/trader or the market official.

Implications for Practice

Discussion with the expert panel is undertaken to achieve the second research objective. Based on the expert panel discussion, the following suggestions are made for the strengthening of the e-trading continued usage in terms of influence factors:

The trust can be improved by incorporating a sound monitoring mechanism for assaying, grading, sorting, delivery, quality checking, and dispute resolution. At present, quality laboratories in many markets are dysfunctional due to a scarcity of equipment or staff. Inclusion and acceptance of authorized private laboratory reports for trade may be a solution. Also, suppose the quality of delivered agri-produce fails to meet the quality standard as per terms and conditions. In that case, a penalty on errant farmer/trader or payment reversal from the escrow account may be provisioned.

In a conventional setup, farmers are compelled to trust the commission agents because of a lack of financing options (e.g. for loans during off-season) and cash payments for products sold. To avoid such a situation, the e-payment in an e-trading transaction is quick on delivery. Moreover, the ATMs are made available in/near the market complex, from where the farmers can withdraw cash. Besides, various financing schemes by the banks (Kisan credit card, livestock credit card, Jan-Dhan cards, FPO credit guarantee, export benefits), the government schemes (PM: crop insurance scheme, annual financial support scheme, irrigation scheme, agri-development scheme, pension scheme, livestock insurance scheme) has reduced the dependence of farmers on the agents for financing.

The online cost is expected to be competitive since online transactions reduce transaction costs and information symmetry. Thus, e-trading provides a cost-effective channel and helps deliver overall welfare. However, the aggregate taxes and market fees at present are high and not uniform across the states. Currently, the taxes and fees at APMC markets (mandis) range from 3 percent (West Bengal) to 19.5 percent (Andhra Pradesh) of Minimum Support Price (Jha, 2007; Subramanian, 2016; Nanda, 2018).

The potential of the e-trading platform is reflected in inter-state and inter-market trade. With uniform taxes and fees, traders with one national license can bid across markets without the need for multiple registrations. In inter-market trade, the market fee goes to the originating/selling market. Thus, the buyer/trader does not have to search for markets with low/discounted market fees in uniformity. The uniformity will also simplify settlement procedures and reduce conflicts. Thus, taxes and fees may be lowered with uniformity across Indian states to promote inter-market trade.

Also, the timelines to full eNAM implementation and functioning must be monitored and controlled. The monitoring may be done by appointing dedicated officers as the users are dissatisfied with the agriculture extension officers, as learned during the field visit.

The Facilitating Conditions may be improved by ensuring the availability of the amenities online and in/around APMCs, e.g. quality assaying labs for grading, electronic weighing, warehousing facility, dedicated user support, and query resolution in coordination with other government schemes.

Smartphone usage is expected to boost access to digital agribusiness, as the rural user base is likely to reach at least 332 million in 2022. Ninety-two percent of the rural users' have access to the Internet, primarily through mobile phones; thus, the Internet user base in rural India shall be more than 305 million (Klynveld Peat Marwick Goerdeler [KPMG], 2020; Mahapatra, 2020; Sharma et al., 2020). The existing intuitive eNAM mobile application, when scaled to support all 22 major local languages with a better graphical user interface, is expected to have a broader user base as the local language Internet user base is growing at more than 13 percent annually. In addition, the features such as local language interface, video and image content with inbuilt Agri-dictionary, mobility, basic bank account (Jan-Dhan), and digital identity (Aadhar number) may further help farmers access digital services without intermediaries (Aravindh & Karthikeyan, 2018; Dhaygude & Chakraborty, 2020).

Each eNAM market has an e-trading room with computers, Internet connectivity, and trained staff (outsourced). The trained staff helps/handhold farmers visiting markets and organize weekly/monthly eNAM awareness sessions both within market premises and at local fairs. The existing limited information technology literacy efforts by eNAM staff, when integrated with the similar e-Kranti and Pradhan Mantri Gramin Digital Saksharta Abhiyaan (PMGDISHA) initiatives by the Government coupled with support from the gram sevaks, Krishi Vigan Kendras, and common service centres, may help strengthen the trained user base of eNAM (Modekurti, 2016; Raja Lakshmi, 2017; Singh, 2017; Department of Agriculture Cooperation and Farmers Welfare [DACFW], 2018; Hindustan Times Digital Content Services, 2018).

The e-trading acceptance is more for inter-market and inter-state trades, considering the higher prices expectation due to competition and an increased number of bids. In such cases, the logistics functionality of the e-trading portal may include preferential freight charges and preferential warehousing. The e-trade of aggregated produce by the farmer-producer organization (FPO) may help the marginal farmers to benefit from the e-trade. The mobile-based application's usage must be promoted to small farmers and

small traders so that e-trading functionality adds to the Mandi premises' offline presence.

The eNAM e-trading system's flexibility may be further enhanced by incorporating the system's capacity to address future uncertainty and risk management. Furthermore, the provision of flexibility in terms of modular design, configurable add-on new features, open-source innovation, and cloud computing functionality may increase the life cycle and return on investment in the system. Furthermore, the problem-solving and non-routine tasks may be augmented by artificial intelligence and the Internet of Things. In addition, the improvement in the customer relationship orientation of the eNAM enabled APMCs is expected to enhance their flexibility to adapt to the changing needs and market conditions, which may improve organizational performance (Tsai & Lasminar, 2021).

In recent times, data privacy and Internet security have raised concerns among users. In this regard, eNAM has taken due precautions to protect user-information digitally. The various methods used are access protection via password, web services with basic security profile 1.0, open standards for interoperability, SSL/TLS authentication mechanism for anti-spoofing, data encryption during storage, and transfer for confidentiality java-based customized security settings. As per the privacy policy, the user data is used only for digital transactions and shared with authorized employees and partners. Such features need to be further enriched from time to time with the emergence of new technological vulnerabilities.

Limitations and Future Research Directions

The ranking process is based on interpretive and judgemental methods and, thus, subjective. A limited number of factors have been considered for this study to avoid complexities associated with paired comparisons.

This research's efficient IRP process can be modified to include multiple interest groups/experts and academicians, and corporate employees approached for this research. In addition, the ordinal weights assigned to various reference factors may be changed as per the changing management priorities, as shown in the sensitivity analysis. In the current ranking process, the reference factors are qualitative. Depending on the situation, a mixed formulation involving both quantitative and qualitative factors can be used. The interpretive ranking may be used for qualitative factors, and the quantitative ranking may be done for quantitative factors.

Conclusion

E-trading is a precursor to the digitalization of the agricultural value chain (Agriculture 4.0). With its pan-India presence, eNAM is expected to deliver in terms of a large number of electronically held market transactions and high trade volume.

This research study ranks the factors influencing e-trading usage in Indian agricultural marketing while considering multiple criteria in the framework.

Indian policymakers and managers may consider the six factors (as per their relative priority in terms of ranks) for increasing the usage of e-trading in the Indian agricultural marketing sector. Therefore, the scarce resources and priority actions may be directed towards the top-ranking factors, e.g. lowering the transaction 'Cost' and increasing 'Trust' in e-trading while improving 'Facilitating Conditions' and 'Ease of Use' for ensuring maximum benefit/desired output.

Declarations

Funding No funds, grants, or other support was received.

Conflict of interest The authors declare that there is no conflict of interest.

References

Aggarwal, N., Jain, S., & Narayanan, S. (2016). *The long road to the transformation of agricultural markets in India: Lessons from Karnataka*. India. Indira Gandhi Institute of Development Research. http://www.igidr.ac.in/pdf/publication/WP-2016-026.pdf.

Ajzen, I. (1991). The Theory of Planned Behaviour. *Organizational Behaviour and Human Decision Processes, 50*(2), 179–211.

Aravindh, K. S., & Karthikeyan, C. (2018). The eminence of mobile agricultural apps in the worldwide Mobile network. *International Journal of Development Extension, 9*(1), 26–38.

Astalin, P. K. (2013). Qualitative research designs: A conceptual framework. *International Journal of Social Science & Interdisciplinary Research, 2*(1), 118–125.

Aulkemeier, F., Iacob, M. E., & Hillegersberg, J. V. (2019). Platform-based collaboration in digital ecosystems. *Electronic Markets, 29*, 597–608.

Auramo, J., Kauremaa, J., & Tanskanen, K. (2005). Benefits of IT in supply chain management: An explorative study of progressive companies. *International Journal of Physical Distribution & Logistics Management, 35*(2), 82–100.

Bender, P. S. (2000). Debunking five supply chain myths. *Supply Chain Management Review, 4*(1), 52–58.

Berg, B. L. (2001). *Qualitative Research Methods for the Social Sciences*. Allyn and Bacon.

Bichler, M., Kalagnanam, J., Katircioglu, K., King, A. J., Lawrence, R. D., & Lee, H. S. (2002). Applications of flexible pricing in business-to-business electronic commerce. *IBM Systems Journal, 41*(2), 287–302.

Bisen, J., & Kumar, R. (2018). Agricultural marketing reforms and national agriculture market (e-NAM) in India: A review. *Agricultural Economics Research Review, 31*, 167–176.

Bryman, A. (2008). *Social Research Methods*. Oxford University Press.

Casalo, L. V., Flavian, C., & Guinaliu,. (2011). The generation of trust in online services and product distribution: The case of Spanish electronic commerce. *Journal of Electronic Commerce Research, 12*(3), 199–2015.

Chaudhary, S., & Suri, P. K. (2021). *Modeling the Enablers of e-Trading Adoption in Agricultural Marketing: A TISM-Based Analysis of eNAM*. Advance online publication. https://doi.org/10.1177/0972262920977979

Chen, R. S., Sun, C. M., Helms, M. M., & Jih, W. J. (2009). Factors influencing information system flexibility: An interpretive flexibility perspective. *International Journal of Enterprise Information Systems, 5*(1), 32–43.

Chircu, A. M., & Kauffman, R. J. (2000). *A framework for performance and value assessment of e-business systems in corporate travel distribution*. Working Paper. http://misrc.umn.edu/workingpapers/fullpapers/2000/0024_112900.pdf.

Chordia, T., Roll, R., & Subrahmanyam, A. (2008). Liquidity and market efficiency. *Journal of Financial Economics, 87*(2), 249–268.

Clasen, M., & Mueller, R. A. E. (2006). Success factors of agribusiness digital marketplaces. *Electronic Markets, 16*(4), 349–360.

Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches*. Sage Publications Inc.

Davis, F. D. (1989). Perceived usefulness, perceived ease-of-use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 318–340.

Dejen, A., & Solomon, N. (2015). Mobile information system for small-scale rural farmers. *IEEE Technological Innovation in ICT for Agriculture and Rural Development (TIAR)*, 79–83.

Department of Agriculture Cooperation and Farmers Welfare (2018). *Agricultural Statistics, at a Glance 2017 Report*. Directorate of Economics and Statistics. Ministry of Agriculture and Farmers Welfare. http://agricoop.nic.in/.

Dey, K. (2016). National agricultural market-rationale rollout and ramifications. *Economic & Political Weekly, 51*(19), 35–39.

Dhaygude, M., & Chakraborty, D. (2020). Rethinking design of digital platforms for emergent users: Findings from a study with rural Indian farmers. *India HCI '20: Proceedings of the 11th Indian Conference on Human-Computer Interaction*. https://dl.acm.org/doi/proceedings/https://doi.org/10.1145/3429290.

Dodds, W. B., Monroe, K. B., & Grewal, D. (1991). Effects of price, brand, and store information on buyers. *Journal of Marketing Research, 28*(3), 307–319.

Dwivedi, Y. K., Rana, N. P., Jeyaraj, A., Marc, C., & Michael, D. W. (2019). Re-examining the unified theory of acceptance and use of technology (UTAUT): Towards a revised theoretical model. *Information System Frontiers, 21*, 719–734.

Engotoit, B., Musa, B., Moya, K. G., & Abima, B. (2016). A Mobile-based communication adoption model for agricultural market information dissemination in Uganda. *Global Journal of Computers & Technology, 5*(1), 249–265.

Engström, A., & Salehi-Sangari, E. (2007). *Assessment of Business-to-Business (B2B) e-marketplaces' Performance* [Doctoral Thesis, The Luleå University of Technology]. LUT Online Repository. https://www.diva-portal.org/smash/get/diva2:990718/FULLTEXT01.pdf.

Evans, S., & Bahrami, H. (2020). Super-Flexibility in practice: Insights from a crisis. *Global Journal of Flexible Systems Management*, *21*(3), 207–214.

Gangotra, A., & Shankar, R. (2016). Strategies in managing risks in the adoption of business analytics practices: A case study of a telecom service provider. *Journal of Enterprise Information Management, 29*(3), 374–399.

Garicano, L., & Kaplan, S. N. (2001). The effects of business-to-business E-commerce on transaction costs. *Journal of Industrial Economics, 49*(4), 463–485.

Gebauer, J., & Schober, F. (2006). Information system flexibility and the cost efficiency of business processes. *Journal of the Association for Information Systems, 7*(3), 122–146.

Girdžiūtė, L. (2012). Risks in agriculture and opportunities of their integrated evaluation. *Procedia-Social and Behavioral Sciences, 62*, 783–790.

Global System for Mobile Communications Association (2020). *The GSMA Agritech Toolkit for the Digitisation of Agricultural Value Chains*. GSMA Agritech Programme. https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2020/03/GSMA_AgriTech_Toolkit_for_the_Digitisation_of_Agricultural_Value_Chains1.pdf.

Goeke, R. J., & Faley, R. H. (2007). Leveraging the flexibility of your data warehouse. *Communications of the ACM, 50*(10), 107–111.

Goodwin, N. C. (1987). Functionality and usability. *Communications of the ACM, 30*(3), 229–233.

Gorod, A., Gandhi, S. J., Sauser, B., & Boardman, J. (2008). The flexibility of the system of systems. *Global Journal of Flexible Systems Management, 9*(4), 21–31.

Govindan, K., Rajendran, S., Sarkis, J., & Murugesan, P. (2015). Multi-criteria decision-making approaches for green supplier evaluation and selection: A literature review. *Journal of Clean Production, 98*, 66–83.

Gupta, P. J., & Suri, P. K. (2018). Analyzing the influence of improved situation, capability level of actors, and flexible process workflow on public value of e-governance projects in India. *Global Journal of Flexible Systems Management, 19*(4), 349–372.

Hindu (2019, July 10). Just 14% of farmers registered on the eNAM platform. *The Hindu Business Line*. https://www.thehindubusinessline.com/economy/agri-business/just-14-of-farmers-registered-on-enam-platform/article28363454.ece.

Hindustan Times Digital Content Services (2018, April 14). Integrate spot, derivative markets to ensure better prices for farmers: Panel. *The Mint*. https://search.proquest.com/docview/2024675783?accountid=173681.

Ho, W., Xu, X., & Dey, P. K. (2010). Multi-criteria decision-making approaches for supplier evaluation and selection: A literature review. *European Journal of Operational Research, 202*(1), 16–24.

Jayashanker, P., Nilakanta, S., Johnston, W. J., & Gill, P. (2018). IoT adoption in agriculture: The role of trust, perceived value, and risk. *Journal of Business & Industrial Marketing, 33*(6), 804–821.

Jha, S. (2007). Indian wheat and rice sector policies and the implications of reform. Economic research report. The United States. *Dept. of Agriculture. Economic Research Service, 41*, 4–5.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus, and Giroux.

Kar, A. K., & Pani, A. K. (2014). How can a group of procurement experts select suppliers? An approach for group decision support. *Journal of Enterprise Information Management, 27*(4), 337–357.

Kar, A. K., & Rakshit, A. (2015). Flexible pricing models for cloud computing based on group decision making under consensus. *Global Journal of Flexible Systems Management, 16*(2), 191–204.

Kaur, T. (2015). Use of E-Marketing in the indian agriculture sector to manage distribution challenges. *Anvesha, 8*(1), 1–6.

Kilpatrick, J., & Factor, R. (2000). Logistics in Canada survey: Tracking the year 2000 supply chain issues and trends. *Materials Management and Distribution, 45*(1), 16–20.

Klynveld Peat Marwick Goerdeler (2020). *Contribution of smartphones to digital governance in India*. https://icea.org.in/wp-content/uploads/2020/07/Contribution-of-Smartphones-to-Digital-Governance-in-India-09072020.pdf.

Kumar, A., & Anbanandam, R. (2020). A flexible policy framework for analysing multimodal freight transportation system in India: SAP–LAP and efficient IRP method. *Global Journal of Flexible Systems Management*, *21*(1), 35–52.

Kuttainen, C. (2005), *the Role of Trust in B2B Electronic Commerce - Evidence from Two e-Marketplaces* [PhD Thesis, The Luleå University of Technology]. LUT Research Repository. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.105.4726&rep=rep1&type=pdf.

Li, Jijian, & Li, Liwei (2005). On the Critical Success Factors for B2B E-Marketplace [Paper Presentation]. *Proceedings of the 7th International Conference on Electronic Commerce*. Association for Computing Machinery, USA (pp. 119–122). https://doi.org/10.1145/1089551.1089576.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Sage.

Madurika, H., & Hemakumara, G. (2015). A GIS-based analysis for suitability location finding in the residential development areas of the greater Matara Region. *International Journal of Scientific & Technology Research, 4*(8), 96–105.

Mahapatra, S. K. (2020). Smartphone Apps for agri-information dissemination during covid19 lockdown. *Biotica Research Today, 2*(5), 116–119.

Marshall, B., Cardon, P., Poddar, A., & Fontenot, R. (2013). Does sample size matter in qualitative research: a review of qualitative interviews is research. *Journal of Computer Information Systems, 54*(1), 11–22.

Michels, M., Bonke, V., & Mußhoff, O. (2019). Understanding the adoption of crop protection smartphone apps: An application of the Unified Theory of Acceptance and Use of Technology. DARE Discussion Papers 1905, Georg-August University of Göttingen. https://ideas.repec.org/p/zbw/daredp/1905.html.

Ministry of Agriculture and Farmers' Welfare (2020, August 08). *National Agriculture Market*. The Government of India. http://www.enam.gov.in/NAM/home/index.html.

Modekurti, D. P. V. (2016). Automation of the modified marketing procedural system to maximize transparency: A case study of vegetables in Madanapalle market. *Journal of Agribusiness in Developing and Emerging Economies, 6*(1), 72–88.

Moorman, C., Zaltman, C., & Deshpande, R. (1992). Relationships between providers and users of market research: The dynamics of trust within and between organizations. *Journal of Marketing Research, 29*, 314–328.

Morse, J. M. (2000). Determining the sample size. *Qualitative Health Research, 10*(1), 3–5.

Muralidharan, C., Anantharaman, N., & Deshmukh, S. G. (2002). A multi-criteria group decision-making model for supplier rating. *Journal of Supply Chain Management, 38*(4), 22–33.

Mustaquim, M. (2017, August 5). *Agriculture marketing reforms: benefiting farmers*. https://www.ruralmarketing.in/industry/agriculture/agriculture-marketing-reforms-benefiting-farmers.

Nabhani, I., & A., Daryanto, A., & Machfud, A. R. (2016). Mobile broadband for the farmers: a case study of technology adoption by cocoa Farmers in Southern East Java, Indonesia. *Agris Online Papers in Economics and Informatics, 8*(2), 111–120.

Naik, Amit Raja (2019, December 16). National Agri Marketplace eNAM Growth Slower Than Expected, Reveals Data. *Inc42*. https://inc42.com/buzz/national-india-agri-marketplace-enam-growth-slower-than-expected-reveals-data/.

Nanda, Kasabe (2018, February 28). 30 Maharashtra APMCs join eNAM. *Financial Express*. https://search.proquest.com/docview/2008663970?accountid=173681.

Narkhede, B. K., Raut, R., Gardas, B., Luong, H. T., & Jha, M. (2017). Selection and evaluation of third-party logistics service providers (3PLSP) by using an interpretive ranking process (IRP). *Benchmarking: An International Journal*, 24(6), 1597–1648.

National Horticulture Research & Development Foundation (2018). *Value Chain Study of Tomato in Karnal, Haryana*. New Delhi, India. https://midh.gov.in/VCS%20Reports/18-Tomato%20Value%20Chain%20Karnal%20Haryana.pdf.

Nirmal, R. (2017). E-NAM: Why inter-state trade is still a non-starter. *Indian Journal of Agricultural Marketing*, 31(3) (conference special.), 73–81.

Nirmal, R. (2019, January 14). How Suryapet has changed after eNAM. *Businessline*. https://search.proquest.com/docview/2263642337?accountid=173681.

Parmeshwar, N., Dhir, S., & Sushil. (2020). Interpretive ranking of choice of interaction of parent firms: Post-international joint venture termination using TISM-IRP. *Global Journal of Flexible Systems Management, 21*, 1–16. https://doi.org/10.1007/s40171-019-00227-4

Patnaik, G. (2011). *Status of Agricultural Marketing Reforms* [Paper Presentation]. Conference on Policy Options and Investment Priorities for Accelerating Agricultural Productivity and Development in India, Indira Gandhi Institute of Development Research, India. http://agritech.tnau.ac.in/amis/pdf/rs/Status_Agrl_Marketing_Reforms.pdf.

Pavithra, S., Gracy, C. P., Saxena, R., & Patil, G. G. (2018). Innovations in agricultural marketing: A case study of the e-tendering system in Karnataka. *India. Agricultural Economics Research Review, 31*(1), 53–64.

Press Information Bureau (2020). Thirty-eight new mandis integrated with eNAM: The pan-India electronic Agri-produce trading portal reaches a milestone of 1000 mandis across 18 States & 3 UTs. *Press Information Bureau. Ministry of Agriculture & Farmers Welfare*. https://pib.gov.in/PressReleasePage.aspx?PRID=1624083.

Rajalakshmi, N. (2017, November 23). The Hindu Business Line: Centre's eNAM initiative yet to break new ground. *Newstex Global Business Blogs, Chatham*. https://search.proquest.com/docview/1967540902?accountid=173681.

Ramesh, A., Banwet, D. K., & Shankar, R. (2012). Modelling the barriers to supply chain collaboration. *Journal of Modelling in Management, 5*(2), 176–193.

Ridings, C., Gefen, D., & Arinze, B. (2002). Some antecedents and effects of trust in virtual communities. *Journal of Strategic Information Systems, 11*, 271–295.

Sajwan, Raju (2020, April 8). COVID-19: Is an e-NAM portal capable of supporting farmers. *DowntoEarth*. https://www.downtoearth.org.in/news/agriculture/covid-19-is-e-nam-portal-capable-of-supporting-farmers–70294.

Salimi, M., Pourdarbani, R., & Nouri, B. (2020). Factors affecting the adoption of agricultural automation using davis's acceptance model (Case Study: Ardabil). *Acta Technologica Agriculture, 23*(1), 30–39.

Seifbarghy, M., & Esfandiari, N. (2013). Modelling and solving a multi-objective supplier quota allocation problem considering transaction costs. *Journal of Intelligent Manufacturing, 24*, 201–209.

Sharma, N. R., Sharma, S., & Sharma, D. (2020). Towards a mobile app technology-enabled sustainable agriculture in India. *Plant Archives, 20*(2), 3065–3071.

Shirzad, S. R., & Bell, D. (2013). A Systematic literature review of a flexible e-procurement marketplace. *Journal of Theoretical and Applied Electronic Commerce Research, 8*(2), 49–70.

Singh, K. M. (2017). *Public-private partnership for effective marketing of agricultural produce. Market-led agricultural extension concept and practices*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2254495.

Singh, R. K., Modgil, S., & Acharya, P. (2019). Assessment of supply chain flexibility using system dynamics modeling. *Global Journal of Flexible Systems Management*, *20*(Suppl 1), S39–S63.

Smart, A., & Harrison, A. (2003). Online reverse auctions and their role in buyer-supplier relationships. *Journal of Purchasing and Supply Management, 9*(5–6), 257–268.

Solaymani, S., Kiomars, S., & Esmaeil, A. Y. (2012). Adoption and use of e-commerce in SMEs: A case study. *Electronic Commerce Research, 12*, 249–263.

Soni, V., Singh, S. P., & Banwet, D. K. (2016). Systematic inquiry for energy security and sustainability (a case study approach for India using SAP-LAP framework). *Journal of Advances in Management Research, 13*(3), 352–367.

Subramaniam, C., & Shaw, M. J. (2002). A study of the value and impact of B2B E-commerce: The case of web-based procurement. *International Journal of Electronic Commerce, 6*(4), 19–40.

Subramanian, Arvind (2016, September 16). *Incentivising Pulses Production through Minimum Support Price (MSP) and Related Policies*. https://finmin.nic.in/sites/default/files/Pulses_report_16th_sep_2016.pdf

Suri, P. K. (2005). NICNET based agricultural marketing information network–A farmers centric portal on agricultural marketing in india and a step towards globalizing Indian agriculture. *Agricultural Marketing Journal, 45*(4), 2–11.

Suri, P.K. (2018) Towards an Effective Agricultural e-Trading System in India. In: Connell J., Agarwal R., Sushil, Dhir S. (eds) Global Value Chains, Flexibility, and Sustainability. *Flexible Systems Management*. Singapore: Springer.

Suri, P. K., & Sushil. (2011). Strategic Insights into e-governance planning and implementation: A study of select agricultural related projects. *International Journal of Electronic Governance, 4*(3), 259–272. https://doi.org/10.1504/IJEG.2011.043817

Sushil. (2005). Interpretive matrix: A tool to aid interpretation of management and social research. *Global Journal of Flexible Systems Management, 6*(2), 27–30.

Sushil. (2009). Interpretive ranking process. *Global Journal of Flexible Systems Management, 10*(4), 1–10.

Sushil. (2012). Multiple perspectives of flexible systems management. *Global Journal of Flexible Systems Management, 13*(1), 1–2.

Sushil. (2018). How to check the correctness of total interpretive structural models? *Big Data Analytics in Operations & Supply Chain Management. Annals of Operation Research*, *270*(1–2), 473–487.

Sushil. (2017a). Multi-criteria valuation of flexibility initiatives using integrated TISM-IRP with a big data framework. *Production Planning & Control, 28*(11–12), 999–1010. https://doi.org/10.1080/09537287.2017.1336794

Sushil. (2017b). Efficient interpretive ranking process incorporating implicit and transitive dominance relationships. *Annals of Operations Research, 283*(1–2), 1489–1516. https://doi.org/10.1007/s10479-017-2608-y

Sushil. (2020). Interpretive multi-criteria ranking of production systems with ordinal weights and transitive dominance relationships. *Annals of Operations Research, 290*(1–2), 677–695. https://doi.org/10.1007/s10479-018-2946-4

Tomar, G., Chauhan, G. S., & Panigrahi, P. K. (2016). Feasibility of m-governance in agriculture: Insights from a multimodal study in rural India. *Transforming Government: People, Process and Policy, 10*(3), 434–456.

Triantaphyllou, Evangelos (2000). Multi-criteria decision-making methods: a comparative study. *Applied Optimization* (pp. 32). Amsterdam: Kluwer Academic Publishers.

Tsai, Y. T., & Lasminar, R. G. (2021). Proactive and reactive flexibility: How does flexibility mediate the link between supply chain information integration and performance. *International Journal of Engineering Business Management, 13*, 1–12.

Tyndall, G. (1998). *Supercharging Supply Chains: New Ways to Increase Value Through Global Operational Excellence*. John Wiley & Sons.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478.

Venkatesh, V., Thong, J., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly, 36*(1), 157–178.

Warfield, J. N. (1974). Toward an interpretation of complex structural models. *IEEE Transactions on Systems, Man and Cybernetics, 4*(5), 405–417.

Wiyada, T., Vatcharaporn, E., & Donyaprueth, K. (2018). A framework to identify factors affecting the performance of third-party B2B e-marketplaces: A seller's perspective. *Electronic Markets, 28*, 129–147.

Yahya, S., & Kingsman, B. (1999). Vendor rating for an entrepreneur development programme: A case study using the analytic hierarchy process method. *Journal of the Operational Research Society, 50*(9), 916–930.

Zhang, Q., Cheng, L., & Boutaba, R. (2010). Cloud computing: State-of-the-art and research challenges. *Journal of Internet Services and Applications, 1*(1), 7–18.

Key Questions

1. How flexibility of a B2B e-Commerce project can be enhanced to support supply chain resilience in the context of agriculture commodities?
2. How could you implement the Interpretive Ranking Process methodology in a specific project?

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

Sanjay Chaudhary is an Associate Professor of MIS and Operations Management at the Delhi School of Business (DSB), Vivekananda Institute of Professional Studies. He is also a Research Scholar at Delhi Technological University. Before DSB, he has seventeen years of international work experience in the corporate sector. An IIT, SUNY, and AIT alumnus, he has expertise in ICT, Corporate, education, and Training. He is awarded a fellowship by the IETE and a Chartered Engineer designation by the Institution of Engineers, India. He is a member of the All India Management Association (AIMA). He has published several academic journal articles.

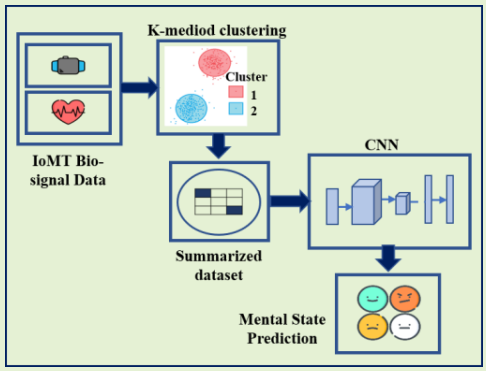
P. K. Suri serves as a Professor at the Delhi School of Management (DSM), Delhi Technological University. His areas of interest for teaching, research, and consultancy include E-Governance, IT Strategy and Management, Project Management and Business Analytics. He has served as Head (DSM) during 2013–2016. He is originally from the 1986 batch of the Indian Statistical Service. He has also served in consulting roles for the UN bodies, viz. WTO and FAO. He is a member of the Governing Council of the Global Institute of Flexible Systems Management (GIFT). He has published numerous journal articles.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2021.3095853, IEEE Sensors Journal

Resolving Data Overload and Latency Issues in Multivariate Time-Series IoMT Data for Mental Health Monitoring

Divya Gupta, MPS Bhatia, Akshi Kumar, *Member, IEEE*

***Abstract*— Pervasive healthcare services have evolved substantially in the recent years with IoMT rapidly changing the pace and scale of healthcare delivery. A promising application of IoMT is to fetch patterns of mental behaviour symptomatology based on bio-signals and transfer it to the corresponding hospital or psychologist for remote monitoring. But the data volume & performance, device diversity & interoperability, hacking & unauthorized use and acceptance & adoption barriers still restrain the practical and competent use of these devices. This research presents a plausible solution to surmount the data overload and processing latency in real-time sensory data collected through wearable devices for mental health monitoring. We propose a modified k-medoid data clustering technique based on time-frame restricted intra-cluster similarity calculations to obtain a summarized version of the original benchmark WESAD dataset for which the degree of information lost is minimum. A CNN is then trained on this summarized dataset for classification of mental state into the baseline, stress and amusement categories. The results show a significant reduction in the average execution time by 34% with a comparable accuracy to the original dataset, thus offering prompt real-time healthcare analytics.**



***Index Terms*— Clustering, CNN, IoMT, k-medoids, Mental State**

I. Introduction

ARTIFICAL intelligence (AI), Internet of Medical Things (IoMT) and big data analytics have revitalized the healthcare industry with an evident paradigm shift from the *'reactive healthcare'* to the *'proactive patient care'*. The applications range from the rudimentary chronic morbid disease management (diabetes, asthma, heart disease, etc.) to remote-assisted living (tele-health), wellness and preventive care (lifestyle assessment), remote intervention and improved drug management. Patients use medical devices, monitoring tools, wearables, and other sensors collectively referred to as the IoMT, that can transmit signals to other devices via the internet and cloud services [1, 2]. It is imperative to synthesize, process, analyze, visualize and integrate this massive amount of sensory data to generate value and insights for chronic disease management and patient care needs. Typically, the acquired data is transmitted over the network to the cloud for storage & analytics. This data received in near real-time allows doctors and caregivers to monitor an array of vitals, dynamically manage treatment plans, and conduct a consult or intervention over a webcam. Further, the data supports predictive analytics, allowing doctors to increase their accuracy of diagnoses by detecting emerging health patterns much faster. Thus, by leveraging the data that IoMT devices collect, AI is proving useful in enabling real-time remote measurement, analysis of patient data, enhancing effectiveness of care and decreasing the overall costs [3]. A promising application of IoMT is to fetch patterns of mental behaviour symptomatology based on bio-signals and transfer it to the corresponding hospital or psychologist for remote monitoring [4, 5]. The primary use of AI in mental healthcare includes:

- Early detection, flagging risks, and prediction
- AI-based chatbots to help patients 24/7

Mental health monitoring is more challenging than the physical health monitoring because firstly, human psychology varies dynamically; secondly, it is difficult to fetch the patterns of mental behavior; thirdly, behavioural symptomatology of mental illness may be visible at a later stage and lastly people with disturbed mental health are constantly in denial of the illness and have associated fears of cultural and social stigma. The use of smart algorithms for IoMT data analytics may provide psychiatrists and therapists with valuable insights which may help faster diagnosis and targeted treatment plans for mental well-being [6, 7].

According to a recent report by Deloitte, the IoMT market is expected to be worth $158.1 billion by 2022 [8], but a few key challenges are standing in the way of its continued growth. These include the network-driven, data-driven,

Manuscript submitted: January 27, 2021

Divya Gupta is pursuing her PhD from Netaji Subhas University of Technology, New Delhi, India (e-mail:divyagupta.cs@nsut.ac.in).

MPS Bhatia is a Professor with the Dept. of Computer Science & Engineering, Netaji Subhas University of Technology, New Delhi, India (e-mail: mpsbhatia@nsut.ac.in).

Akshi Kumar is with the Dept. of Computer Science & Engineering Department, Delhi Technological University, New Delhi, India (e-mail: akshikumar@dce.ac.in).

XXXX-XXXX © XXXX IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
1530-437X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:49:03 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2021.3095853, IEEE Sensors Journal

device-driven and user-drives issues as given in fig.1.

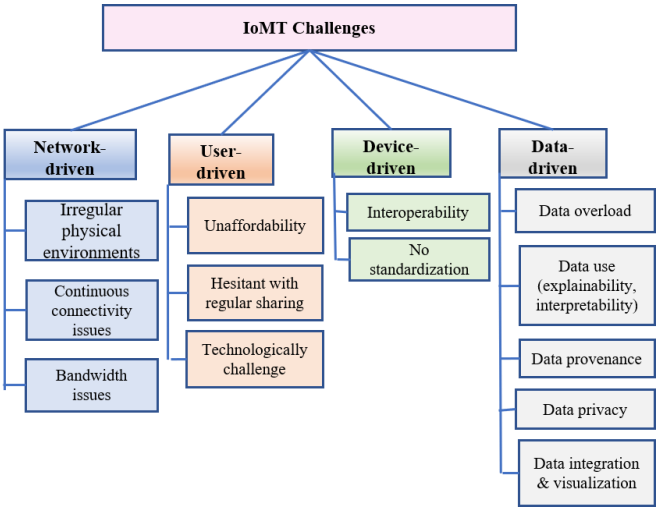


Fig.1. Challenges in IoMT adoption

- *Network-driven issues:* The large volume of data transmitted can strain the network resulting in bottlenecks [9]. Healthcare applications can't tolerate latency and require significant bandwidth for ingesting data. While integrating IoT into the healthcare industry, it is mandatory that the devices be movable and with a guaranteed continuous connectivity. But this is interrupted due to the irregular physical environments.
- *Data-driven issues:* IoMT devices record tons of data and utilize it to gain vital insights. However, the volume of data is so colossal that gaining insights from it is turning out very challenging for practitioners and ultimately affects the efficacy of decision-making. Moreover, this concern is rising as more devices are connected which record more and more data making data overload a primary issue. Also, while AI technologies are extraordinarily powerful, adoption of these algorithms in healthcare has been slow because doctors and regulators cannot verify the results making data use, data provenance, interpretability and explainability a rising concern. Further, healthcare data may contain sensitive attribute values which may lead to identify disclosure. Ensuring informational privacy (secrecy, confidentiality, data protection, or anonymity) of healthcare data is imperative for ensuring quality of care, enhanced autonomy, and preventing economic harm, embarrassment, and discrimination [10]. Lastly, although there are vast amounts of data in the healthcare sector, this data is often very fragmented and dispersed. By collecting information through sensors, integrating such information on a single platform, and visually presenting it through a smart healthcare command centre, healthcare staff can quickly acquire meaningful information that enables them to improve the quality and efficiency of healthcare services. Thus, data overload, lack of explainability, privacy, data integration & visualization are the key data-driven concerns.
- *Device-driven issues:* Data formats and communication protocols may be different for larger devices. This can significantly increase the time it takes for these devices to integrate with existing IoMT devices. Lack of standardization in terms of definitions of devices, data format and communication protocols affects data interchange and interoperability impeding the expediency and effectiveness of IoMT [2].
- *User-driven issues:* Unaffordability is one of the key user concerns where the cost of IoMT is higher for patients. Another major concern is that the patients may not be comfortable with regular sharing of their health information or may be technologically challenged.

At its core, the IoMT are equipped with state-of-the- art sensors that have the ability to record and transfer vital healthcare data thus enhancing the competence of healthcare delivery and ensuring improved patient outcomes [11]. But quite clearly, the data management systems must cater to the phenomenal volume, variety and velocity of data which may need huge storage and make analytical, process and retrieval operations cumbersome and time-consuming. Though the recent advancements in 5G wireless technology and edge computing enable connected devices to process data closer to where it is created, it may partially improve the response time [12]. Still the overwhelming amount of data can create havoc and the healthcare practitioners might miss on some important information. To resolve this problem, a plausible solution is to condense/ summarize the dataset such that it's still an informative version of the entire dataset [13]. This research uses the k-medoid clustering [14, 15] to obtain the condensed version of the original dataset for which the degree of information lost is minimum and then trains a convolution neural network (CNN) on the summarized data set for diagnostic classification. This would help resolve the data overload issue. Though k-medoid suffers from scalability issues, the intra-cluster similarity calculations have been restricted using a time-frame window mitigating the computational cost. The condensed data eventually decreases the processing latency and delivers prompt real-time analytics.

We consider the publicly available WESAD (Wearable Stress and Affect Detection) dataset [16], which contains bio-signal data of 15 subjects measured through two IoMT devices, namely RespiBAN and Empatica E4 for a period of 2 hours. The multivariate time-series data consists of 12 bio-signals: electrocardiogram (ECG), electromyogram (EMG), body temperature (TEMP), respiration (RESP), blood volume pressure (BVP), electrodermal activity (EDA) and three axis acceleration (ACC). ECG, EDA, and EMG are analog time-series data and to analyze continuous time-series data it is fragmented into finite intervals. For example, to identify the mental state of an individual, the time-series data in the benchmark WESAD dataset is broken at an interval of 1 second with the sliding window of 0.25 seconds, as the effect of a stimulus can occur over a short interval of 2 seconds. The conversion of analog data to discrete values makes a total of 63000000 instances just for a period of 2 hours in this dataset.

1530-437X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:49:03 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2021.3095853, IEEE Sensors Journal

This huge amount of data is not easy to analyze for any medical practitioner even for a few patients on a daily basis. By using k-medoid cluster analysis, summary can be generated as it partitions n-observations into k-clusters in which each observation belongs to the cluster with the closest medoid. The original WESAD data is initially transformed by converting the analog bio-signals into discrete values and further into numeric values. Clustering is then used to generate the centroid for each cluster and these centroids now characterize the summarized WESAD dataset. Finally, CNN is trained using this summarized WESAD data containing k-medoids, to predict the mental state of an individual into three categories as stress, amusement or baseline. Thus, the primary contributions of this research are:

i. A solution to surmount data overload and processing latency in real-time sensory data for mental health monitoring.
ii. To assess the efficiency of a deep learning model trained using the summarized version for mental state prediction.

Learning algorithms offer unprecedented insights into diagnostics, care processes, treatment variability, and patient outcomes [17, 18]. Pertinent studies to resolve various IoMT issues are available. Fadlullah et al. [19] identified that network and computational congestion problems may impact the real-time analytics of the healthcare data and proposed a deep learning based IoT edge analytics approach to support intelligent healthcare for residential users. Chen et al. [20] proffered a cognitive computation based smart healthcare system on the edge to resolve the multimodality, latency and resource issues. In 2020, Alfarraj and Tolba [21] introduced a responsive model for effectively handling IoMT data regardless of the time factor. In 2021, Goswami et al. [22] proposed a method for efficient power utilization for remote health monitoring.

Recent studies on mental health monitoring using WESAD dataset have reported the use of various machine learning and deep learning algorithms [23-26]. In 2021, Sharma et al. [27] proposed a model for computational intelligence on the edge. The authors reportedly summarized the dataset using an optimized fuzzy C-means clustering algorithm to finally train a deep hierarchical model for affective state detection in the WESAD data.

The next section outlines the details of the proposed model followed by the results and discussion in section 3. Finally, the conclusion and future work is given in the last section.

II. The Proposed Model for Fast Predictive Analytics

A CNN based deep learning model is put forward which facilitates predictive analytics of mental health to detect and monitor disturbed mental states. The model is trained using a summarized [26] version of the original IoMT dataset created using cluster analysis, which directly curbs the data overload issue and simultaneously reduces the processing time and consequent decision latency. Fig.2 depicts the architecture of the proposed model.

The proposed fast analytical model for detecting the mental state of an individual consists of the following two architectural components:

- K-medoid clustering technique for summarized dataset
- CNN trained using summarized dataset to detect the mental state

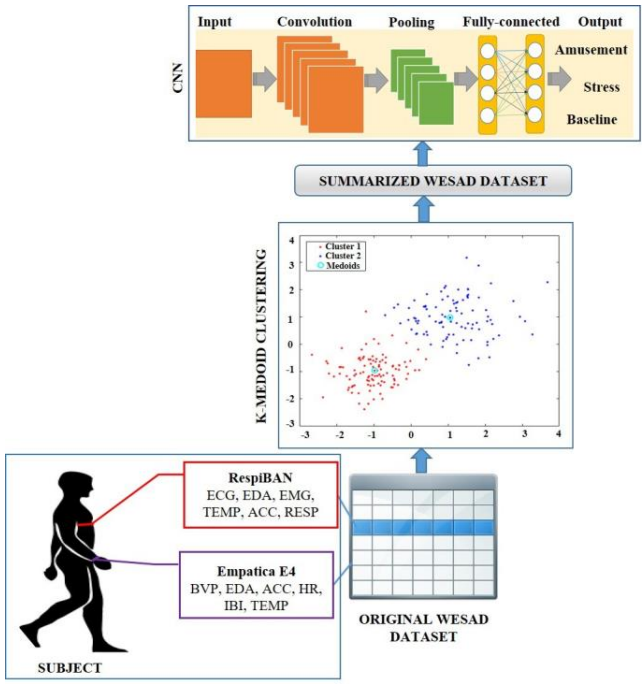


Fig.2. Architecture of proposed model

Clustering is counted amongst the key unsupervised learning techniques. Clustering of multivariate time-series data is specifically useful in exploratory data analysis and summary generation as it groups objects into several clusters by measuring the similarity between objects through distance and identifies the characteristics of each cluster [28]. Typically categorized as partition-based, density-based, grid-based hierarchical and model-based clustering techniques [29], each of these has its pros and cons while being used for clustering of time-series data. Though hierarchical clustering is a strong contender owing to its visualization capabilities and with no prerequisite of cluster initialization, it suffers due to its quadratic complexity and lack in scalability. Alternatively, the partition-based clustering techniques such as k-means and k-medoids (partitioning around medoid, PAM) are faster.

In this research, we use the k-medoids cluster analysis for data reduction. The 'K-Medoids Clustering' combines the k-Means and the medoid shift algorithms aiming to partition n-observations into k clusters in which each observation belongs to the cluster with the closest medoid. That is, instead of taking the mean value of the object in a cluster as a reference point, medoids can be used, which is the most centrally located object in a cluster. The medoid of a finite data set is a data point (one of the observations) from this set, whose average dissimilarity to all data points is minimal. It offers an improvement on the k-means clustering, in terms of execution time and non- sensitivity to outliers or noise. It is effective as it does not depend on the order in which data points are

1530-437X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:49:03 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2021.3095853, IEEE Sensors Journal

examined. Moreover, the cluster center is part of the dataset, unlike k-means where the cluster center is gravity based.

Algorithm 1: K-medoid

Input: WESAD dataset-D, number of clusters-k
Output: Medoids M
1: Arbitrary choose k object as initial medoids
2: Assign each remaining object to nearest medoids
3: Calculate objective function: *the sum of dissimilarities of all objects to their nearest medoid*
4: Randomly select a non-medoid object, O_{ramdom}
5: If objective function improved.
6: Swapping O and O_{ramdom}
7: Compute Total Cost of swapping
8: Repeat (3-7) Until no change

The second component of the proposed model uses a CNN to predict the mental state of an individual into three categories, namely stress, amusement and baseline for real-time monitoring. CNN is a neural network with a sequence of convolutional layers (often with a pooling step) and then followed by one or more fully connected layers. A convolutional layer has a number of filters that does convolutional operation [31]. CNN is faster to train as convolutions can be done in parallel, thus utilizing full advantage of GPU parallelism. The functionalities of each layer in CNN is as follows:

- *Convolution layer:* computes the output of neurons that are connected to local regions in the input, each computing a dot product between their weights and a small region they are connected to in the input volume. The convolution layer's parameters consist of a set of learnable filters
- *RELU layer:* applies an element-wise activation function, such as the max (0, x) thresholding at zero. This leaves the size of the volume unchanged.
- *Pooling layer:* performs a down sampling operation along the spatial dimensions (width, height). Its function is to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network. The pooling layer often uses the Max operation to down-sample the previous layers feature map.
- *Fully-connected (FC) layer:* Fully connected layers are the normal flat feed-forward neural network layers and may have a non-linear activation function or a Softmax activation in order to predict classes. The FC layer basically computes the class scores, resulting in volume of size [1x1xN], where each of the N numbers correspond to a class score, such as among the N categories.

Both the clustering operation and the predictive model have been trained on a system with a 2.7 GHz Intel core i5 processor with 16GB RAM over the WESAD dataset. The WESAD data was collected by Schmidt et al. in 2018 [16] at the University of Siegen, Germany for analysing the affective state of 17 individuals by monitoring their physiological changes through various biomarkers collected through Empatica E4 (wrist-worn) and RespiBAN (chest-worn) devices in a controlled lab environment for a period of 2 hours. The final data for only 15 subjects was made publically available with entries of 2 subjects considered as invalid. WESAD tracks the bio-signals like ECG, EDA, and EMG for determining the stress levels which are analog signals and for training a deep learning model this time-series analog data is discretized by breaking the data at finite intervals. Each subject after discretization has 40 lakh instances approximately, as the time-series data is broken into a finite interval of 1 second with 0.25 second as the sliding window to monitor the change [32].

In general, clustering is used to identify different types of instances available, i.e. to categorize the instances using unsupervised learning techniques. Clustering groups the entities of similar nature together, but in time-series data especially in healthcare, even with similar instances, data cannot be grouped together, as the context in terms of pre and post condition have an effect on the data. Therefore, to resolve this constraint the K-medoid clustering algorithm has been modified to accomplish the cluster generation for a particular time frame only. In a basic K-medoid clustering algorithm, K data points are chosen at the beginning, and then clusters are formed by analyzing the similarity with rest of the data with these k- medoids and after calculating the similarity with each medoid, the point is assigned to the nearest medoid. Once all the data has been categorized into k-partitions, the medoid (centralized) point of each partition is varied from the same partition iteratively until all the points have been chosen once as the medoid point, and then the optimized clusters are provided as the output. There are two issues with this basic approach:

- In each iteration, the similarity is calculated again with each instance in the dataset, for large datasets like WESAD, calculating similarity with 40 lakh entries for a single subject in each iteration is a tedious and time-consuming task.
- A person can have identical bio-signals at different time instance, for example, a person sitting on a chair is in baseline state, on receiving a work assignment, and while performing the task his physiological state changes, but once the task is over and the person relaxes again, he again goes into the baseline state, that is the person has identical bio-signals pre and post the task. In basic k-medoid clustering technique, the states pre and post the task will be put under the same cluster based on their similarity without taking the time-constraint into account.

Therefore, the modified K-medoid clustering algorithm has been used, in which rather than calculating the similarity function with all the data points, in each iteration, the similarity of the medoid is calculated with only 80 data points before the medoid and 80 points after the medoid in the continuous data, ensuring that no two-time varied instances should be clubbed together. Thus, to resolve the data overload issue in real-time IoMT data processing, we reduce the size of the data by performing clustering on this time-series data and the centroid of a cluster is used for analysis rather than every instance of the cluster.

As per the observation made by Siirtola in 2019 [23], the 'affect' perceived depends upon the time window taken. The use of these time segments (windows) capture more discriminative information for the same input signal. Affective states are typically short and need a time window of seconds.

1530-437X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:49:03 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2021.3095853, IEEE Sensors Journal

But analysing every single second of data may not be effective and therefore we consider the window size of 3 seconds to detect any physiological change in the body and maintain the uniformity of our experiment. The base hypothesis for identifying the optimal clusters for data reduction is empirically analysed for performance accuracy using varying time window. Based on this analysis, to initialize the model, 2400 clusters for WESAD (2 hours => 7200 seconds) are used, and 200 extra cluster points are chosen for any outlier or frequent change in the body. As in real-time analysis the clustering process will be performed for streaming data, waiting for upcoming data can be ruinous in certain scenarios and therefore the optimized segment of 3 second was chosen for time coherence. This 3 second time segment cannot be generalized for every health condition prediction model.

III. Results & Discussion

Various bio-signals measure different physiological changes in the body and each indicates some sort of psychological or emotional affect a person is experiencing. That is, to understand the current mental state of a person each bio-signal plays an important role. The bio-signals in the WESAD dataset are:

- *RespiBAN:* Electrocardiogram (ECG), Electromyogram (EMG), Body Temperature (TEMP), Respiration (RESP), Blood Volume Pressure (BVP), Electrodermal Activity (EDA), Three Axis Acceleration (ACC)
- *Empatica E4:* Body Temperature (TEMP), Blood Volume Pressure (BVP), Electrodermal Activity (EDA), Three Axis Acceleration (ACC)

As observed from the work carried by different researchers on WESAD, each person acts differently even under similar circumstances and their bodies show different physiological changes for stress, amusement, depression and for neutral state. Therefore, the model should be accustomed as per the individual. To train the model for each individual, the prediction model was implemented separately on each individual. Each individual has around 40, 00000 instances for two hours with 12 different bio-signals. The data was initially summarized by clustering using the modified K-medoid approach and after generating the clusters, the mediod point of each cluster was chosen as a representative for each cluster. In K-medoid the number of clusters and the initial medoid point of the cluster has to be provided. In execution, k was provided as 2600, first medoid was initialized with the 80th instance, afterwards every 120th instance was chosen as the next medoid point. During execution, the similarity of each medoid point was calculated only with 80 before instances and 80 post instances, a total of 200 iterations were performed, by setting the terminating condition, to avoid falling in the local optima. Table 1 and table 2 depict the original WESAD and the summarized WESAD data of subject 5.

The reduced dataset was then split into a 70:30 ratio for training and testing the deep-learning model, with 4-cross fold validation. Each subject was trained and tested individually, before and after summarizing the data. To analyze the effectiveness of the proposed model, the model is evaluated using accuracy and execution time, to understand the reduction in decision-latency time. Fig.3 and Fig.4 show the comparison of pre- and post-summarization on predicting the mental state of a person using CNN in terms of accuracy and execution time (in seconds) respectively.

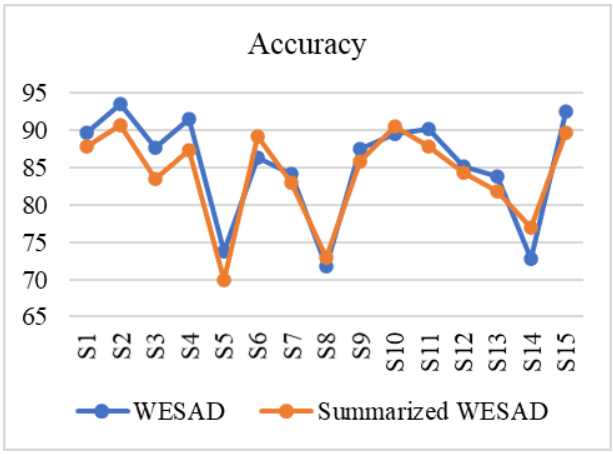


Fig. 3. Model accuracy using original and summarized WESAD

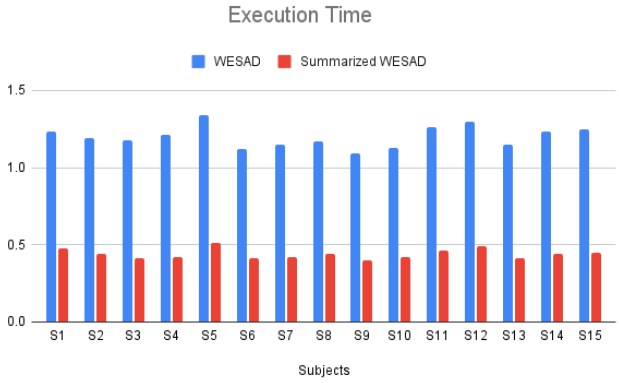


Fig. 4. Model accuracy using original and summarized WESAD

To handle the data overload issue and to fasten the processing time, the features (bio-signals) can also be reduced. As experimented by Schmidt et al. in 2018 [16] & Siirtola in 2019 [23], by clubbing different bio-signals for WESAD it was observed that a subset of bio-signals can predict the mental state with close accuracy as to considering all the bio-signals. To test the hypothesis, rather than sub-selecting the bio-signals to predict the mental state on WESAD data, we experimented by taking the bio-signals from each device individually. 4 bio-signals measured by both the devices are common and training a deep neural network on the identical result generated by two different devices is a waste of the processing power and increases the computational power.

Fig. 5 and fig. 6 shows the accuracy and execution time of the model on complete WESAD data, on only Empatica E4 data of each subject, and on only RespiBAN data of each subject respectively. It was observed that Empatica E4 has predicted the state better in comparison to RespiBAN with 81.77 and 80.57 accuracy respectively, but both have performed poorly in comparison to complete WESAD data which resulted in 85.37 accuracy.

1530-437X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:49:03 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2021.3095853, IEEE Sensors Journal

TABLE 1.
ORIGINAL WESAD DATA OF SUBJECT 5

	Empatica E4								RespiBAN							
Time	**BVP**	**EDA**	**ACC-A**	**ACC-B**	**ACC-C**	**HR**	**IBI**	**TEMP**	**ECG**	**EDA**	**EMG**	**TEMP**	**ACC-X**	**ACC-Y**	**ACC-Z**	**RESP**
0	0	0	0	5	63	60	0.85941	27.61	31600	9942	32567	26.533	37303	32479	31367	34015
0	0	0.40031	0	5	63	60.5	0.85941	27.61	31636	9923	32739	26.542	37291	32495	31377	34015
1	0	0.47577	0	5	63	60	0.85941	27.61	31709	9954	32850	26.551	37287	32497	31364	34015
1	0	0.486	0	5	63	59.75	0.96879	27.61	31763	9925	32672	26.549	37279	32487	31342	34025
1	0	0.48856	0	5	63	59.8	0.96879	27.59	31810	9955	32867	26.539	37273	32489	31373	34049
2	0	0.48984	-1	5	63	60.67	0.90629	27.59	31881	9971	32661	26.534	37282	32485	31367	34039
2	0	0.48984	-1	5	63	61.43	0.95317	27.59	31982	9927	32693	26.555	37269	32488	31365	34036
2	0	0.48856	0	5	63	62.12	0.85941	27.59	32140	9955	32783	26.565	37271	32487	31356	34039
3	0	0.486	0	5	63	62.67	0.75003	27.59	32267	9929	32828	26.578	37268	32490	31357	34055
3	0	0.48344	0	5	63	63.1	0.84379	27.59	32415	9937	32553	26.549	37267	32491	31355	33822
3	0	0.48856	0	5	63	63.45	0.82816	27.59	32439	10005	32757	26.549	37277	32485	31340	34049
4	0.01	0.48984	0	5	63	63.75	0.76566	27.59	32401	9939	32788	26.539	37292	32475	31335	34069

TABLE 2.
SUMMARIZED ORIGINAL WESAD DATA OF SUBJECT 5

		Empatica E4								RespiBAN							
Cluster	**Time**	**BVP**	**EDA**	**ACC-A**	**ACC-B**	**ACC-C**	**HR**	**IBI**	**TEMP**	**ECG**	**EDA**	**EMG**	**TEMP**	**ACC-X**	**ACC-Y**	**ACC-Z**	**RESP**
C1	1	0	0.39	-0.167	5	63	60.1	0.90369	27.6	31733	9945	32726	26.541	37286	32489	31365	34026
C2	4	0.002	0.488	-0.167	5	63	62.8	0.83337	27.59	32274	9949	32734	26.556	37274	32486	31351	34012
C3	7	-0.02	0.491	-0.667	4.8333	62.833	64.8	0.83077	27.57	31955	9942	32775	26.561	37279	32478	31343	34049
C4	10	0.875	0.49	0.1667	5.5	62.833	65.4	0.90629	27.57	32053	9959	32725	26.549	37270	32485	31357	34021

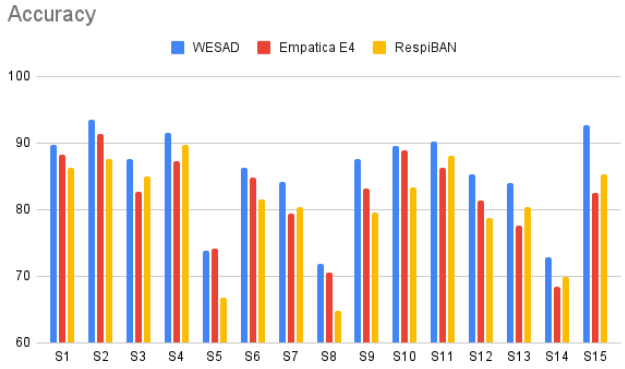


Fig. 5. Accuracy using WESAD vs. individual wearable data

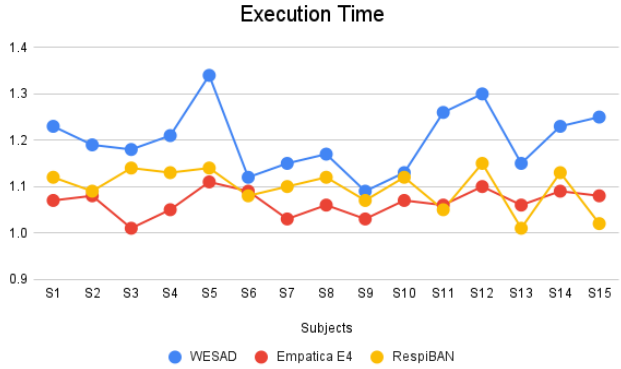


Fig. 6. Execution time using WESAD vs. individual wearable data

Although, the model can perform better by choosing a subset of combined features of the two devices. To evaluate the efficiency of the proposed data reduction strategy, the improved k-medoid clustering algorithm has been used and evaluated on Empatica E4 and RespiBAN data separately too. Table 3, summarizes the accuracy and execution time for all the 3 datasets for pre- and post- summarization.

Fig. 7 and fig. 8 represents graphically the comparison of accuracy and execution time for all the scenarios, that is, original WESAD, summarized WESAD, Empatica E4, summarized Empatica E4, RespiBAN and summarized RespiBAN.

TABLE 3.
SUBJECT-WISE ACCURACY AND EXECUTION TIME

Subjects	**WESAD**		**Summarized WESAD**		**Empatica E4**		**Summarized Empatica E4**		**RespiBAN**		**Summarized RespiBAN**	
	Accuracy	**Time**	**Accuracy**	**Time**	**Accuracy**	**Time**	**Accuracy**	**Time**	**Accuracy**	**Time**	**Accuracy**	**Time**
S1	89.71	1.23	87.94	0.48	88.19	1.07	87.21	0.39	86.31	1.12	84.97	0.38
S2	93.54	1.19	90.71	0.44	91.37	1.08	90.29	0.36	87.59	1.09	85.49	0.34
S3	87.67	1.18	83.54	0.41	82.68	1.01	83.21	0.33	84.96	1.14	81.65	0.38
S4	91.49	1.21	87.39	0.42	87.19	1.05	83.47	0.35	89.73	1.13	88.71	0.36
S5	73.87	1.34	69.89	0.51	74.21	1.11	69.34	0.38	66.81	1.14	65.32	0.39

1530-437X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:49:03 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2021.3095853, IEEE Sensors Journal

S6	86.35	1.12	89.13	0.41	84.73	1.09	85.92	0.35	81.59	1.08	79.08	0.34
S7	84.19	1.15	83.02	0.42	79.38	1.03	76.75	0.38	80.36	1.1	78.41	0.35
S8	71.83	1.17	72.91	0.44	70.64	1.06	68.78	0.32	64.81	1.12	62.84	0.37
S9	87.56	1.09	85.83	0.4	83.15	1.03	79.95	0.35	79.62	1.07	80.03	0.36
S10	89.49	1.13	90.56	0.42	88.91	1.07	87.27	0.38	83.37	1.12	82.78	0.35
S11	90.28	1.26	87.91	0.46	86.27	1.06	84.89	0.37	88.06	1.05	87.46	0.31
S12	85.24	1.3	84.28	0.49	81.34	1.1	81.08	0.4	78.81	1.15	78.07	0.38
S13	83.91	1.15	81.84	0.41	77.63	1.06	76.92	0.39	80.45	1.01	76.7	0.34
S14	72.77	1.23	76.91	0.44	68.46	1.09	65.87	0.41	69.98	1.13	67.16	0.37
S15	92.62	1.25	89.79	0.45	82.53	1.08	80.79	0.39	85.23	1.02	84.23	0.34
Average	85.36	1.2	84.11	0.44	81.77	1.06	80.11	0.37	80.51	1.09	78.86	0.35

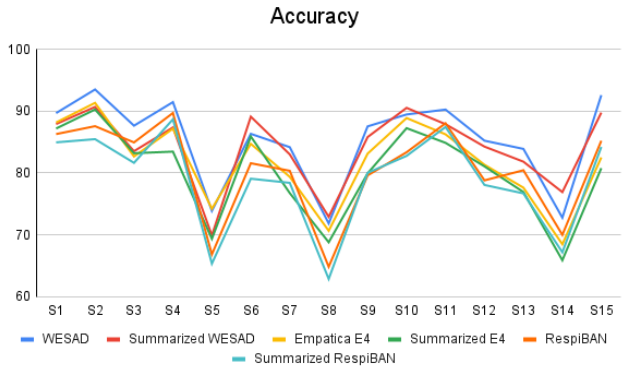


Fig. 7. Accuracy comparison for all scenarios

Fig. 8. Execution time comparison for all scenarios

As evident from the results, the summarization technique used has reduced the average execution time by 34%, whereas the accuracy has declined only by 0.98% for all the three scenarios. In real-time healthcare, this can reduce the decision latency time for critical medical services and provide an accurate detection of the current mental health condition of an individual. As a solution to mitigate the risk associated with the minor reduction in accuracy, the model can be trained using original (non-summarized) data and while analyzing the health condition in real-time to avoid the latency in decision making, the k-medoid clustering can be used for summarizing the continuous incoming data. Also, as the model is currently trained on only 2 hours of bio-signals data, to achieve better accuracy data can be collected for a longer period of time. Further, as smart healthcare services in real-time need high precision, as inefficiency can result in a loss of life, therefore, to increase the effectiveness/accuracy of prediction of health analysis in real-time, each feature can be trained explicitly with computationally less expensive and well-learned models.

IV. Conclusion

This work substantiates that integrating stakeholders within the pervasive healthcare paradigm which include practitioners, patients, processes, and connected medical devices along with the use of disruptive technologies like AI, blockchain, and robotics can improve the diagnostic methods and increase the quality of patient care. Although IoMT devices facilitate disease prevention, fitness promotion, and remote intervention in emergency situations, the overwhelming amount of data make timelines and decision-making difficult for medical experts. In this research, a CNN model was trained using a summarized version of the original IoMT based mental health WESAD dataset. We used the k-medoid technique for cluster analysis to create the summarized version of the WESAD dataset, but as k-medoid suffers from scalability issues we restricted the intra-cluster similarity calculations using a time-frame window and mitigated the computational cost. Evidently the data clustering technique reduced the average execution time.

As possible future direction, we intend to work on other data-driven issues pertaining to IoMT, especially as the devices and data are highly vulnerable to cybersecurity risks. Moreover, healthcare diagnostic models are critical for human life and need proper justification and ‘context’ to the diagnosis. Explainable AI can be used for data interpretability and improving human understanding, for determining the justifiability of the decision made by the machine, introducing trust and reducing bias in healthcare decision making.

References

[1] G. J. Joyia, R. M. Liaqat, A. Farooq and S. Rehman, “Internet of Medical Things (IOMT): applications, benefits and future challenges in healthcare domain”. *J Commun*, vol. 12 no.4, pp. 240-247, 2017. DOI: 10.12720/jcm.12.4.240-247

[2] K.T. Kadhim, A.M. Alsahlany, S.M. Wadi and H.T. Kadhum, “An overview of patient’s health status monitoring system based on Internet of Things (IoT). *Wireless Personal Communications*, 114, pp.2235-2262, 2020. DOI: 10.1007/s11277-020-07474-0.

[3] L. Greco, G. Percannella, P. Ritrovato, F. Tortorella and M. Vento, “Trends in IoT based solutions for health care: Moving AI to the edge”. *Pattern recognition letters*, 135, pp.346-353, 2020. DOI: 10.1016/j.patrec.2020.05.016.

[4] S. Graham, C. Depp, E. E. Lee, C. Nebeker, X. Tu, H.C. Kim and D.V. Jeste, “Artificial intelligence for mental health and mental illnesses: an overview”. *Current psychiatry reports*, vol.21, no.11, pp.1-18, 2019. DOI: 10.1007/s11920-019-1094-0.

[5] L. Rachakonda, S.P. Mohanty, and E. Kougianos, “iLog: an intelligent device for automatic food intake monitoring and stress detection in the IoMT”. *IEEE Transactions on Consumer Electronics*, vol.66, no.2, pp.115-124, 2020. DOI: 10.1109/TCE.2020.2976006.

[6] D. Gupta, MPS Bhatia, and A. Kumar. “Real-Time Mental Health Analytics Using IoMT and Social Media Datasets: Research and

1530-437X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:49:03 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2021.3095853, IEEE Sensors Journal

Challenges" (May 10, 2021). Available at SSRN: https://ssrn.com/abstract=3842818 or http://dx.doi.org/10.2139/ssrn.3842818

[7] S. Yang, B. Gao, L. Jiang, J . Jin, Z. Gao, X. Ma and W.L.Woo, "IoT structured long-term wearable social sensing for mental wellbeing". *IEEE Internet of Things Journal*, vol. 6, no. 2, pp.3652-3662, 2018. DOI: 10.1109/JIOT.2018.2889966.

[8] N. Mani, A. Singh and S.L. Nimmagadda, "An IoT guided healthcare monitoring system for managing real-time notifications by fog computing services". *Procedia Computer Science*, 167, pp. 850-859, 2019. DOI: 10.1016/j.procs.2020.03.424.

[9] S. C. Mukhopadhyay, "Wearable sensors for human activity monitoring: A review". *IEEE sensors journal*, vol. 15, no.3, pp.1321-1330, 2014. DOI: 10.1109/JSEN.2014.2370945.

[10] L. Cilliers, "Wearable devices in healthcare: Privacy and information security issues". *Health information management journal*, vol. 49, no. 2-3, pp.150-156, 2020. DOI: 10.1177/1833358319851684.

[11] F. Al-Turjman, M.H. Nawaz and U.D. Ulusar, "Intelligence in the Internet of Medical Things era: A systematic review of current and future trends". *Computer Communications*, 150, pp. 644-660, 2020. DOI: 10.1016/j.comcom.2019.12.030.

[12] R Ranjan and A. Sharma, "Voice-Controlled IoT Devices Framework for Smart Home". *In Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019*), pp. 57-67, Springer, Singapore, 2020. DOI: 10.1007/978-981-15-3369-3_5.

[13] D. Happ and S. Bayhan, S, "On the impact of clustering for IoT analytics and message broker placement across cloud and edge." In *Proceedings of the Third ACM International Workshop on Edge Systems, Analytics and Networking*, pp. 43-48, 2020.DOI: 10.1145/3378679.3394538.

[14] L. Kaufman and P J. Rousseeuw. "Clustering by means of medoids" *Statistical Data Analysis based on the L1 Norm and Related Methods*, pp. 405–416, 1987.

[15] L. Kaufman and P. J. Rousseeuw. "Partitioning around medoids (program pam)". *Finding groups in data: an introduction to cluster analysis*, pp. 68–125, 1990.

[16] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger and K.V. Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection". In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, pp. 400-408, 2018. DOI: 10.1145/3242969.3242985.

[17] A. Kumar, "Using cognition to resolve duplicacy issues in socially connected healthcare for smart cities". *Computer Communications*, vol. 152, pp.272-281, 2020. DOI: 10.1016/j.comcom.2020.01.041.

[18] A. Kumar, A. Sharma and A. Arora, "Anxious depression prediction in real-time social data". In *International Conference on Advances in Engineering Science Management & Technology (ICAESMT)-2019*, Uttaranchal University, Dehradun, India, 2019. DOI: 10.2139/ssrn.3383359.

[19] Z. M. Fadlullah, A. K. Pathan and H. Gacanin, "On Delay-Sensitive Healthcare Data Analytics at the Network Edge Based on Deep Learning," *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)*, Limassol, 2018, pp. 388-393. DOI: 10.1109/IWCMC.2018.8450475.

[20] M.Chen, W. Li, Y. Hao, Y. Qian and I. Humar, "Edge cognitive computing based smart healthcare system". *Future Generation Computer Systems*, vol. *86*, pp.403-411, 2018. DOI: 10.1016/j.future.2018.03.054.

[21] O. Alfarraj,and A. Tolba, "Unsynchronized wearable sensor data analytics model for improving the performance of smart healthcare systems". *Journal of Ambient Intelligence and Humanized Computing*, 1-12, 2020. DOI: 10.1007/s12652-020-02576-w.

[22] P., Goswami, A., Mukherjee, B., Sarkar, & L., Yang. Multi-agent-based smart power management for remote health monitoring. *Neural Computing and Applications*, 1-10, 2021. DOI: 10.1007/s00521-021-06040-4.

[23] P. Siirtola, "Continuous stress detection using the sensors of commercial smartwatch". *In Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, pp. 1198-1201, 2019. DOI: 10.1145/3341162.3344831.

[24] D. Bajpai and L. He, "Evaluating KNN Performance on WESAD Dataset". *In 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)*, pp. 60-62, IEEE, 2020. DOI: 10.1109/CICN49253.2020.9242568.

[25] J. Lin, S. Pan, C.S. Lee and S. Oviatt. "An explainable deep fusion network for affect recognition using physiological signals". *In Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 2069-2072, 2019. DOI: 10.1145/3357384.3358160.

[26] F.I. Indikawati and S. Winiarti. "Stress detection from multimodal wearable sensor data". *In IOP Conference Series: Materials Science and Engineering*, Vol. 771, No. 1, p. 012028. IOP Publishing, 2020. DOI: 10.1088/1757-899X/771/1/012028.

[27] A. Kumar, K. Sharma, A. Sharma. "Genetically optimized Fuzzy C-means data clustering of IoMT-based biomarkers for fast affective state recognition in intelligent edge analytics." *Applied Soft Computing*, 107525, 2021. DOI: 10.1016/j.asoc.2021.107525.

[28] A. Kumar & A. Sharma. "Systematic literature review of fuzzy logic based text summarization". *Iranian Journal of Fuzzy Systems*, *16(5)*, 45-59, 2019. DOI: 10.22111/IJFS.2019.4906.

[29] R. Xu and D. Wunsch, Clustering (Vol. 10). John Wiley & Sons, 2008.

[30] A. Ghosal, A. Nandy, A.K. Das, S. Goswami and M. Panday, "A short review on different clustering techniques and their applications". Emerging technology in modelling and graphics, pp.69-83, 2020. DOI: 10.1007/978-981-13-7403-6_9.

[31] A. Kumar and A. Jaiswal. "A Deep Swarm-Optimized Model for leveraging Industrial Data Analytics in Cognitive Manufacturing". *IEEE Transactions on Industrial Informatics*, vol. 17, no.4, pp.2938-2946, 2020. DOI: 10.1109/TII.2020.3005532.

[32] A. Kumar, K. Sharma, A. Sharma. "Hierarchical deep neural network for mental stress state detection using IoT based biomarkers". *Pattern Recognition Letters*, *145*, 81-87, 2021. DOI: 10.1016/j.patrec.2021.01.030.

Divya Gupta is pursuing her Ph.D from the Department of Computer Science & Engineering at Netaji Subhas University of Technology, New Delhi, India. She received her post graduate degree in software engineering from Delhi Technological University, Delhi, India in 2018. Her research interests are in the area of artificial intelligence, pervasive healthcare, data science & social media analytics.

MPS Bhatia received his Ph.D. degree in computer engineering from the Faculty of Technology at the University of Delhi, New Delhi, India. He is currently a full professor with the department of computer science and engineering, Netaji Subhas University of Technology, New Delhi, India. His research interests include the area of data science, pervasive healthcare and cybersecurity.

Akshi Kumar (Member, IEEE) is currently an assistant professor with the department of computer science and engineering, Delhi Technological University, Delhi, India. Her research interests are in the area of affective computing, data science, social media analytics, natural language understanding and pervasive healthcare.

1530-437X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:49:03 UTC from IEEE Xplore. Restrictions apply.

Arabian Journal for Science and Engineering
https://doi.org/10.1007/s13369-021-05846-7

REVIEW--MECHANICAL ENGINEERING

Review on Spray-Assisted Solar Desalination: Concept, Performance and Modeling

Anil Kumar[1,2] · **Ravi Kant**[1] · **Samsher**[3]

Received: 27 March 2021 / Accepted: 9 June 2021
© King Fahd University of Petroleum & Minerals 2021

Abstract

This article presents a review on performance, thermal modeling and numerical modeling of spray-assisted solar desalination systems (SLTD). Spray assisted has the benefits of higher rates of heat and mass transfer, the ability to work at moderately low-temperature differences, ease of system design, less scaling and fouling issues, and lower initial cost than conventional thermal desalination. Thermal models are handy tools to expect the performance of well-designed SLTD before its fabrication and saves cost and time. Software applications play a crucial role in developing and analyzing the mathematical model and assessing the performance of SLTD. CFD simulation is useful in the analysis as it can show the precise distribution of temperature inside the evaporation tower. It is used to model and analyze the evaporation process of brine spray and a powerful tool for guiding the selection of operating conditions. Highest productivity obtained as 9 L/m^2 per day, and the maximum daily efficiency was 87%. Maximum value of the performance ratio reached 1.42 in the double stage constant heat source spray assisted desalination system. Steam jet ejector contributes over 40% of the total energy degradation. Rate of production and performance ratio increases with the increase in top brine temperature. Productivity of SLTD was more compared to conventional distillation systems due to higher rate of heat transfer and a higher rate of evaporation.

Keywords Solar · Spray · Desalination · Performance · Modeling

List of Symbols

A	Area, m^2
BPE	Boiling point elevation
C	Specific heat, kJ/kg
C_p	Specific heat, kJ/kg
C_{ps}	Specific heat of seawater (J/kgK)
C_{pv}	Specific heat of vapor (J/kgK)
C_{pw}	Specific heat of water (J/kgK)
D	Diameter (m)
$\dot{D}$	Production rate, (kg/s)
$\dot{E}$	Exergy destruction rate, (kw)
g	Gravitational acceleration constant (m/sec^2)
h	Specific enthalpy, (kJ/kg), enthalpy (J/kg), heat transfer coefficient(W/m^{2}K)
h_{fg}	Latent heat of vaporization (J/kg)
H	Solar radiation, (W/m^2)
I	Solar irradiance (W/m^2)
k	Thermal conductivity (W/mK)
L	Latent heat of water, J/kg
m	Flow rate, l/h
M	Productivity (l/h), mass
$\dot{m}$	Mass flow rate, (kg/s)
$\dot{M}$	Mass flow rate, (kg/s)
$(\mathrm{MC})_d$	Thermal capacity
Nu	Nusselt number, dimensionless
P	Pressure (Pa), pumping power (W)
p	Pressure (Pa)
PCF	Motive steam pressure correction factor
P_p	Pump power (W)
Pr	Prandtl number, dimensionless
PR	Production ratio, performance ratio
Q	Heat flux (kW), heat transfer rate between condenser and cooling seawater (W)
R	Universal gas constant (J/k-mol)
r	Recycling ratio, dimensionless
Re	Reynolds number, dimensionless
s	Specific entropy (kJ/kg-K)
Sc	Schmidt number, dimensionless
Sh	Sherwood number, dimensionless

✉ Anil Kumar
anilkumar76@dtu.ac.in

[1] Department of Mechanical Engineering, Delhi Technological University, Delhi 110 042, India

[2] Centre for Energy and Environment, Delhi Technological University, Delhi 110 042, India

[3] Harcourt Butler Technical University, Kanpur 208002, India

Published online: 10 July 2021

T	Temperature, °C
t	Time, min
TCF	Entrained vapor temperature correction factor
Δt	Differential time element
U	Overall heat transfer coefficient (W/m^{2}K)
w	Entrainment ratio
w	Power, W
X	Salinity (g/kg

Superscripts

$t, t+\Delta t$	Value of the parameter at this time instant

Subscripts

0	Ambient
a	Ambient
amb	Ambient
Av	Average
ave	Average
b	Brine
c	Condenser, condensation
ci	Cold inlet
c, in	Condenser inlet
c, out	Condenser outlet
co	Cold outlet
cond	Condenser
cs	Condenser surface
cw	Cooling water
d	Distillate, destruction, daily
dist	Distillate
e	Evaporator, entrainment, evaporation
ev	Evaporator water column
evap	Evaporator
f	Feed
H	Heating
h	Heating, hourly
hi	Hot inlet
ho	Hot outlet
i	iTh stage, inner
in	Inlet
l	Liquid
loss	Temperature loss in demister, heat loss
o	Outlet, outer
p	Positive steam
ps	Distillate column
R	Heat recovery
s	Surface
S	Solar water collector
sat	Saturation
sc	Solar collector
sc, in	Solar collector inlet
sc, out	Solar collector outlet
sl	Solar collector
st	Storage Tank
sw	Seawater
u	Useful
v	Vapor
v, eq	Vapour at equilibrium
vs	Vapor space
w	Water
II	Second-law

Greek Letters

η	Efficiency
θ	Dimensionless temperature difference
Δ	Difference
μ	Dynamic viscosity (pa s)
v	Kinematic viscosity (m^2/s)
ρ	Density (kg/m^3

Abbreviations

CC	Cooling coil
CV	Control valve
EC	Condensation tower
ET	Evaporation tower
FM	Flow meter
FWR	Fresh water reservoir
GOR	Gain output ratio
HDH	Humidification-dehumidification
SS	Solar still
SLTD	Spray assisted low-temperature desalination
HEX	Heat exchanger
MED	Multiple effect distillation
MSF	Multi-stage flash
PR	Productivity rate
SWC	Solar water collector
TC	Thermocouples
TDS	Total dissolved solids
TES	Thermal energy Storage

1 Introduction

Water is essential to survive human life on earth. It is abundantly accessible on earth but under human use, significantly less of its accessibility. Most prominent problem at the moment is fresh and drinking water. Approximately 71% of the earth is surrounded by water, 96.5% of which is seawater, and the rest is found in rivers and lakes, local ponds, ice caps and glaciers, soils and aquifers, etc. Only less than 1% is worth of fresh human water [1]. That small amount of fresh water has been the primary water source to fulfill the need for agriculture and mechanical activities. Freshwater is not as safe as suggested by the national standard. It includes harmful micro-organisms and infections that cause various water-borne diseases, such as cholera, intestinal looseness, malaria, typhoid, and others, which regularly execute more than 3 million people[2]. It is essential to reduce the gap between the requirements and supply of drinking water

through different refining water methods to resolve the water shortage problem. Renewable energy is environment-friendly and is known as green energy. Solar energy is a helpful way to find small quantities of pure water from salt-water. For the last few years, solar desalination (SD) has become a subject of significant interest [3].

Figure 1 illustrates the global use of freshwater. Drinking water consumes typically 2 to 4L/day by adults and 0.5 L/day by offspring. Water consumed for a domestic purpose like washing and cooking, etc., significantly varies from country to country between 50 to 500 L/day. The need for freshwater is comparatively large in the agriculture area. A large amount of agricultural water is required to produce high-value products like grain-fed cattle in hot climatic conditions [4].

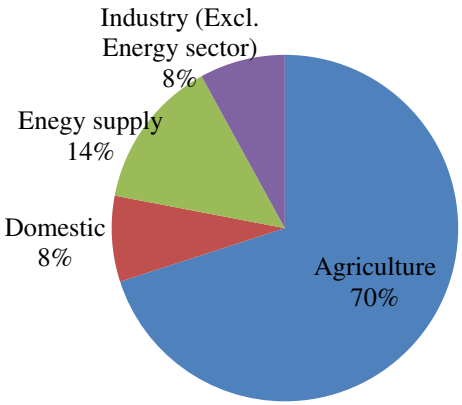


Fig. 1 Global use of fresh water

Desalination is the process that converts saline or brackish water into pure or freshwater. Desalination systems use one of the following procedures to remove or decrease salt from saline water [1]

- Phase change methods: The most prevalent techniques are multi-effect boiling (MEB), multi-stage flashing (MSF) distillation, and Membrane distillation (MD), which are driven by thermal energy. Humidification-dehumidification (HD), compression of mechanical vapors (MVC), and freezing are methods of phase change powered mainly by mechanical energy.
- Reverse osmosis (RO) and nano-filtration are the processes of a membrane that are mainly powered by mechanical energy (pressure).
- Electro dialysis (ED) or polarization of ion concentration is the process that is powered by electrical energy and utilized to remove salt from saline water.

Different types of desalination methods are shown in Fig. 2. Solar desalination is the technology that uses solar energy as a fuel to convert saline water into pure or freshwater.

There are many processes available to convert salty water into pure water; one is solar desalination. Conventional solar

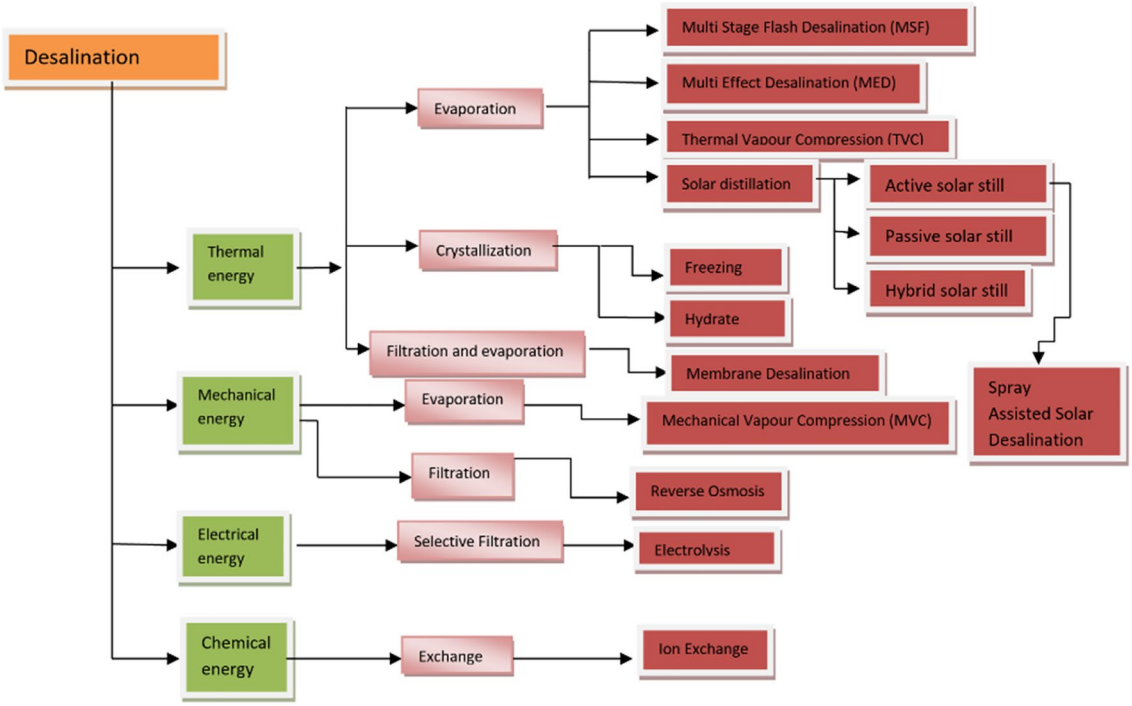


Fig. 2 Desalination processes

stills are simple solar stills that work on the principle of solar desalination. The productivity (yield) and thermal efficiency of conventional solar stills are very low. For conventional solar still, the daily freshwater productivity at a depth of 20 mm is 4.5 L and 2.31 kg/m^2 (Fathy et al. 2018) [5]. Sohani et al. [6] investigate the effect of side mirrors and sun tracking system on the performance of active and passive solar still. Performance of active and passive solar still has increased by using side mirrors and sun tracking. Kariman et al. [7] performed energy and exergy analysis to determine the primary source of energy consumption and exergy destruction. Results showed that boiler and heat exchanger are the main components of exergy degradation. Jamil and Ali [8] presented a detailed analysis of various freshwater collection techniques and performed a cost analysis of the device to predict system performance.

Integration of spray assisted process could be the solution to improve the productivity and efficiency of the solar desalination system. Spray flash evaporation is a successful technique of desalination driven by solar energy. Figure 3 shows the spray-assisted low-temperature desalination system. In this invention, ocean water is preheated in the feed tank by solar energy and then pumped into the evaporation chamber and sprayed by a spray nozzle at a lower pressure than the seawater atmospheric pressure. A small amount of ocean water transforms into vapor due to tremendous flash evaporation in the evaporator. When the vapor is condensed, freshwater is subsequently acquired. These attractive characteristics have drawn growing attention in research at spray flash evaporation.

Many researchers have conducted studies about spay assisted solar thermal desalination systems. Phenomena of spray flash evaporation were studied, and the temperature profiles of sprayed water were mapped axially [9]. Performance of a vacuum spray flash desalination system was investigated using theoretical and experimental methods. Study demonstrates how the vacuum chamber's design parameters (nozzle type, diameter, and length) influence the evaporation rate [10]. Mathematical models were introduced to evaluate the energetic, environmental, and economic performances of spray assisted low-temperature desalination (SLTD) systems [11]. Thermal model was established to evaluate the thermodynamic performance of the hybrid SLTD system. Energy and exergy efficiencies have been evaluated of the hybrid system with the help of the model under varying design and operating conditions [12]. Developed solar still was compared with conventional solar still in active and passive mode and describes that the productivity of developed solar still was more than conventional solar still by 51–148% [13]. A pilot plant was designed and constructed to evaluate one-stage technique to enhance productivity by increasing evaporation and condensation. Outcomes show that the inlet temperature of hot water and flow rates affects productivity, efficiency, and the output ratio of the desalination unit [14]. Ikegami et al. [15] compared upward jet flash evaporation to downward jet flash evaporation. Productivity of the downward jet technique was lower than upward jet flash evaporation.

Many of the theoretical models published in the literary work concentrated on steady-state performance study without transient developments. Because of intermittent solar radiation, it is difficult to keep constant operating conditions in solar-powered systems. Hence, an unsteady-state mathematical model for predicting dynamic performance is still to be developed.

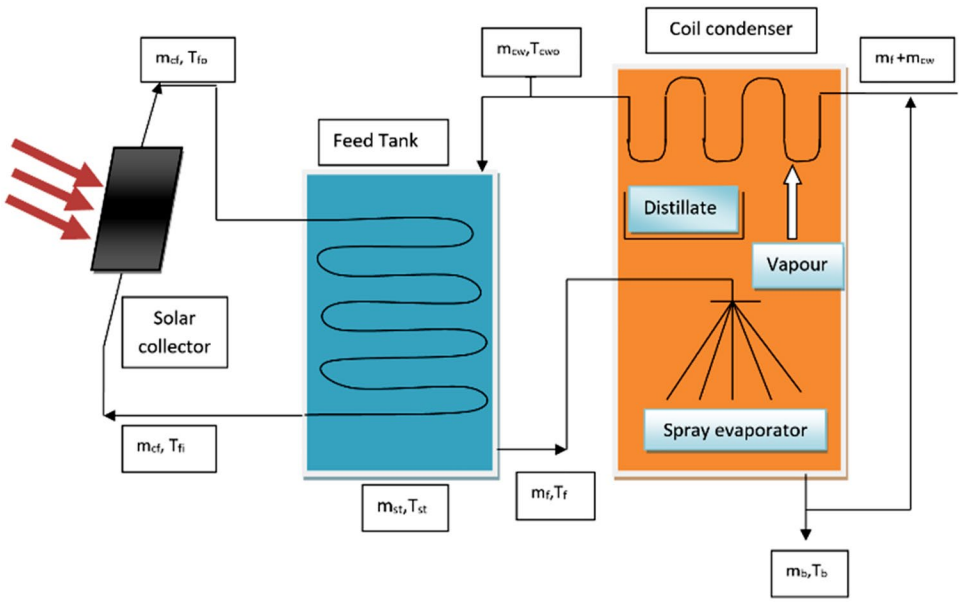


Fig. 3 Spray assisted low temperature solar thermal desalination system

The present study aims to provide comprehensive information about the performance, various available thermal and numerical models for the SLTD. There is not a single review paper on performance and modeling of spray-assisted system. Therefore, there is a need for present hour to compile this field's advancement to help researchers develop an energy-efficient SLTD system. Review consists of working principle, performance, thermal modeling, and numerical modeling of the spray-assisted solar desalination system. This study will be constructive to researchers and scientists for further research and improvement in designing spray-assisted solar desalination systems.

2 Advancement in Performance of Spray-Assisted Solar Desalination Systems

Performance of the system was evaluated by calculating the productivity and thermal efficiency**.** Thermal performance of SLTD systems evaluated by various authors is discussed in this section. Chen et al. (2018) measured the performance of SLTD system with 7.6 m^2 area of solar collector and 305 L water storage tank. Yield of the system reached to 30 L per day [16]. Miyatake et al. (1981) experimentally investigated the impact of liquid temperature on the spray flash evaporation by the varying inlet jet temperature from 40 to 80 °C. Spray flash evaporation has a higher evaporation rate and higher evaporation performance, even at a lower temperature than flash evaporation occurring in other systems [17]. Eltawil and Omara (2014) experimentally investigated the performance of hybrid (developed) solar still and compared it with conventional solar still. Hybrid solar still is shown in Fig. 4. The developed solar still consists of a flat plate collector, spray unit, and external condenser. Galvanized steel metal sheet with 2 mm thickness is used to make the basin of both of the solar stills (conventional and developed solar still) having an area 0.50 m^2 (1.0 m × 0.50 m) each. Experimental results show that the productivity of convention solar still (CSS) varies between 3 to 4 L/m^2. Productivity of developed solar (DSS) still was 51–148% more than the CSS (Fig. 5). The DSS productivity is enhanced by 56% and 82% by spraying circulated hot water in active and passive modes without using a condenser [13].

El-Agouz et al. (2014) performed an experimental study on the spray evaporation solar desalination system (Fig. 6). Outcome shows that the inlet temperature of hot water and flow rates affects the productivity, efficiency, gained output ratio of the desalination unit. Maximum daily productivity is 9 L/m^2, which is higher than humidification—dehumidification desalination system. Maximum daily efficiency was 87% [14].

Kabeel et al. (2018) conducted an experimental study on a hybrid solar desalination system (SS-HDH). Solar air heater is integrated with the SS-HDH desalination unit. In this study, productivity, humidification efficiency, gained output ratio are calculated. Maximum yield of the system was 18.25 L/m^2 per day at water feed rates of 0.03 kg/s. Humidification efficiency was 79%. It influenced much by the increase in water feed rate compared to the air feed rate. Overall efficiency of the unit ranges from 21 to 39%. Gained output ratio was 2.57 [18]. Muthunayagam et al. (2005) developed a pilot plant that demonstrates the feasibility of desalination by low-pressure vaporization of seawater between 26 °C and 32 °C temperatures. Plant worked at a vacuum pressure of between 1.3 kPa and 2.3 kPa. Salty water sprayed into the

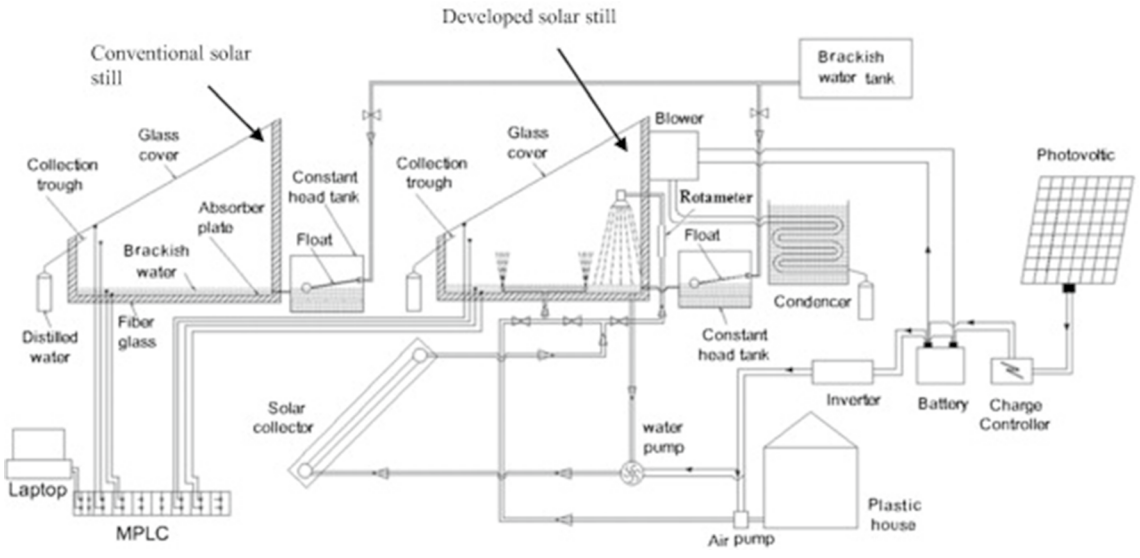


Fig. 4 Hybrid solar still (conventional and developed solar still) [13]

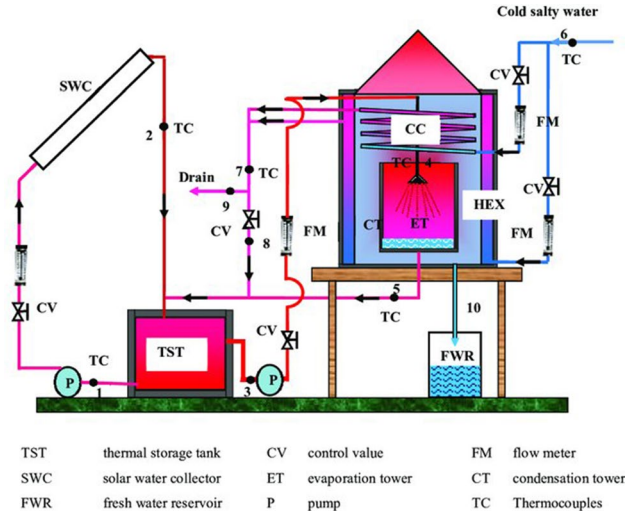


Fig. 5 Developed solar desalination system [14]

Fig. 6 Experimental setup [14]

evaporation tower using a spray nozzle and evaporated at low pressure. Yield increases with an increase in temperature of feed water for all values of vaporizer pressure [19]. Significant remarks on performance of spray-assisted solar desalination systems are given in Table 1.

2.1 Performance Parameters

Expressions to evaluate the hourly efficiency, productivity rate, gained output ratio and overall efficiency are given as [14]:

Hourly efficiency is calculated as:

$$\eta_h = 100 \times \frac{(M_w \times L_{w-\mathrm{av}})/3600}{A_s H + P_p} \tag{1}$$

Daily efficiency is calculated as:

$$\eta_d = \frac{1}{n} \sum_{1}^{n} \eta_h \tag{2}$$

$$L_{w-\mathrm{av}} = 10^3 \times \left(2501.9 - 2.40706 T_w + 1.192217 \times 10^{-3} T_w^2 - 1.5863 \times 10^{-5} T_w^3\right) \tag{3}$$

where P_p is the energy of the pump which is given by $P_p = V \times I$.

Ratio of pure water to the inlet flow rate output of salty water is defined as the productivity rate (PR) and is calculated as:

$$\mathrm{PR} = \frac{M_w}{m_{w,\mathrm{hi}}} \times 100 \tag{4}$$

Gained Output Ratio (GOR) is expressed as:

$$\mathrm{GOR} = \frac{M_w \times L_{w-\mathrm{av}}}{\left[(m \times h)_{w,\mathrm{hi}} - (m \times h)_{w,\mathrm{ho}}\right]} \tag{5}$$

where $\left[(m \times h)_{w,\ \mathrm{hi}} - (m \times h)_{w,\ \mathrm{ho}}\right]$ describes the total enthalpy difference between entering and leaving the hot water system. The flow rate of hot water leaving the system is expressed as:

$$m_{w,ho} = m_{w,hi} - M_w \tag{6}$$

3 Modeling of Spray-Assisted Solar Desalination System

3.1 Thermal Modeling

Thermal modeling is an operational approach to calculating the efficiency of any thermal system for theoretical research. Thermal models for solar desalination systems are developed based on the energy balances of its various parts. These models are beneficial to the pre-manufacturing design of systems and save time.

Chen et al. (2018) developed detailed thermodynamic, environmental, and economic models for the SLTD system to predict the brine temperature and specific energy consumption. Multi-objective optimization is conducted using these mathematical models to determine the optimal parameters of design and process based on the genetic algorithm. The main sources of primary investment and greenhouse gas emissions will also be recognized and key viable solutions will be suggested to boost system performance [11].

Wahab et al. (2012) developed a mathematical model for a multi-stage flash desalination system to predict the brine temperature, distillate and cooling brine, productivity and

Table 1 Performance of spray-assisted solar desalination

Authors and references	Productivity	Salient feature
Chen et al. [16]	30 L per day	Using low-grade heat sources, the SLTD system can produce pure water of high quality Production rate increased with an increase in feed flow rates, feed water levels, feed water Temperature, and differences in temperature between cooling water and feed water
Eltawil and Omara [13]	51–148% more than the conventional system	Productivity increases by 51% when the external condenser is utilized with basin solar still When flat plate solar collector is coupled with basin solar still, the yield increases in comparison to conventional solar still by: 56% and 82% for sprayed hot water (passive and active mode) 76% for hot water jet (active circulation) When water solar collector and condenser are paired with basin solar still, the productivity increases in comparison to conventional solar still by: 104% and 141%, for spray hot water (passive and active mode) For hot water jet—132% (active mode)
El-Agouz et al. [14]	9 (L/m^2)	Desalination unit's productivity, efficiency, and output ratio are affected by the flow rate and inlet temperature of the hot water Yield increases with the increase in water flow for the same inlet average temperature of hot water Highest yield of the system was achieved to 9 L/m^2 per day Maximum yield of SLTD system is more than the humidification-dehumidification desalination system Highest daily efficiency reached 87% in the new desalination unit After desalination, the TDS of saline water is 40 ppm Distilled water cost reached $ 0.029 per liter
Muthunayagam et al. [19]	–	Experimental results show excellent consistency with the theoretical model Productivity of plant was found to 4% theoretically. The quality of the distilled water is excellent
Miyatake et al. [17]	–	Spray flash evaporation has a higher evaporation rate and higher evaporation performance even at a lower temperature than flash evaporation

brine outlet flow rate. Author discussed the enthalpy balance of flashing brine, heat transfer, and mass balance for each stage (Fig. 7) [20].

Maroo and Goswami (2009) studied a single-stage and two-stage solar-powered desalination system and developed model to predict the productivity and performance ratio. Productivity of two-stage system was more compared to single-stage system. Single- and multi-stage systems are shown in Figs. 8 and 9, respectively. Mass, energy and salt balance equations were developed for each component in this model [21].

Chen et al. (2018) developed a thermal model to determine the energy losses and sources of inefficiencies with the heat balance, mass balance, and salt balance. Relations of energy utilization efficiencies for the system, as shown in Fig. 3, are also discussed in this study [16].

Chen et al. (2019) established a thermodynamic model to determine energy and exergy efficiency of the SLTD- TVC system. Impact of design and operational parameters on energy and exrgy efficiency and major variable for exergy degradation was discussed. Production ratio was improved by 10–35% of the system and affected by the cooling water flow rate, TBT, and motive steam pressure (Fig. 10) [12].

Chen et al. (2018) developed a numerical model focused on droplet analysis and applied it to translate the temperature profiles into the mean spray droplet diameters. Equations of heat and mass transfer were developed and discussed the process of splash evaporation of a jet of hot water splashed into a chamber of low pressure in this study (Fig. 11). Specific considerations were given to the flash evaporation method under lower temperature variations, resulting in increased thermal efficiency and improved desalination energy usage [9].

Major remarks on various thermal models for spray-assisted solar desalination systems are given in Table 2. Energy balance equations used in developed thermal models for the various SLTD are described in Table 3.

4 Numerical Modeling of Spray-Assisted Solar Desalination System

Software plays a key role in solar desalination system's analysis and simulation. Numerical simulation tools save time and cost in comparison to an experimental analysis of systems. This section discusses the different software used to design and test different solar desalination systems assisted by spray [1].

The mapping method from a given input to output is considered fuzzy, dependent on expert knowledge. Artificial

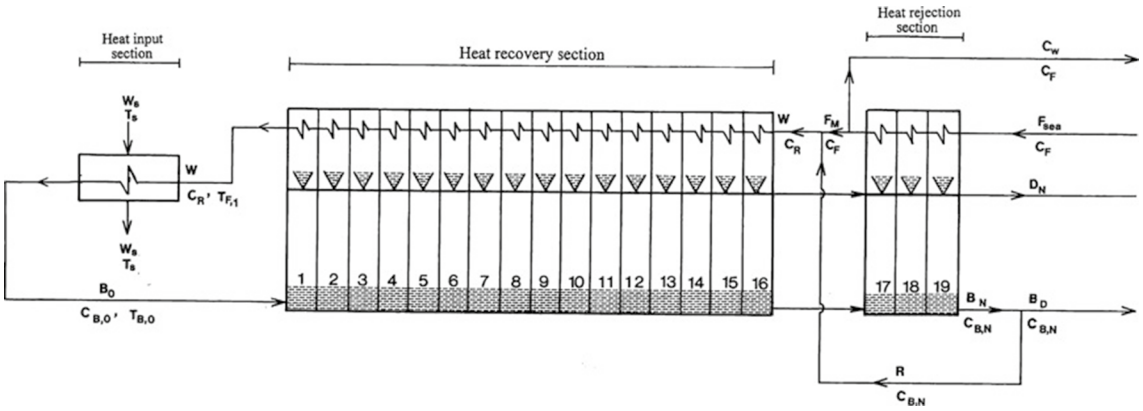


Fig. 7 Multi-stage spray flash desalination system [20]

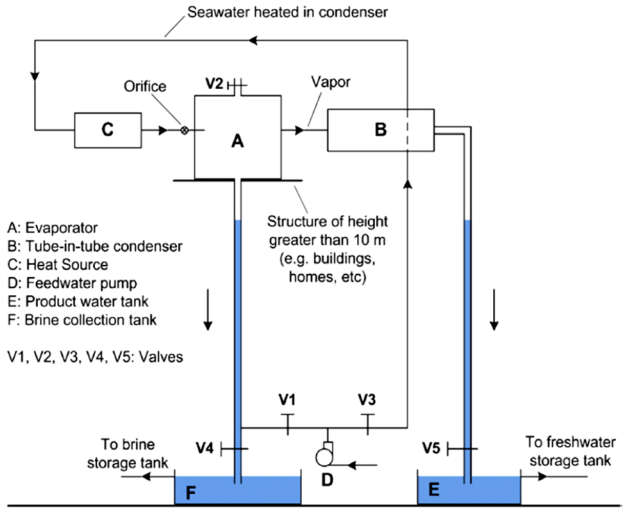


Fig. 8 Single-stage spray flash desalination system[21]

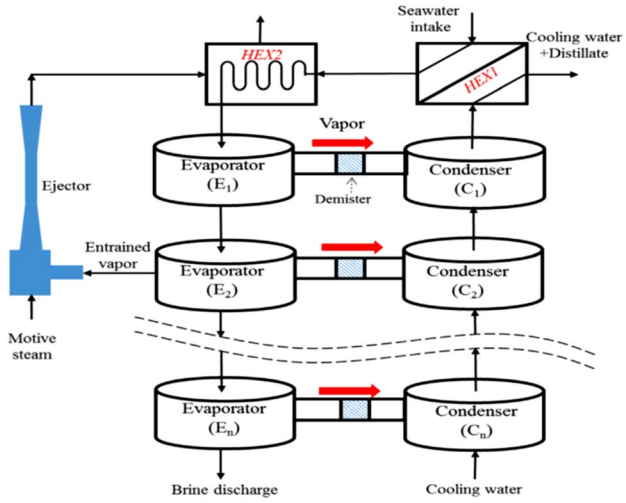


Fig. 10 SLTD-TVC system [12]

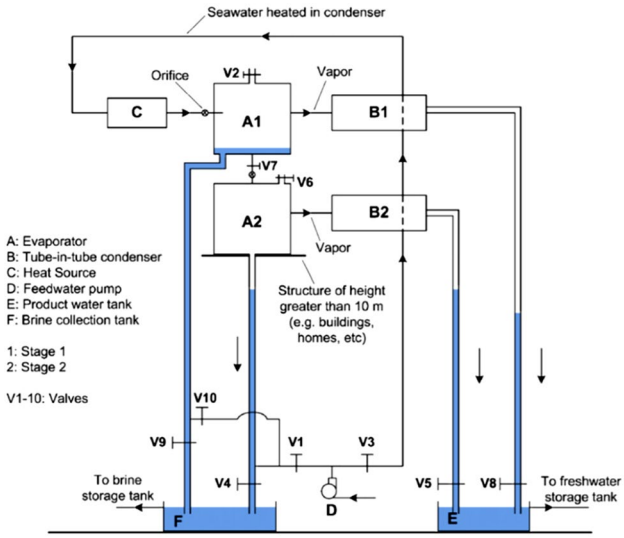


Fig. 9 Double stage spray flash desalination system [21]

Neural Networks (ANN) are a method for optimistic simulation [26]. Heat transfer and evaporation processes are involved in the desalination system. In such a condition, computational fluid dynamics (CFD) techniques are used.

Xuening et al. did CFD modeling of the spray evaporation system, which is combined with the solar parabolic trough collector for brine disposal [22]. Ra and Reitz established a model of vaporization for sprays of multi-component fuel. The existing vaporization models are incorporated in a multi-dimensional CFD framework and used to measure evaporation processes for droplets and sprays of single and multi-component fuel for varying ambient and droplet temperatures [23]. Wellmann et al. presented a low-temperature desalination model based on observational data and technical details. It included developing energy mass and material balance of media streams concerning the calculated data from the demonstration plan. The design is applied using the "Epsilon Professional" energy cycle simulation software

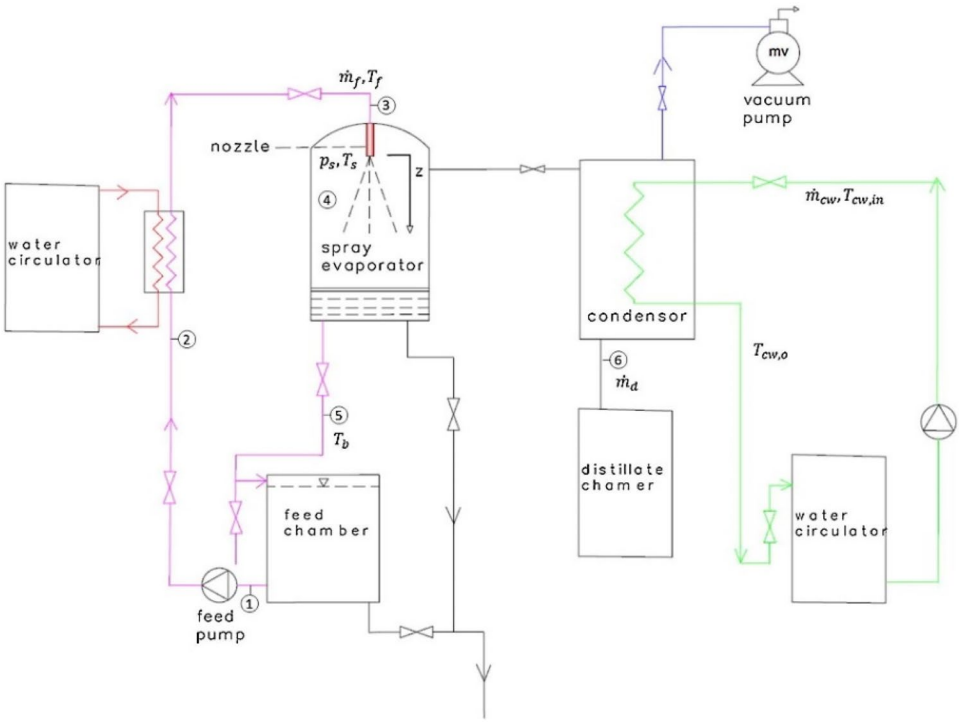


Fig. 11 Spray flash desalination system [9]

and paired with a solar power concentration (CSP) plant of 10 MW. Outcomes show the cogeneration in the CSP plant of 2.2 MW and the desalination of 520 m^3/d pure water [24]. Hoseinzadeh et al. (2020) solved differential equations related to solar collector's three-dimensional simulations using a finite volume method [25].

Araghi et al. (2017) did numerical modeling of spray flash desalinator with CFD software. Suggested desalinator was designed by utilizing the CFD FLUENT 16.2 kit, and some tests were compared to thermal model found in ASPEN/HYSY 8.0. Defined thermodynamic models based on the vapor–liquid balance in Aspen and Fluent have been shown to predict evaporation rates with average errors of 5% and 17%, respectively. A discrete phase model (DPM) approach can also evaluate the desalinator thermo-fluid environment with an appropriate accuracy of approximately 9%. Profiles of droplets are modeled in size, speed, temperature, and concentration and fundamental mechanics are discussed concerning the structure of VSFE. The simulation process is shown in Fig. 12 [10].

Depending on the base case operating conditions and the present spray angle (90°), Fig. 13a and b illustrates the temperature of the spray and the decrease of the velocity of each streamline, respectively. A wide spray angle decreases the length of the trajectories of the droplets. As a result, temperature and speed decreased significantly relative to the steep angle. A steeper angle to spray results in a higher decline in temperature but imposes the need for a larger vacuum chamber to avoid contact with the demister by droplets. Further suggests the significance of determining the optimal vacuum chamber diameter and duration for developing an effective spray flash desalinator with a proper spray angle.

Concentration of saline water in the vapor is shown in Fig. 14 as a continuous phase. Also, it shows cross section of Z-axis and cut of first interval at 5 cm above nozzle, respectively. The concentration across the nozzle outlet is very dense. It reduces as the water spray breaks down from the orifice to the chamber wall over a streamline, resulting in the spray burst evaporation phenomenon. Depending on the profile obtained, it should be observed that with a drop in the concentration value of DPM, the percentage rate of evaporation within the vacuum drum is enhanced.

Turbulent frequency across the vacuum desalinator's vertical position appears in Fig. 15a. Peak turbulent strength occurs along the spray direction, and from the center of the chamber to the suction, exit begins to decrease. Contours of pressure near the spray, on the other hand, show that the positive pressure carried to the vacuum chamber by pressurized injected saline water has no significant effect on the stable vacuum conditions within the chamber, as illustrated in Fig. 15b. The positive pressure changes take place closer to the nozzle outlet and the magnitude stays constant within +1.5% [10].

Xuening et al. used CFD fluent software to simulate brine droplets' evaporation and diffusion process in high-temperature airflow. A discrete phase model is established in the spraying evaporation system and based on the theory of gas–liquid phrase and hydromechanics. A simulation

Table 2 Salient features of thermal models of spray-assisted solar desalination system

S.No	Authors	Thermal model	Salient features	References
1	Chen et al. (2019)	Energy balance, mass balance, and exergy destruction rate equation based thermal modeling for the evaporator, condenser, and steam jet ejector	Overall production rate increases by 10–35 percent with the incorporation of TVC into the SLTD system. The hybrid system's production ratio rises monotonously with increased motive vapor pressure and more operating levels Leading cause of exergy degradation was the steam jet ejector, contributing over 40% of the total exergy degradation A medium flow rate of cooling water, more operating stages, and lower motive pressure maximized the system's overall second–law efficiency	[12]
2	Chen et al. (2018)	Heat and mass transfer, heat and mass balance	Optimization profits a 16-stage process with a maximum temperature of the brine of 93 °C and 242 m^2 area of HEX1 when the highest priority is given to the specific energy consumption This configuration of the device results in low specific energy consumption of 78.52 kWh/m^3 with relatively high CO_2 emissions and primary investment	[11]
3	Chen et al. (2018)	Heat balance, mass and salt balance, thermal efficiency and solar efficiency	A maximum yield of 8.97 L/day was achieved in Feb Solar collector's inefficiency contributes to a major part of the energy loss, whereas the desalination unit's energy losses are also high. There is an excellent feed flow rate, which improves the solar collector's efficiency and minimizes the desalination unit's inefficiencies System can provide a continuous supply of pure water of 30L per day with 7.6 m^2 area of solar collector, cooling, and feedwater flow rate of 3.8 LPM	[16]
4	Chen et al. (2018)	Heat and mass transfer	Hot water sprayed through a circular nozzle into the vacuum chamber breaks into droplets because of the bursting of Microbubbles produced by nucleation Droplet's average diameter was found to be much smaller than the diameter of the nozzle. These fine droplets allow the process of evaporation to be finished within a span of less than 50 cm, ensuring an evaporator of reduced size	[9]
5	Wahab et al. (2011)	Mass and energy balance equations	TBT and the brine recycle flow rate were the key variables influencing the plant's output TBT plays an essential role in the MSF plant's output. Both the rate of production and the performance ratio will increase with raising the TBT Due to a rise in the flashing distance, the production rate rises. In contrast, the performance ratio rises due to the reduction in the latent heat of water vaporization at a higher temperature. Simultaneously, higher TBT results in increased development of the scale Increasing the recycle stream flow rate, on the other hand, increases the output of distillate, which adversely affects the performance ratio	[20]

Table 2 (continued)

S.No	Authors	Thermal model	Salient features	References
6	Maroo and Goswami (2009)	Energy, mas and salt balance equations, performance ratio	Productivity of two-stage system is 35–40% more than the single-stage system for the same feed water flow rate Performance ratio of two-stage system is also more than the single-stage system for the same feed flow rate System efficiency achieved is much higher than that of a traditional solar system	[21]

experiment is performed. This model is tested by comparing system temperature to experimental temperature under various feed rate conditions, the temperature at 280 °C, hot air flux at 66 m^3/h, and compressed air pressure at 0.2 MPa. Simulation results were highly consistent compared to calculate ones and the relative errors were below 5%. A fluent technique is used to get the temperature range distribution of complete towers at various feed rates. Temperature distribution of the evaporation tower at a feed rate of 13 L/h appears in Fig. 16a. A horizontal cross section of temperature distribution at 1.5 m in z-direction from the base appears in Fig. 16b [22].

A model of vaporization is defined for the sprays of multi-component fuel. Discrete multi-component (DMC) fuel methodology has been utilized to simulate the gasoline and diesel system fuel properties. It also stimulates the composition for standard and splash-boiling evaporation. Unsteady evaporation of droplets and sprays for single and multi-component fuel was assumed, as shown in Figs. 17 and 18. The initial size of the droplet for gasoline spray was set at 30 µm, to uncouple the impacts on the process of evaporation due to superheated drop conditions. Both fuels were assumed to have a cone angle of the spray of 15°. Position of the injection for both fuels is the same [23].

An unsteady internal heat flux model and a model to determine the droplet's surface temperature were proposed. An estimated solution for the quasi-stable equation of energy was utilized to describe the heat flux from the atmospheric gas to the droplet-gas interface, taking into account the inter-diffusion of fuel vapor surrounding gas. Current vaporization models were introduced in a CFD code to compute the evaporation processes of droplets and sprays of single and multi-component fuel for different atmospheric and droplet temperatures. Differences are discussed between the representation of model fuels utilizing single and multi-component fuel [23].

Researchers used numerous software for numerical modeling of spray-assisted solar desalination system. Specific features/utility of each software is compiled in Table 4.

5 Conclusions

Thermal modeling is a useful tool for designing energy-efficient spray-assisted solar desalination systems. For the spray-assisted solar desalination system, several researchers and scientists have used various modeling techniques. Most thermal models are based on energy balance, mass transfer, and heat transfer calculations in the evaporation tower and condensation tower. These are beneficial in the pre-manufacturing design, simulation of different SD systems that saves time and capital investment. Thermal models predicted performance of spray assisted SD systems. Operating and design parameters significantly impact the productivity and performance of solar spray-assisted desalination systems. Spray-assisted solar desalination provides better distillate output than conventional solar still.

Many researchers and scientists have used various software for spray-assisted solar desalination systems in optimizing design parameters. CFD (ANSYS, FLUENT) and ASPEN/HYSY8.0 simulation software are commonly used for thermal performance analysis. Partial differential equations based on energy balance have been widely used for various components of the spray-assisted solar desalination system. CFD (ANSYS, FLUENT) was used to simulate the spraying unit in the evaporation tower because it gives better results than other software. The significant findings of spray-assisted solar desalination systems to follow:

- Highest daily yield of the SLTD system is 9 L/m^2.
- Productivity of the upward jet injector spray desalination system was more than downward jet injector spray desalination system
- Major part of exergy degradation was found in the steam jet ejector, contributing over 40% of the total exergy degradation.
- Maximum second law efficiency of the desalination system could be achieved by reducing the motive vapor pressure, increasing operating stages, and keeping a medium flow rate of cooling water.

Table 3 Various thermal developed for SLTD

S. No	Thermal model	References
1	Temperature differences in the evaporator and condenser are identified by entry and exit conditions, based on the assumption of the lumped parameter and represented as:	[16]

$$(\mathrm{MC})_d \frac{\mathrm{d}T}{\mathrm{d}t} = \dot{Q}_{\mathrm{evap}} - \dot{Q}_{\mathrm{cond}} - \dot{Q}_{\mathrm{loss}}$$

$$\dot{Q}_{\mathrm{loss}} = h_{\mathrm{loss}} A_d \left(T_e - T_{\mathrm{amb}}\right)$$

Heat rate of evaporation is represented as:

$$\dot{Q}_{\mathrm{evap}} = \dot{m}_f C_p \left(T_f - T_b\right)$$

$$T_b = T_e + \theta \left(T_f - T_e\right)$$

Condensed rate is expressed as:

$$\dot{Q}_{\mathrm{cond}} = \left(\dot{m}_f + \dot{m}_{\mathrm{cw}}\right) C_p \left(T_{\mathrm{cw},o} - T_{\mathrm{cw,in}}\right)$$

$$T_{\mathrm{cw,in}} = \frac{r\dot{m}_b T_b + \left(\dot{m}_f - r\dot{m}_b + \dot{m}_{\mathrm{cw}}\right) T_{\mathrm{amb}}}{\dot{m}_f + \dot{m}_{\mathrm{cw}}}$$

$$T_{\mathrm{cw},o} = T_c - \left(T_c - T_{\mathrm{cw},i}\right) \exp\left(-\frac{U_c A_c}{\left(\dot{m}_f + \dot{m}_{\mathrm{cw}}\right) C_p}\right)$$

$$T_c = T_e - \mathrm{BPE}$$

$$\mathrm{BPE} = \left(-4.584 \times 10^{-4} T_e^2 + 0.2823\, T_e + 17.95\right) X_{\mathrm{st}} 2 + \left(1.536 \times 10^{-4} T_e^2 + 0.05267\, T_e + 6.56\right) X_{\mathrm{st}}$$

Temperature difference in the feed tank is represented as:

$$(\mathrm{MC})_{\mathrm{st}} \frac{\mathrm{d}T_{\mathrm{st}}}{\mathrm{d}t} = \dot{m}_f C_p \left(T_f - T_{\mathrm{st}}\right) + \dot{m}_{\mathrm{cf}} C_p \left(T_{\mathrm{cf},o} - T_{\mathrm{cf},i}\right) + h_{\mathrm{loss}} A_{\mathrm{st}} \left(T_{\mathrm{amb}} - T_{\mathrm{st}}\right)$$

$$T_{\mathrm{cf},i} = T_{\mathrm{st}} - \left(T_{\mathrm{st}} - T_{\mathrm{cf},o}\right) \exp\left(-\frac{U_{\mathrm{st}} A_{\mathrm{st}}}{\dot{m}_{\mathrm{cf}} C_p}\right)$$

$$T_{\mathrm{cf},o} = T_{\mathrm{cf},i} + \frac{\dot{Q}_{\mathrm{sl}}}{\dot{m}_{\mathrm{cf}} C_p}$$

where $\dot{m}_f$ is the flow rate of the inside fluid, $\dot{Q}_{\mathrm{sl}}$ is the heat gained from the solar collector and is expressed as:

$$\dot{Q}_{\mathrm{sl}} = \eta_{\mathrm{sl}} A_{\mathrm{sl}} - I$$

$$\eta_{\mathrm{sl}} = 0.822 - 2.77 \frac{T_{\mathrm{ave}} - T_{\mathrm{amb}}}{I} - 0.017 \frac{\left(T_{\mathrm{ave}} - T_{\mathrm{amb}}\right)}{I}$$

$$T_{\mathrm{ave}} = \left(T_{\mathrm{cf},i} + T_{\mathrm{cf},o}\right)/2$$

where A_{sl} is the area of solar collector and η_{sl} is the efficiency of the solar collector

Rate of pure water output is expressed as:

$$\dot{D} = \frac{\dot{Q}_{\mathrm{evap}}}{h_{\mathrm{fg}}}$$

Brine flow rate is calculated as:

$$\dot{m}_b = \dot{m}_f - \dot{D}$$

Salinity of the feed tank varied according to the inlet and exit streams and represented as:

$$M_{\mathrm{st}} \frac{\mathrm{d}X_{\mathrm{st}}}{\mathrm{d}t} = \dot{m}_f X_{f,\mathrm{in}} - \dot{m}_f X_{\mathrm{st}}$$

Salinity of the brine is computed by the salt conversation and expressed as:

$$X_b = X_{\mathrm{st}} \frac{\dot{m}_b}{\dot{m}_f}$$

Salinity of inlet flow is identified by the salinity of the brine and brine recycling ratio. It is given by the following expression:

$$X_{f,\mathrm{in}} = \frac{r\dot{m}_b X_b + \left(\dot{m}_f - r\dot{m}_b + \dot{m}_{\mathrm{cw}}\right) X_f}{\dot{m}_f + \dot{m}_{\mathrm{cw}}}$$

Productivity and energy utilization efficiencies examined the system performance. Productivity is achieved as:

Table 3 (continued)

S. No	Thermal model	References
	$D = \int \dot{D}\mathrm{d}t$ Thermal efficiency and solar efficiency determine the effectiveness of energy consumption. The ratio of the distillate's corresponding evaporative energy to the solar collector's thermal energy is known as the thermal efficiency and is given as: $\eta_T = \frac{\int \dot{D} h_{\mathrm{fg}}\mathrm{d}t}{\int \dot{m}_{\mathrm{cf}} c_p (T_{\mathrm{fi}} - T_{\mathrm{fo}})\mathrm{d}t}$ Ratio of the distillate's corresponding to evaporative energy to the overall energy input is described as overall efficiency $\eta_s = \frac{\int \dot{D} h_{\mathrm{fg}}\mathrm{d}t}{\int (\mathrm{IA}_{\mathrm{sl}} + P)\mathrm{d}t}$ In the above expression P is the power of the pump, which can be determined from mass flow rates, pumping efficiency, and pressure gradients by the following expression: $P = \frac{\dot{m}_f}{\rho}\frac{\Delta p_f}{\eta} + \frac{\dot{m}_{\mathrm{cf}}}{\rho}\frac{\Delta p_{\mathrm{cf}}}{\eta}$	
2	*Energy and mass balance in SLTD* For each evaporator and condenser, the mass balance equations are described as: $\dot{m}_{f,i+1} = \dot{m}_{f,i} - \dot{D}_{e,i}$ $\dot{m}_{\mathrm{cw},i} = \dot{m}_{\mathrm{cw},i+1} + \dot{D}_{c,i}$ In each stage, the energy balance can be written as: $\dot{m}_{f,i} h_{f,i} = \dot{m}_{f,i+i} h_{f,i+1} + \dot{D}_{e,i} h_{v,i}$ $\dot{m}_{\mathrm{cw},i} h_{\mathrm{cw},i} = \dot{m}_{\mathrm{cw},i+i} h_{\mathrm{cw},i+1} + \dot{D}_{c,i} h_{v,i}$ Balance of mass between the evaporator and condenser is stated as: $\dot{D}_{e,i} = \{\dot{D}_{c,i}(i \neq N_e)$ $\dot{D}_{e,i} = \{\dot{D}_{c,i} + \dot{m}_e(i = N_e)$ Ocean and cooling water temperatures exiting in each stage are expressed as: $T_{f,i+1} = T_{e,i} + (T_{f,i} - T_{e,i})\theta_{e,i}$ $T_{\mathrm{cw},i} = T_{c,i} - (T_{c,i} - T_{\mathrm{cw},i+1})\theta_{c,i}$ Condenser temperature is expressed as: $T_{c,i} = T_{e,i} - \mathrm{BPE}_i - T_{\mathrm{loss},i}$ Boiling temperature elevation is calculated as: $\mathrm{BPE}_i = (-4.584 \times 10^{-4} T_{e,i}^2 + 0.2823\, T_{e,i} + 17.95) X_{f,i}^2 + (1.536 \times 10^{-4} T_{e,i}^2 + 0.05267\, T_{e,i} + 6.56)\, X_{f,i}$ Temperature loss across the demister is calculated as: $T_{\mathrm{loss},i} = \exp\left[1.885 - 0.02063(1.8 T_{e,i} + 32)\right]/1.8$ Salt balance is written as: $\dot{m}_{f,i} X_{f,i} = \dot{m}_{f,i+1} X_{f,i+1}$ Heat recovery process is described as: $\dot{Q}_R = \dot{m}_{f,\mathrm{in}} c(T_{f,r} - T_{f,\mathrm{in}}) = \dot{m}_{\mathrm{cw},o} c(T_{\mathrm{cw},o} - T_{\mathrm{cw},r})$ In HEX2, further heating of seawater is done to its required temperature (TBT): $\dot{Q}_H = \dot{m}_{f,\mathrm{in}} c(\mathrm{TBT} - T_{f,r}) = \dot{m}_h h_{\mathrm{lv}}$ *Energy and mass balance of steam ejector* Mass balance is written as: $\dot{m}_h = \dot{m}_p + \dot{m}_e$ Flow rate of entrained vapor can be expressed as: $\dot{m}_e = w\dot{m}_p$ where the entrainment ratio is represented by w which is expressed as: $w = 0.296 \frac{P_e^{1.19}}{P_h^{1.04}} \left(\frac{P_p}{P_h}\right)^{0.015} \frac{\mathrm{PCF}}{\mathrm{TCF}}$ PCF can be calculated as: $\mathrm{PCF} = 3 \times 10^{-7} P_p^2 - 0.0009 P_p + 1.6101$	[12]

Table 3 (continued)

S. No	Thermal model	References
	TCF can be calculated as $\mathrm{TCF} = 2 \times 10^{-8} T_h^2 - 0.0009 T_h + 1.0047$ *Exergy balance equation* Exergy destruction rate for evaporator is written as: $\dot{E}_{d,e} = T_o \left[\left(\dot{m}_{f,i} - \dot{D}_{e,i} \right) c \ \ln \left(\frac{T_{f,i+1}}{T_{f,i}} \right) + \dot{D}_{e,i} \left(s_{v,i} - s_{f,i} \right) \right]$ Exergy destruction rates for the heat exchanger are written as: $\dot{E}_{d,\mathrm{HEX1}} = T_o \left[\dot{m}_{f,\mathrm{in}} c \ \ln \left(\frac{T_{f,r}}{T_{f,\mathrm{in}}} \right) + \dot{m}_{\mathrm{cw},o} c \ln \left(\frac{T_{\mathrm{cw},r}}{T_{\mathrm{cw},o}} \right) \right]$ $\dot{E}_{d,\mathrm{HEX2}} = T_o \left[\dot{m}_{f,\mathrm{in}} c \ \ln \left(\frac{\mathrm{TBT}}{T_{f,r}} \right) + \dot{m}_h c \ln \left(s_{h,i} - s_{h,o} \right) \right]$ Rate of exergy destruction for the condenser is expressed as: $\dot{E}_{d,c} = T_o \left[\left(\dot{m}_{\mathrm{cw},i} - \dot{D}_{c,i} \right) c \ \ln \left(\frac{T_{\mathrm{cw},i}}{T_{\mathrm{cw},i+1}} \right) + \dot{D}_{c,i} \left(s_{v,i} - s_{\mathrm{cw},i} \right) \right]$ Rate of exergy destruction for the steam jet ejector is expressed as: $\dot{E}_{d,\mathrm{TVC}} = T_o \left[\dot{m}_m s_m + \dot{m}_e s_e - \dot{m}_h s_h \right]$ Exergy destruction rate for brine disposal is expressed as: $\dot{E}_{d,\mathrm{brine}} = \dot{m}_{f,N+1} \left[c \left(T_{f,N+1} - T_o \right) + T_o \left(s_o - s_{f,N+1} \right) \right]$ Exergy destruction due to distillate removal is expressed as: $\dot{E}_{d,\mathrm{dist}} = \dot{D} \left[c \left(T_{\mathrm{cw},o} - T_o \right) + T_o \left(s_o - s_{\mathrm{cw},o} \right) \right]$ Performance ratio is calculated as: $\mathrm{PR} = \frac{\dot{D}}{\dot{m}_h}$ In the above equation $\dot{D}$ is the overall production rate, which is computed as: $\dot{D} = \sum \dot{D}_{e,i}$ Overall exergy losses are measured by the overall exergy destruction rate, which is written as: Second law efficiency of the system is expressed as: $\eta_{\parallel} = \frac{\dot{D}e_d}{\dot{D}e_d + \dot{E}_{d,\mathrm{system}}}$	
3	Conductive function in the spherical coordinate defines the heat conduction within the water droplet as: $\rho_l c_p \frac{\partial T_l}{\partial t} = K_l \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial T_l}{\partial r} \right)$ Conduction of heat in drop is symmetrical and boundary at the middle of the droplet is expressed as: $K_l \frac{\partial T_l}{\partial r} \ (\text{at } r = 0) = 0$ Conduction of heat on the surface of the droplet is expressed as: $K_l \frac{\partial T_l}{\partial r} \ (\text{at } r = d/2) = -h \left(T_l - T_s \right) - h_m \left(\rho_l - \rho_v \right) h_{\mathrm{fg}}$ Heat and mass transfer coefficients are calculated from the following relations: $\mathrm{Nu} = \frac{\mathrm{hd}}{k_v} = 2 + 0.0495 \ \mathrm{Re}_v^{0.55} \ \mathrm{Pr}_v^{0.33}$ $\mathrm{Sh} = \frac{h_m d}{D_v} = 2 + 0.0495 \ \mathrm{Re}_v^{0.55} \mathrm{Sc}_v^{0.33}$	[9]
4	Evaporator *Mass balance:* $\begin{bmatrix} \text{Mass flow rate of} \\ \text{feed saline water} \end{bmatrix} = \begin{bmatrix} \text{mass flow rate of} \\ \text{vapor generated} \end{bmatrix} + \begin{bmatrix} \text{mass flow rate} \\ \text{of falling brine} \end{bmatrix}$ $\dot{M}_{\mathrm{sw}}^t = \dot{M}_v^t + \dot{M}_b^t$ Energy balance in the evaporator ids expressed as:	[21]

Table 3 (continued)

S. No	Thermal model	References

$$\begin{bmatrix}\text{energy of feed}\\ \text{saline water}\end{bmatrix} = \begin{bmatrix}\text{energy of}\\ \text{vapor generated}\end{bmatrix} + \begin{bmatrix}\text{energy of}\\ \text{falling brine}\end{bmatrix}$$

$$\dot{M}^t_{\text{sw}} h^t_{\text{sw}} = \dot{M}^t_v h^t_v + \dot{M}^t_b h^t_b$$

where

$$\dot{M}^t_v = \frac{\dot{M}^t_{\text{sw}} \int_{T^t_v}^{T^t_{\text{sw}}} C^t_{\text{ps}} \mathrm{d}T}{h^t_{\text{fg}}}$$

Salt balance:

$$[\text{salt in feed saline water}] = \left[\text{salt in falling brine}\right]$$

$$\dot{M}^t_{\text{sw}} X^t_{\text{sw}} = +\dot{M}^t_b X^t_b$$

where

$$X^t_b = \left(\frac{\dot{M}^t_{\text{sw}}}{\dot{M}^t_{\text{sw}} - \dot{M}^t_v}\right) X^t_{\text{sw}}$$

Flashing efficiency is calculated as:

$$\eta_{\text{flash}} = \frac{\text{Actual vapor generation rate}}{\text{maximum possible vapor generation rate}}$$

$$\eta^t_{\text{flash}} = \frac{\dot{M}^t_{v,\text{actual}}}{\dot{M}^t_{v,\max}} = \frac{\int_{T^t_v}^{T^t_{\text{sw}}} C^t_{\text{ps}} \mathrm{d}T / h^t_{\text{fg}}\left(T^t_v\right)}{\int_{T^t_{v,\text{eq}}}^{T^t_{\text{sw}}} C^t_{\text{ps}} \mathrm{d}T / h^t_{\text{fg}}\left(T^t_{v,\text{eq}}\right)}$$

Concentrated brine column

Mass balance:

$$\begin{bmatrix}\text{change in pressure in}\\ \text{the system}\end{bmatrix} \propto \begin{bmatrix}\text{change in mass of water}\\ \text{in the column}\end{bmatrix}$$

$$M^{t+\Delta t}_{\text{ev}} = M^t_{\text{ev}} - \frac{A_{e,p}}{g}\left[P^{t+\Delta t} - P^t\right]$$

Salt balance:

$$\begin{bmatrix}\text{salt in the brine column}\\ \text{at the time } t+\Delta t\end{bmatrix} = \begin{bmatrix}\text{salt in the brine column}\\ \text{at the time of } t\end{bmatrix} + \begin{bmatrix}\text{salt added because of falling brine in}\\ \Delta t\end{bmatrix} - \begin{bmatrix}\text{salt in brine rejected from the system in}\\ \Delta t\end{bmatrix}$$

$$M^{t+\Delta t}_{\text{ev}} X^{t+\Delta t}_{\text{ev}} = M^t_{\text{ev}} X^t_{\text{ev}} + \dot{M}^t_b X^t_b \Delta t - \left\{\frac{A}{g}\left[P^{t+\Delta t} - P^t\right] X^t_{\text{ev}} + \dot{M}^t_b X^t_{\text{ev}} \Delta t\right\}$$

$$\frac{M^{t+\Delta t}_{\text{ev}}\left(X^{t+\Delta t}_{\text{ev}} - X^t_{\text{ev}}\right)}{\Delta t} = \dot{M}^t_b\left(X^t_b - X^t_{\text{ev}}\right)$$

Energy balance:

$$\begin{bmatrix}\text{energy of the brine column}\\ \text{at the time } t+\Delta t\end{bmatrix} = \begin{bmatrix}\text{energy of the brine column}\\ \text{at the time of } t\end{bmatrix} + \begin{bmatrix}\text{energy added by falling brine in}\\ \Delta t\end{bmatrix} - \begin{bmatrix}\text{energy of brine rejected from the system in}\\ \Delta t\end{bmatrix}$$

$$M^{t+\Delta t}_{\text{ev}} h^{t+\Delta t}_{\text{ev}} = M^t_{\text{ev}} h^t_{\text{ev}} + \dot{M}^t_b h^t_b \Delta t - \left\{\frac{A}{g}\left[P^{t+\Delta t} - P^t\right] h^t_{\text{ev}} + \dot{M}^t_b h^t_{\text{ev}} \Delta t\right\}$$

$$\frac{M^{t+\Delta t}_{\text{ev}} \int_{T^t_{\text{ev}}}^{T^t_{\text{ev}}} C^t_{\text{ps}} \mathrm{d}T}{\Delta t} = \dot{M}^{t+\Delta t}_{\text{ev}} \int_{T^t_{\text{ev}}}^{T^t_{\text{ev}}} C^t_{\text{ps}} \mathrm{d}T$$

vapor space

Mass balance:

Table 3 (continued)

S. No	Thermal model	References

$$\begin{bmatrix}\text{mass of vapor at time}\\ t+\Delta t\end{bmatrix} = \begin{bmatrix}\text{mass of vapor}\\ \text{at time } t\end{bmatrix} + \begin{bmatrix}\text{mass of vapor added in}\\ \Delta t\end{bmatrix} - \begin{bmatrix}\text{mass of vapor condensed in}\\ \Delta t\end{bmatrix}$$

$$M_{\text{vs}}^{t+\Delta t} = M_{\text{vs}}^{t} + \dot{M}_{v}^{t}\Delta t - \dot{M}_{d}^{t}\Delta t$$

$$\frac{M_{\text{vs}}^{t+\Delta t} - M_{\text{vs}}^{t}}{\Delta t} = \dot{M}_{v}^{t} - \dot{M}_{d}^{t}$$

Energy balance:

$$\begin{bmatrix}\text{energy of vapor}\\ \text{in vapor space at}\\ \text{time } t+\Delta t\end{bmatrix} = \begin{bmatrix}\text{energy of vapor}\\ \text{in vapor spaceat time } t\end{bmatrix} + \begin{bmatrix}\text{energy of vapor added in}\\ \Delta t\end{bmatrix} - \begin{bmatrix}\text{energy of condensed water in}\\ \Delta t\end{bmatrix} - \begin{bmatrix}\text{energy rejected}\\ \text{to condenser}\end{bmatrix}$$

$$M_{\text{vs}}^{t+\Delta t}h_{\text{vs}}^{t+\Delta t} = M_{\text{vs}}^{t}h_{\text{vs}}^{t} + \dot{M}_{v}^{t}h_{v}^{t}\Delta t - \dot{M}_{d}^{t}h_{d}^{t}\Delta t - \dot{Q}^{t}\Delta t$$

where $\dot{Q}^{t}$ heat transfer rate between the condenser and cooling water

$$\frac{M_{\text{vs}}^{t+\Delta t}h_{\text{vs}}^{t+\Delta t} - M_{\text{vs}}^{t}h_{\text{vs}}^{t}}{\Delta t} = \dot{M}_{v}^{t}h_{v}^{t} - \dot{M}_{d}^{t}h_{d}^{t} - \dot{Q}^{t}$$

Condenser

Heat transfer coefficient of condensation is calculated as:

$$h_{i} = 0.555\left[\frac{g\rho_{l}(\rho_{l}-\rho_{v})k_{l}^{3}h_{\text{fg}}^{\cdot}}{\mu_{l}(T_{\text{sat}}-T_{s})D}\right]^{1/4}$$

In the above equation $h_{\text{fg}}^{\cdot}$ is known as the modified latent heat, which is given as:

$$h_{\text{fg}}^{\cdot} = h_{\text{fg}} + \frac{3}{8}C_{p,l}\left(T_{\text{sat}} - T_{s}\right)$$

Overall heat transfer coefficient is determined as:

$$U^{t} = \frac{h_{i}h_{o}}{h_{i}+h_{o}}$$

Rate of heat transfer from the condensing vapors to cooling salty water is calculated as:

$$\dot{Q}^{t} = U^{t}A_{\text{cs}}\frac{\left(T_{\text{out}}^{t} - T_{\text{in}}^{t}\right)}{\ln\left(\frac{T_{\text{sat}}^{t}-T_{\text{in}}^{t}}{T_{\text{sat}}^{t}-T_{\text{out}}^{t}}\right)} = \dot{M}_{\text{sw}}C_{\text{ps}}^{t}\left(T_{c,\text{out}}^{t} - T_{c,\text{in}}^{t}\right)$$

$$T_{c,\text{out}}^{t} = T_{\text{sat}}^{t} - \left(T_{\text{sat}}^{t} - T_{c,\text{in}}^{t}\right)e^{\left(-\frac{U^{t}A_{\text{cs}}}{\dot{M}_{\text{sw}}C_{\text{ps}}^{t}}\right)}$$

In the context of the current system, two performance ratios can be specified, one based on the total heat falling on the solar collectors, and the other based on the assumption that heat input is the useful heat added to the system

$$\text{PR}_{1} = \frac{\text{heat utilised for evaporation}}{\text{useful heat added}} = \frac{\sum_{\Delta t} m_{d}h_{\text{fg}}}{\Delta t\sum_{\Delta t}\dot{Q}_{u}}$$

$$\text{PR}_{2} = \frac{\text{heat utilised for evaporation}}{\text{energy incident on collector}} = \frac{\sum_{\Delta t} m_{d}h_{\text{fg}}}{\Delta t\sum_{\Delta t}\dot{Q}_{u}}$$

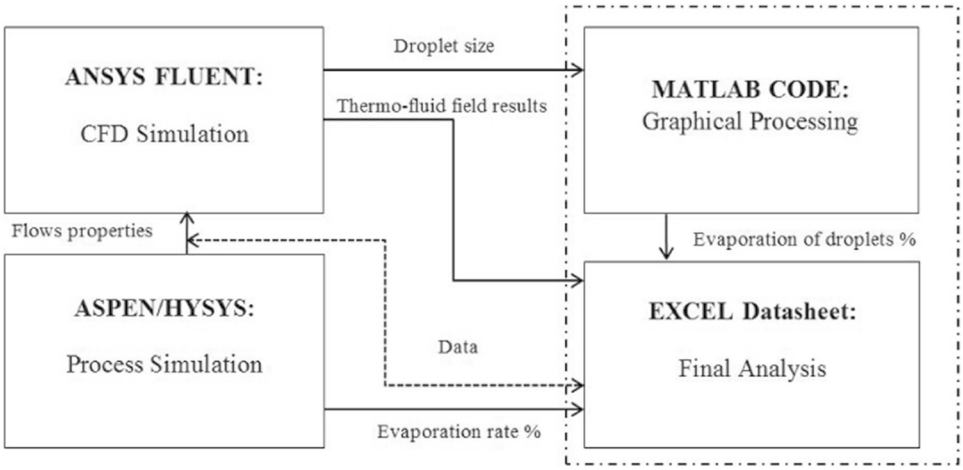


Fig. 12 Simulation algorithm [10]

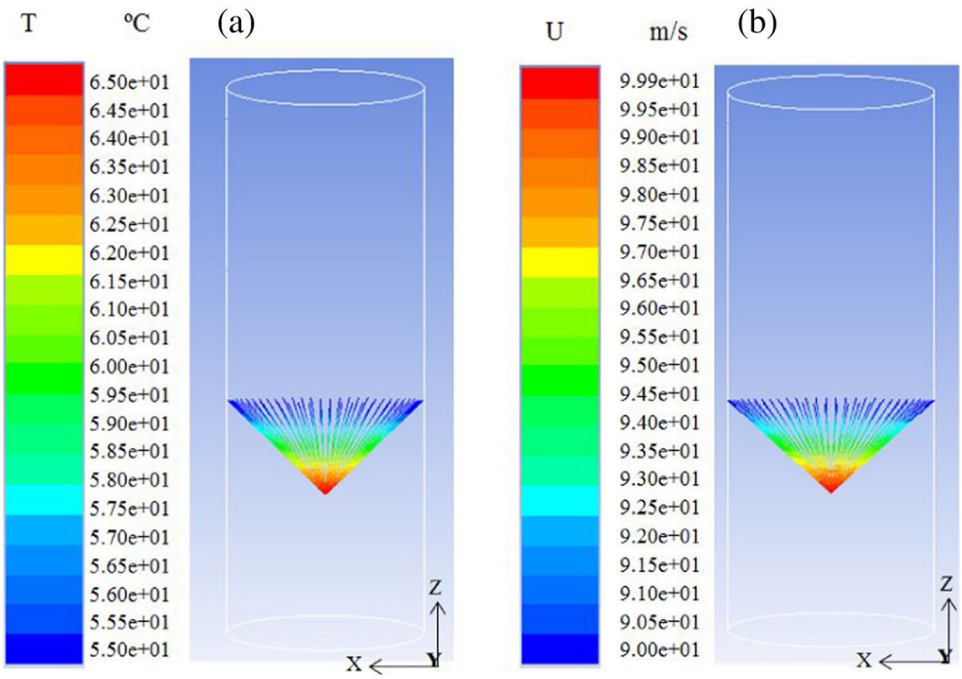


Fig. 13 Pattern of spray; **a** Temperatures of droplet streams **b** Velocities of droplet streams [10]

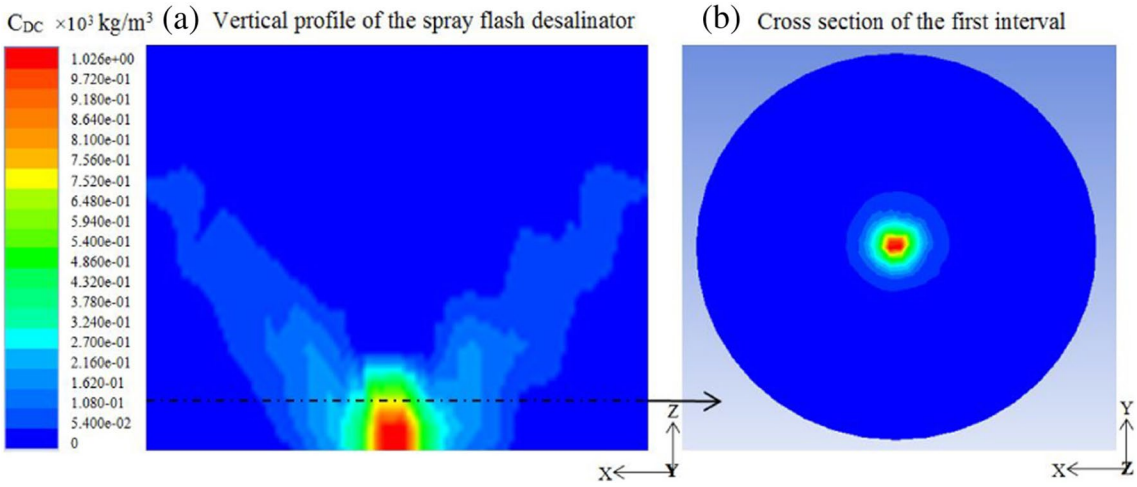


Fig. 14 Contours of the DPM concentration profile [10]

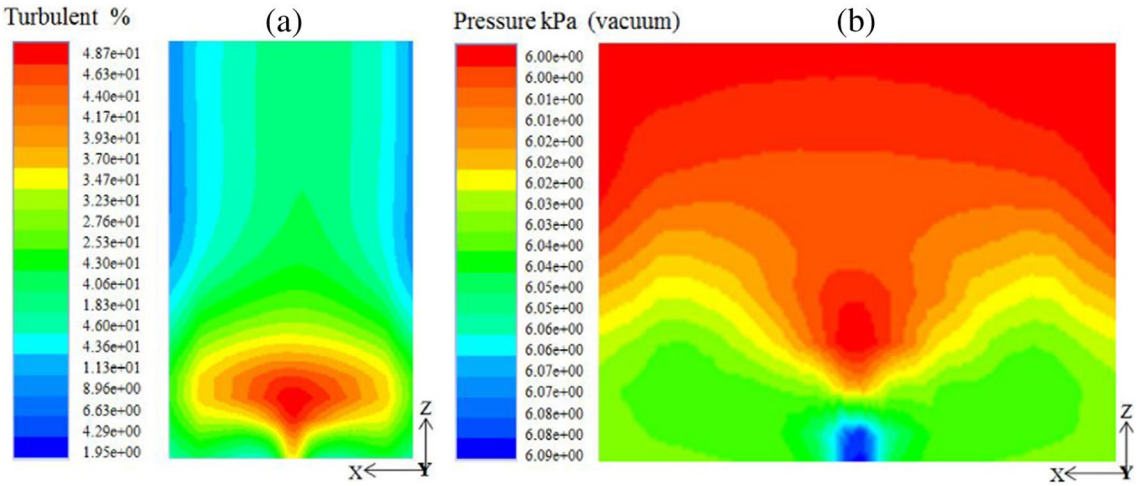


Fig. 15 **a** Turbulent frequency **b** Vacuum pressure across the chamber [10]

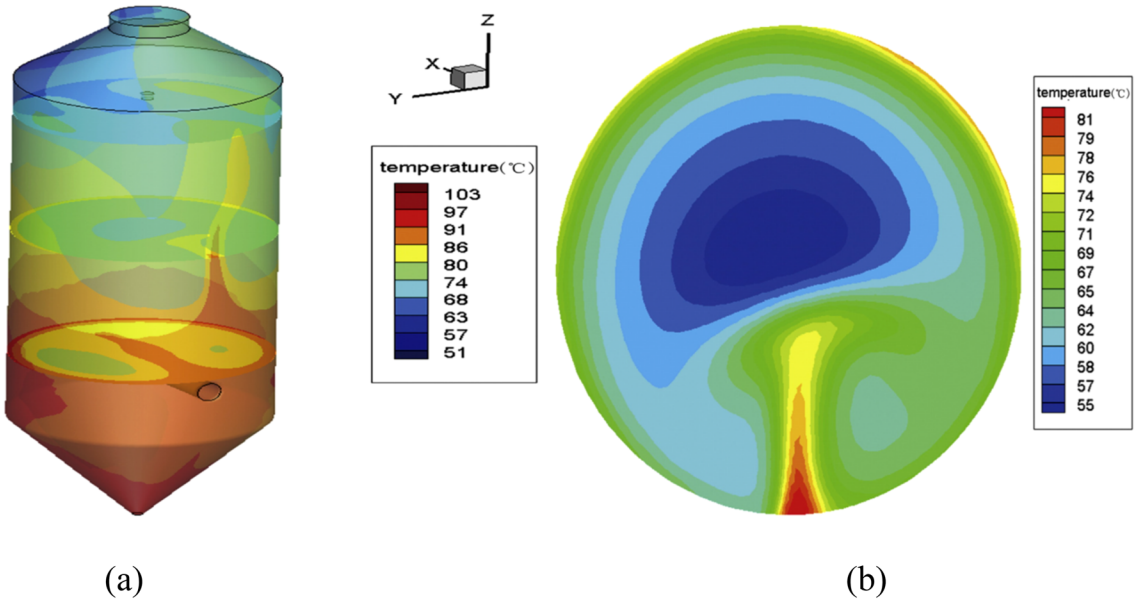


Fig. 16 **a** Temperature distribution in spray evaporation tower **b** Temperature distribution at $z = 1.5$ m from bottom [22]

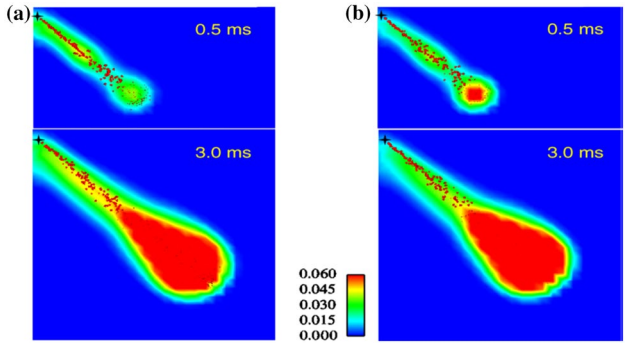


Fig. 17 **a** Droplets distribution and fuel mass fractions for multi-component gasoline **b** Droplets distribution and fuel mass fractions for single component iso-octane [23]

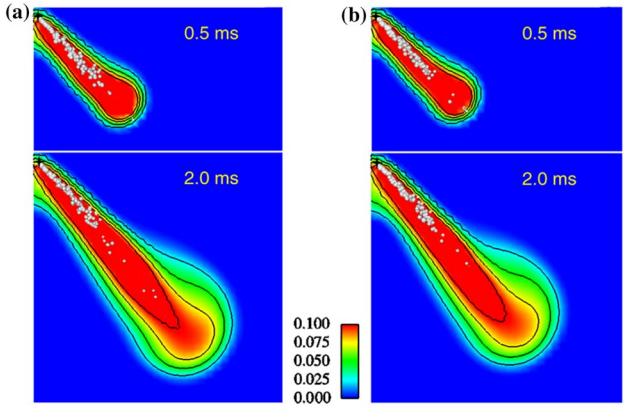


Fig. 18 **a** Droplets distribution and fuel mass fraction for multi-component diesel **b** Droplets distribution and fuel mass fraction for single component tetra-decane [23]

- Solar collector's efficiency contributes to a significant part of the energy loss, whereas the desalination unit's energy losses are also high.
- TBT and brine recycle flow rates were the key variables influencing the plant's output.
- Simulation outcomes based on various thermodynamics models embedded in ASPEN and FLUENT are consistent with an average variation of 5% and 17%, respectively, in experimental data of the evaporation rate. ENRTL model is more accurate than the other state equations.

Spray assisted has the benefits of higher rates of heat and mass transfer, the ability to work at moderately low-temperature differences, ease of system design, less scaling and fouling issues, and lower initial cost than conventional thermal desalination.

Table 4 Salient features of numerical models of spray-assisted solar desalination

S. No	Authors	Software	Salient features	References
1	Araghi et al. (2017)	CFD ANSYS FLUENT16.2, ASPEN/HYSY 8.0	Simulation outcomes based on the various thermodynamics models embedded in ASPEN and FLUENT are consistent with an average variation of 5% and 17%, respectively, in experimental data of evaporation rate. ENRTL model is more accurate than the other state equations CFD design has sufficient precision (about 9 percent) to determine the proposed desalinator's temperature profile Design parameters of the vacuum chamber influence-Evaporation rate, for example, the size of the nozzle, width, and diameter. However, more research is needed to identify the best model to achieve a higher rate of evaporation DPM-based CFD model suggests that the amount of evaporation rate within the vacuum chamber improves by reducing the DPM concentration value and peak turbulent frequency	[10]
2	Xueninig et al. (2015)	CFD ANSYS Fluent	A model was proposed to simulate the tower temperature. Experiments validated the numerical simulations with less than 5% error in observed temperatures in the evaporation tower. Highest performance is achieved under compressed air pressure 0.3 MPa, 330° C hot air temperature, and 14 L/h brine feed rate CFD simulation is useful in analysis as it can show the precise distribution of temperature inside the tower by which the test process can understand correctly CFD is used to model and analyze the evaporation process of the brine spray and can become a powerful tool for guiding the selection of operating conditions	[22]
3	Ra and Reitz (2009)	CFD	Current DMC evaporation model has successfully simulated the processes of vaporization for multi-component automotive fuels for different conditions of evaporation, involving super-critical and flash-boiling vaporization Reduced ambient pressure in droplet's evaporation of single multi-component fuel raises the rate of evaporation in the initial vaporization stages. It raises the constants of evaporation in later drop life period	[23]

Acknowledgements Authors are thankful to Department of Science and Technology, Govt. of India, to support this study under "India-Iran Bilateral Scientific and Technological Cooperation" (DST/INT/Iran/P-11/2018).

Author Contributions AK was involved in conceptualization, visualization, writing—review and editing, supervision, investigation. RK contributed to methodology, writing—original draft, S was involved in supervision.

Funding Department of Science and Technology, Govt. of India for supporting this study under "India-Iran Bilateral Scientific and Technological Cooperation" (DST/INT/Iran/P-11/2018) (information that explains whether and by whom the research was supported).

Availability of Data and Material This is a review article of previous work. Critical analysis has been done with reference to earlier research work. Hence, there is no data used.

Declarations

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

Human and Animal Rights Additional declarations for articles in life science journals that report the results of studies involving humans and/or animals.

References

1. Kumar, A.; Prakash, O.: Solar Desalination Technology. Springer, Berlin (2019) https://doi.org/10.1007/978-981-13-6887-5
2. Kant, R.; Prakash, O.; Tripathi, R.; Kumar, A.: Exergy Analysis of Active and Passive Solar Still. In: Kumar, A.; Prakash, O. (Eds.) Solar Desalination Technology. Green Energy and Technology. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-6887-5_12
3. Rani, A.; Suresh, S.; Kumar, A.: Review on thermal modeling of solar desalination systems. Res. J. Chem. Environ **23**, 90–102 (2019)
4. Pugsley, A.; Zacharopoulos, A.; Mondol, J.D.: Solar Desalination Potential Around the World. Renewable Energy Powered Desalination Handbook, p. 47–90. Elsevier, Amsterdam (2018) https://doi.org/10.1016/B978-0-12-815244-7.00002-7
5. Fathy, M.; Hassan, H.; Ahmed, M.S.: Experimental study on the effect of coupling parabolic trough collector with double slope solar still on its performance. Sol. Energy **163**, 54–61 (2018)

6. Sohani, A.; Hoseinzadeh, S.; Berenjkar, K.: Experimental analysis of innovative designs for solar still desalination technologies; An in-depth technical and economic assessment. J. Energy Storage (2020). https://doi.org/10.1016/j.est.2020.101862
7. Kariman, H.; Hoseinzadeh, S.; Heyns, P.S.: Energetic and exergetic analysis of evaporation desalination system integrated with mechanical vapor recompression circulation. Case Stud. Thermal Eng. **16**, 100548 (2019). https://doi.org/10.1016/j.csite.2019.100548
8. Jamil, F.; Ali, H.M.: Sustainable desalination using portable devices: a concise review. Sol. Energy **194**, 815–839 (2019). https://doi.org/10.1016/j.solener.2019.10.085
9. Chen, Q.; Kum, J.M.; Li, Y.; Chua, K.J.: Experimental and mathematical study of the spray flash evaporation phenomena. Appl. Therm. Eng. **130**, 598–610 (2018). https://doi.org/10.1016/j.applthermaleng.2017.11.018
10. Araghi, A.H.; Khiadani, M.; Sada, M.H.; Hooman, K.: A numerical model and experimental veri fi cation for analysing a new vacuum spray fl ash desalinator utilising low grade energy. Desalination **413**, 109–118 (2017). https://doi.org/10.1016/j.desal.2017.03.014
11. Chen, Q.; Kum Ja, M.; Li, Y.; Chua, K.J.: Energy, economic and environment(3E) analysis and multi-objective optimization of a spray-assisted low-temperature desalination system. Energy (2018). https://doi.org/10.1016/j.energy.2018.03.051
12. Chen, Q.; Kum Ja, M.; Li, Y.; Chua, K.J.: Energy, exergy and economic analysis of a hybrid spray-assisted low- temperature desalination/thermal vapor compression system. Energy **166**, 855–871 (2019)
13. Eltawil, M.A.; Omara, Z.M.: Enhancing the solar still performance using solar photovoltaic, flat plate collector and hot air. Desalination **349**, 1–9 (2014). https://doi.org/10.1016/j.desal.2014.06.021
14. El-agouz, S.A.; El-aziz, G.B.; Awad, A.M.: Solar desalination system using spray evaporation. Energy **76**, 276–283 (2014). https://doi.org/10.1016/j.energy.2014.08.0090360-5442
15. Ikegami, Y.; Sasaki, H.; Gouda, T.; Uehara, H.: Experimental study on a spray flash desalination (influence of the direction of injection). Desalination **194**, 81–89 (2006). https://doi.org/10.1016/j.desal.2005.10.026
16. Chen, Q.; Kum Ja, M.; Li, Y.; Chua, K.J.: Evaluation of a solar-powered spray-assisted low-temperature desalination technology. Appl. Energy **211**, 997–1008 (2018). https://doi.org/10.1016/j.apenergy.2017.11.103
17. Miyatake, O.; Tomimura, T.; Ide, Y.; Fuj, T.: An experimental study of spray flash evaporation. Desalination **36**, 113–128 (1981)
18. Kabeel, A.E.; El-said, E.M.S.: Experimental study on a modified solar power driven hybrid desalination system. Desalination **443**, 1–10 (2018). https://doi.org/10.1016/j.desal.2018.05.017
19. Muthunayagam, A.E.; Ramamurthi, K.; Paden, J.R.: Low temperature flash vaporization for desalination. Desalination **180**, 25–32 (2005). https://doi.org/10.1016/j.desal.2004.12.028
20. Abdul-wahab, S.A.; Reddy, K.V.; Al-weshahi, M.A.; Al-hatmi, S.; Tajeldin, Y.M.: Development of a steady-state mathematical model for multistage flash (MSF) desalination plant. Int. J. Energy Res. **36**, 710–723 (2012)
21. Maroo, S.C.; Goswami, D.Y.: Theoretical analysis of a single-stage and two-stage solar driven fl ash desalination system based on passive vacuum generation. Dealination **249**, 635–646 (2009). https://doi.org/10.1016/j.desal.2008.12.055
22. Xuening, F.; Lei, C.; Yuman, D.; Min, J.; Jinping, F.: CFD modeling and analysis of brine spray evaporation system integrated with solar collector. Desalination (2015). https://doi.org/10.1016/j.desal.2015.02.027
23. Ra, Y.; Reitz, R.D.: A vaporization model for discrete multi-component fuel sprays. Int. J. Multiph. Flow **35**, 101–117 (2009). https://doi.org/10.1016/j.ijmultiphaseflow.2008.10.00
24. Taylor, P.; Wellmann, J.; Neuhäuser, K.; Behrendt, F.; Lehmann, M.: Modeling an innovative low-temperature desalination system with integrated cogeneration in a concentrating solar power plant. Desalin. Water Treat. (2014). https://doi.org/10.1080/19443994.2014.940212
25. Hoseinzadeh, S.; Sohani, A.: Using computational fluid dynamics for different alternatives water fl ow path in a thermal photovoltaic (PVT) system. Int. J. Numer. Methods Heat Fluid Flow **31**, 1618–1637 (2020). https://doi.org/10.1108/HFF-02-2020-0085
26. Prakash, O.; Laguri, V.; Pandey, A.; Kumar, A.; Kumar, A.: Review on various modelling techniques for the solar dryers. Renew. Sustain. Energy Rev. **62**, 396–417 (2016). https://doi.org/10.1016/j.rser.2016.04.028

Turkish Journal of Computer and Mathematics Education Vol.12 No.13 (2021), 4110-4120

Research Article

Rumour Stance Classification using A Hybrid of Capsule Network and Multi-Layer Perceptron

Akshi Kumar*[1], Meghna Upadhyay*[2]

*Department of Computer Science & Engineering, Delhi Technological University,
Delhi, India

[1]akshikumar@dce.ac.in
[2]umeghna23@gmail.com

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 4 June 2021

Abstract- The accessibility and comfort of using social media have provided an optimal environment for people to expeditiously spread the information they have and sometimes without any knowledge of the authenticity of the information. Consequently, people inspect the stances reflected in the corresponding responses. To discover the certainty of rumour, stances are generally classified into 4 classes: support, deny, query and comment. This paper brings forward a model for the Stance Classification of Rumours on a Twitter dataset which utilizes the newly introduced Capsule Network along with Multilayer Perceptron. The rule-based strategy is used to merge the output of both the networks in a way that utilizes the strength of the two networks. The performance of the proposed model is surpassing the state-of-the-art with regard to the macro average F1-score indicating better results across different sets of classes.

Keywords: Capsule Network, MLP, Neural Network, Rumours, Stance Classification

I. Introduction

Misdirecting and deceiving individuals for personal or other reasons such as financial or political has become quite effortless with the advancement in social media which for the most part brings about hurting the people. Apart from sharing the information, social media has facilitated the creation of false information and manipulation of the information shared/exchanged by individuals. Rumour can be described as a fragment of data about a certain event/incident whose truth value is undetermined. It's circulating news without adequate information and proof to help it, in this way putting a question on its credibility. Therefore spreading rumours by the means of social media has become inevitable. Consequently, there is a growing requirement to analyze and work on rumours circulating expeditiously through social media.

Conventional newsgathering sources such as newspapers, news channels, or radio are still used by people and are reliable. But the boost in internet access speeds and smartphones has increased the use of social media and days of viewing social networking as a meaningless activity is a distant memory. In 2020, there are 3.96 billion people actively using social media, and this is an increase of 10.9% year-on-year from 3.48 billion in 2019. In case of a catastrophic event, the general public, media, and emergency responders make use of social networking platforms to search and circulate information related to that uncertain event. Correspondingly, this catastrophe-related information flow often results in the origination of rumours.

For the most part, it is not easy for an individual to precisely determine the authenticity of rumours by only going through the text contents of a post. Thus, people have started to focus on the stances given by other users in the responses to the particular post. Stance classification can be defined as the matter of recognizing the attitude captured by a user in a brief response text within a rumour microblog [1]. Commonly, stances reflected in the responses display support or deny with the source information. Although, for an uncertain circumstance people tend to express more than just support or deny. The RumorEval dataset [2] labels users' stance into four groups-support, deny, question, and comment (SDQC) [3]. Table I. shows example of rumours along with the responses.

- Support: The response text precisely reflects support towards the source rumour.
- Deny: The response text precisely denies the source rumour
- Query: The response text doesn't reflect either support or deny but demands some supplementary information concerning the rumour.
- Comment: The response text is a comment without any precise contribution of data which may lead to the determination of the veracity of rumour.

This paper presents an approach for the advancement of the stance classification system which is critical to therefore deciding the truthfulness of rumours. The task of determining the stance of rumours can be viewed as a

text classification problem where a rumour is considered as a plain text which needs to be classified into four target classes (SDQC). Consistently, text classification problems are accomplished by recurrent neural networks (RNNs) and convolutional neural networks (CNNs). The state-of-art utilizes LSTM based approach to determine the stance of the rumours. The paper brings forward a new model for determining the stance of the rumour. The model comprises Capsule Neural Network and Multilayer Perceptron (MLP) and the result achieved by this model is comparable to the state-of-the-art methods.

II. Related Work

Researchers have since quite a while ago took a gather at the characteristics of rumours to comprehend their propagation style and to recognize them from different sorts of data that are frequently shared by individuals. Two main factors of rumour diffusion claimed by Allport and Postman [4] are that people are always looking for some significance in matters and, when encountered with some sort of vagueness, people attempt to seek meaning by saying some convincing tale [5]. This claim supports why rumours with time generally change by turning out to be more limited, sharper, and more reasonable. However, Rosnow [7] declared that there are four salient factors responsible for the propagation of rumour. It essentially should be related to the end result of the audience, required to surge the anxiety of the person, be acceptable to a certain degree, and be undetermined. On contrary, Guerin and Miyazaki [6] proclaim that a rumour is a way of talking and interacting to build up a good relationship with people. Accordingly, it fulfils the purpose of creating and conserving relationships. Therefore, the task of rumour stance classification turns out to be significantly more relevant with the rampant increase of social media users.

The problem of rumour stance classification started to be extensively investigated in RumourEval task at SemEval 2017 [1]. Team Turing's paper describes a sequential approach to deal with the classification problem. After performing pre-processing steps which were removing non-alphabetic characters, converting them into lowercase, and tokenizing, they used word embeddings from Google News [8]. Along with this, additional features such as count of negative and swear words, punctuation marks, URLs, and pictures were also utilized to describe the rumour. The conversational structure of rumour was modeled by applying a LSTM-based sequential model and an accuracy of 0.784 on RumourEval test set was achieved [1] and thereby sets the state-of-art for the same.

One of the earliest works in the field of analyzing stances of rumours is done by Qazvinian in 2011 [9]. The work done was categorized into two parts: (1) Extracting rumours. (2) Belief Classification classifies the users who consider the rumour to be true versus those who do not believe the rumours to be true or question the rumour. Hence, it was considered as a binary classification task (support/ deny) as the deny class and the questioning class was consolidated into one class [1]. The approach used was based on building different Bayes classifiers as high-level features and then learning a linear function of these classifications for the retrieval in the first task and classification in the second task [9].

TABLE I: EXAMPLE OF RUMOURS ALONG WITH THEIR RESPONSES

Source Tweet 1: Coup in #Russia? Good article by @[user1]. [link1] #RamzanKadyrov #Putin #putindead [link2]"
Response 1:#RamzanKadyrov- A Sunni Muslim Russian/Chechen Ultranationalist w/ ties to the \"United Russia\" party(Putinism); #Russia's FSB(current KGB).
Response 2:So if #RamzanKadyrov had #BorisNemtsov assassinated was it at the direction of Putin or is/was Putin at the direction of FSB/Kadyrov? Hmm
Source Tweet 2: Latest on #Germanwings crash: Pilots signaled 911 before dropping out of midair; airline CEO calls this a \"dark day.\"[link1]
Response 1:"@[user1] Signalled 911? Called 'Mayday' would be more appropriate, factual reporting...
Response 2: @[user1] signaled 911?
Response 3: @[user1] you might want to change the use of 911 in this context."

Since 2015, there has been a sudden increase in enthusiasm for this task. Work done by Xiaomo Lie [10] proved to be better than Qazvinian. He implemented a rule-based method using an extensive list of positive, negative, and negation keywords with a set of language rules such as " negative words not preceded by a negation word

imply that the user denies the event" [10]. Further, Hamidian and Diab [11] employed a Tweet Latent Vector (TLV) feature and created a 100-d vector representation of each tweet [11] to access the capability of carrying out the task of binary classification of stances. Thereafter, a three-way classification that separates the denying and questioning classes was carried out by Lukasik [5]. He developed an automated, supervised classifier based on Gaussian Processes that use multi-task learning to classify the stance expressed in each individual tweet as either supporting, denying, or questioning the rumour [5]. In 2016, the set of labels considered in previous work were expanded to incorporate a new label commenting and thus making it a four-way classification. This was done by Michal Lukasik [12]. Along with this, he introduced the Hawkes process for the stance classification and demonstrated the importance of using temporal information of tweets along with textual content [12].

Moving away from the two-way classification, Zubiaga and Kochkina [13] adopted the four-way stance classification and the approach evaluated the tree structure of Twitter conversation instead of determining the stance on the basis of a single tweet alone and gave a considerably better performance in terms of macro-averaged F1-score. Similar sequential approaches exploiting the conversational threads have been analyzed and concluded that LSTM provides the best performance [14]. Eventually, deep neural network-based work started to rise to tackle the stance classification task. Lozano, Lilja, Tj¨ornhammar†, and Karasalo [15] combined CNNs with automatic rule mining and manually written rules. Jing Ma, Gao, and Wong [16] used Gated Recurrent Unit (GRU) for representing hidden units rather than LSTM to improve efficiency. In SemEval 2019 competition, Fajcik, Burget, and Smrz [17] built the pre-trained end-to-end Bidirectional Encoder Representations from Transformers (BERT) architecture and finished second in the competition. In 2020, A Stochastic Attention Convolutional Neural Network (SACNN) was introduced by Na Bai, Zhixiao Wang, and Fanrong Meng [18]. The model captures the different habits of the public for the rumour stance classification task [18] and the outcome of the model was higher than the state-of-art results. In addition to the deep learning neural networks, some scholars have utilized machine learning models for the stance classification task. Kaizhou Xuan and Rui Xia [3] implemented classical machine learning classifiers such as Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Linear Support Vector Machine (SVC), and Naive Bayes (NB). These classifiers were combined with 18 defined features including text features, user features, and propagation features. Logistic Regression classifier gave the best performance.

The work presented in this paper has an objective towards improving the stance classification system by taking advantage of Capsule Networks first introduced by [19].

A. The Capsule Network Model

In spite of the fact that various models built on deep neural networks have been offered for the problem of stance classification, there exists demand development in this project. This research is focused on using a Capsule Network to upgrade the accuracy of the Rumour Stance Classification system. In the introductory paper of Capsule Network, image classification was implemented on the MNSIT dataset and it revealed that the proposed model works better than Convolutional Neural Networks. A Convolutional Neural Network recognizes objects by detecting image features. Initial layers are responsible for detecting simple features such as edges and more complicated features such as eyes or nose (in case of face detection) are discovered by layers that are deep into the architecture. Therefore, the combination of all these features makes the ultimate prediction. It can be concluded that CNN does not make use of any spatial information and also the connection of layers is done by the use of the pooling function. "The pooling function used in the Convolution Neural Networks is a big mistake and the fact that it works so well is a disaster"- Geoffrey E. Hinton.

A Capsule Neural Network aims to perform inverse graphics. Given that, the method ventures to reverse engineer a technique that is responsible for the production of the desired image [20]. Equivariance is known to be one of the crucial features of the Capsule Network. Its main objective is to preserve the comprehensive information of the location and pose of the object all over the network. For example, considering the event of marginally rotating the target, there is also a marginal change in the activation vectors.

As defined by Geoffrey E. Hinton, a capsule is a group of neurons whose activity vector represents the instantiation parameters of a specific type of entity such as an object part [19]. The probability of the existence of an entity is defined by an activity vector's length. Also, the instantiation parameters are depicted by the orientation of the activity vector. The Capsule Network consists of many capsules at different levels. The particular level of capsules that are activated generates predictions by utilizing transformation matrices and are responsible for the initiation of the higher level capsule's parameters. A higher-level capsule gets activated as soon as multiple predictions agree. A repetitive procedure accountable for routing that decides the credit attribution among nodes of low and high level was also introduced [21].

Lately, Capsule Networks have gained substantial consideration and have been successful even in the field of NLP inclusive of text classification [22], sentimental analysis [23], fake news detection [20], and identifying toxic comments [24].

III. Proposed Model Architecture

The model put forward by the research done for the Rumour Stance Classification problem is given by Figure 1. The model consists of two neural networks, Capsule Neural Network and Multilayer Perceptron Neural Network. The Capsule Network takes sentences converted into vectors of a fixed length as input while the input to the Multilayer Perceptron is the features extracted from the text sentences. The combined result from both the neural networks produces the final class label for the rumour sentences. The layered architecture of the Capsule Network is shown in the Figure 2.

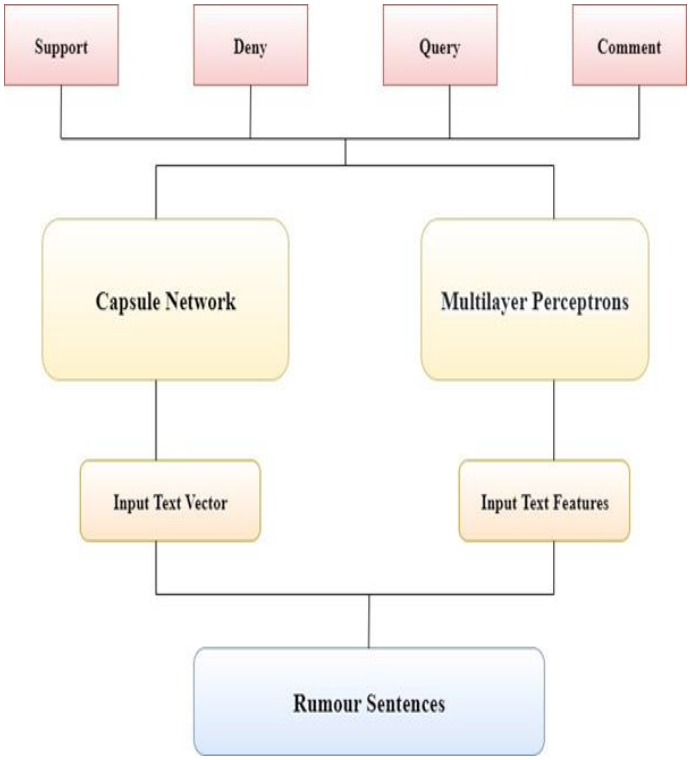


FIGURE 1: ARCHITECTURE OF THE PROPOSED MODEL

A. Embedding Layer

This is the first layer of the model. The layer is responsible for converting the integer encoded data into fixed-size vectors. The integer encoded data is an individual word represented by unique integers. The layer is loaded with arbitrary weights and it learns embedding for all the words present in the vocabulary. The output of the layer is fixed size dense vector embedding for each word.

B. Convolutional Layer

This layer is responsible for obtaining various features from the input sentence that are placed at different positions in the sentence. Convolution filters are intended for the same.

C. Primary Capsule Layer

The aim of this first capsule layer in the architecture is to reform the end product of the convolution layer into a capsule vector representation. This maintains the semantic meaning of the words in the sentence. 8 dimension capsules are present in this layer with 32 channels. Moving a layer up in the network depends upon an algorithm called routing by agreement. Said algorithm activates the capsule present in the above layer when multiple capsules of the current layer vote for the specific capsule of the above layer.

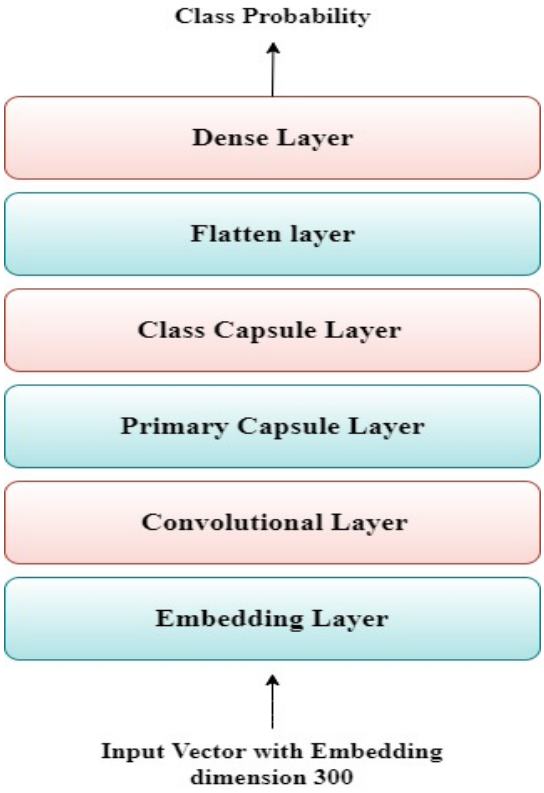


FIGURE 2: LAYERED ARCHITECTURE OF CAPSULE NETWORK

D. Class Capsule Layer

The intake of the layer is the output from the primary capsule layer. Every capsule present in the previously mentioned layer is in connection with the local region of the layer below. The connection between this layer and the previous layer is provided by the mechanism of dynamic routing. The capsules are of dimension 16 in this layer and the number of routing used in the iterative dynamic routing algorithm is 3.

Figure 3 depicts the architecture of Multilayer Perceptron Neural Network. The input to the network is a normalized text feature vector with a size equal to the number of features. Feature Extraction is an essential process and the context-based features used in the training of the network are explained below.

Sentence Length: This counts the number of words present in the rumour post excluding all URLs and special characters.

Upper Case Count: To express strong feelings, some people use all capital letters in the word, and therefore the count of such words is used as a feature.

Negation words: It is the count of the presence of negation in the rumour post. A vocabulary is built containing all the common words that can be utilized to express deny and words that represent a contradiction. Some of the words in the vocabulary are ‘never’, ‘isn’t’, ‘barely’, ‘no’, ‘shouldn’t’.

Abusive words: Most often people use dirty or abusive words to convey some strong feelings. A collection of such words is created which include words like ‘morons’, ‘fuck’, ‘stupid’, ‘bastard’, ‘bitch’. This collection also takes into account the slang representation of these words like ‘stfu’, ‘wtf’. Count of these words in the rumour post is used as a feature. These words appear more in deny or comment responses rather than query or support.

Question words: A small collection of question words such as ‘what’, ‘where’, ‘why’ is used and the count of occurrence of these words is used as a feature.

Denial words: The existence of words like ‘liar’, ‘unconfirmed’, ‘contradicting’ can be used to express the feeling of denial. These words are rarely present in support or query responses. Count of these words as a feature is used to distinguish deny class from other classes.

Belief words: A group of positive words such as 'agree', 'believe', 'sure', 'absolutely' is generated and the count of such words in a rumour post is used as a feature. These words appear more frequently in support responses.

Sentiment Score: This feature is constructed using VADER Sentiment Analysis tool. A compound score of rumour posts is used as a feature. The compound score ranges from -1 to +1. -1 demonstrates the most extreme negative score and +1 demonstrates the most extreme positive score. A chance of a low compound score is more in deny class.

Question Mark: For this feature, the presence of a question mark in the rumour post is taken as '1' and its absence as '0'. It plays an important role as the chances of the presence of a question mark in query response is higher than the other three classes.

Exclamation Mark: People tend to use exclamation marks to express emotions. The presence of an exclamation mark is assigned '1' and its absence '0'.

Presence of URLs: The rumour post containing URL links has a high possibility to show support towards the rumour as it can provide extra information in the favour of the rumour post. Therefore, the presence of URL links is significant to the support responses.

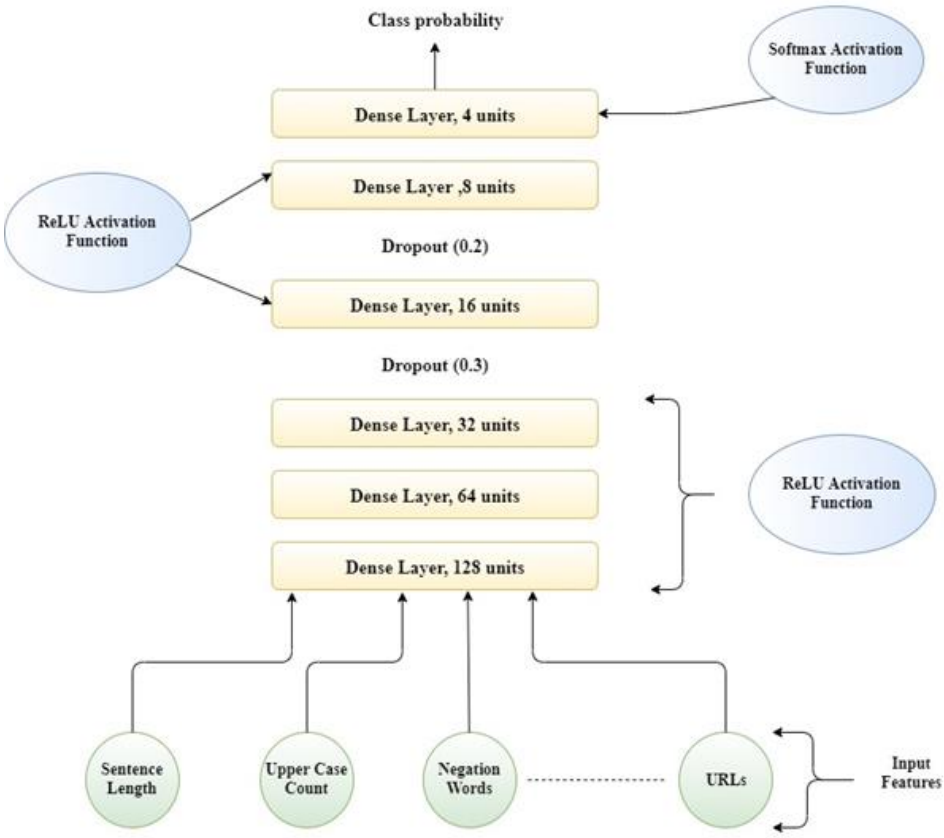


FIGURE 3: ARCHITECTURE OF MUTILAYER PERCEPTRON NEURAL NETWORK

VI. Dataset

To assess the efficiency of the suggested model for the Rumour Stance Classification task, the experiment was conducted on the RumorEval dataset formally came from real Twitter data. Twitter is a liberal wellspring of reports of breaking news. The dataset is composed of Tweets from eight separate incidents which include Ebola Essien, Ottawa shooting, Charlie Hebdo, Sydney Seige, Prince Toronto, Putin missing, Germanwings, and Ferguson riots [25].

Every rumour in each incident in the dataset is tagged as SDQC. The dataset incorporates a total of 5568 tagged rumours. Out of 5568 total tweets, 4238 tweets are used in the training set and 281 tweets are used in the development set. And for testing, 1049 tweets were used. The per-class distribution of tweets in the training, development, and testing is shown in Table II. It can be precisely observed that there is a big class imbalance in

the dataset. More than 60% are tagged as 'comment'. At the same time, 'deny' and 'query' combined make less than 16% of the total rumours. Because of this skewed nature of the dataset, class weights are used to compensate for the imbalance of the data. The calculated class weights are [S, D, Q, C] = [0.157, 0.396, 0.399, 0.048].

TABLE II: DISTRIBUTION OF THE TWEETS IN THE DATASET INTO INDIVIDUAL SETS

	S	**D**	**Q**	**C**
Training	841	333	330	2734
Development	69	11	28	173
Testing	94	71	106	778
Total	1004	415	464	3685

A. Data Inconsistency

Upon careful examination of the dataset, it has been found that for some few tweets the labelling of the class is slightly inconsistent. Consider the tweets presented in the Table III. The first and the second tweet convey somewhat similar meanings and are in response to the same source tweet but the first one is labelled as a comment and the second one is labelled as support. Similarly, there are few tweets that are labelled as query but there is no question asked in the tweet. To remove this data inconsistency, some of the tweets are relabelled after careful examination of tweet context and the source tweet. Training and development set combined consists of 4519 tweets. Out of these 4519 tweets, 260 tweets are changed which means only 5.75% of the tweets are relabelled. No change is made in the testing data set.

TABLE III. EXAMPLES SHOWING INCONSISTENCY IN THE TWITTER DATASET

I just feel sick RT @[user]: At least 12 dead in Paris shooting. Updated story:[link]	Comment
Awful. RT @[user]: At least 12 dead in the Paris shooting. [link]	Support
I'm in London right now, feel free to make an appointment @[user] @[user] @[user]	Query
@[user] @[user] good thats a start	Query
@[user] @[user] @[user] @[user] How would you know the pilot was Muslim?	Deny

B. Data Pre-processing

Before feeding the data into the neural network, there is a need to apply some transformation and this process is known as data pre-processing. The structure of the data must be in a proper form to get a better outcome. The pre-processing steps performed on the dataset are as follow:

- All the sentences in the dataset are tokenized. The purpose of tokenization is to split the sentences into individual units or words.
- Using all the tokens, a vocabulary is created which consists of all the distinct words in the dataset.
- The input is a sequence of words. So, a tokenizer is built to modify them into an integer sequence. Hence, every word is mapped to a unique integer.
- The dataset contains sentences of variable length. Therefore, all the sentences were converted into fixed-length sentences. Sentences of shorter size as compared to the maximum size sentence are packed with trailing zeros. These sentences are then fed to Capsule Neural Network.
- For Multilayer Perceptron Neural Network, all the feature data is normalized by using MinMaxScaler which scales each feature to a given range with a default range of 0 to 1.

V. Experiment

The proposed architecture for Rumour Stance Classification is given in the Figure 1. It consists of a single convolution layer before the capsule layer. Other architectures which include two convolution layers, three convolution layers, and parallel convolution layers were also tested. It was observed that these architecture models were not able to perform any better than the proposed one. The results were comparable and more layers only added to the complexity of the architecture with no noticeable improvement in the result. Similarly, an attempt with parallel capsule layers architecture was also made but didn't show any observable improvement. Hence the architecture with the least complexity was preferred over others as they all generated comparable results.

To apprehend the semantic meaning from the words in the sentences, word embeddings are used. It's a technique to represent separate words as a predefined vector. To initialize word embedding vectors, Keras Embedding Layer with a vector dimension of 300 is used. Further, these vectors are fed into the neural network. Consecutive to the embedding layer, there is an N-gram convolutional layer with N values as 9 and with a filter of size 256. The filter is set to move one unit at a time and therefore the stride is set as 1. Following all layers are capsule layers with capsule dimensions set as 8 in the first layer and 16 in the next layer. The first one is the primary layer followed by the class capsule layer. The connection of these capsule layers is achieved by using the transformation matrices and is multiplied by the coefficient generated by dynamic routing. The output generated by the class capsule layer is then flattened to a dense layer with softmax activation which brings out the class probabilities. Similarly, in the Multilayer Perceptron network, extracted input features go through a number of hidden dense layers and produce the class probabilities. To get the final class label, the per-class performance of both networks is carefully considered and merged in a way that brings out the best from both networks. The Multilayer Perceptron network provided better results for the query class in comparison to the capsule layer network. On the other hand, both the networks were not able to detect the deny class as well as the other classes. For the improvement of the same, few rules are developed after examining the training and development dataset and these were used to merge the result from the neural networks. It has been observed that phrases like 'stop spreading these lie', 'I believe otherwise', 'there is no truth', are commonly used in tweets which are labelled as deny. So the occurrence of phrases with similar meaning implies that the probability of 'deny' class is greater than the other classes. Eventually, the class with maximum probability from both the networks is assigned as the final class label.

TABLE IV: COMPARISON WITH TURING MODEL: ACCURACY, MACRO AVERAGE F1-SCORE, AND F1-SCORE OF INDIVIDUAL CLASSES

	Accuracy	**Average Macro F1-Score**	**S**	**D**	**Q**	**C**
Turing	78.4	0.43	0.40	0.00	0.46	0.87
Proposed Model	77.6	**0.55**	**0.49**	**0.27**	**0.57**	0.86

VI. Result

TABLE V: PRECISION AND RECALL VALUES OF EACH CLASS

	S	**D**	**Q**	**C**
Precision	0.52	1.00	0.57	0.83
Recall	0.46	0.15	0.58	0.90

The achievement of the proposed model on the testing dataset is displayed in Table IV. As the dataset is skewed, the average macro F1-score and per class F1-score are presented in the result table. Macro average F1-score shows how the model has performed overall across various classes present within the dataset. The model discussed in this paper has slightly less accuracy than Turing which has set the state of the art for the Rumour Stance Classification problem but the average macro F1-score of the proposed model is more than the Turing. The model

also achieves better results in predicting the support, deny, and query classes. Table V shows the per class precision and recall values.

TABLE VI: CONFUSION MATRIX FOR TESTING SET

	S	**C**	**Q**	**D**
S	43	48	3	0
C	37	700	41	0
Q	1	44	61	0
D	2	56	2	11

The proposed model anticipates the comment class as the majority class. Table VI demonstrates the confusion matrix for the model. Although, the performance of the model in picking out the deny class is not as good as compared to the other classes. Most of the deny class data are wrongly predicted and are labelled as comment. This can be improved with the addition of more data labelled as deny. However, the majority of 'query' class data got the right prediction label and the Multilayer Perceptron Network plays a crucial role in this.

CNN baseline models are tested on the same dataset to demonstrate the comparison between the Capsule Network and CNN base models. The first CNN model is a parallel architecture which includes three parallel convolutional layers with max pooling followed by a fully connected dense layer. The architecture of this model is shown in Figure 4.

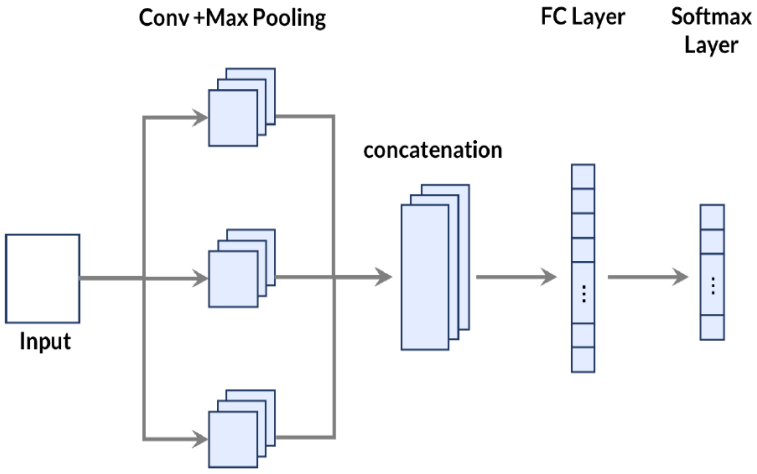


FIGURE 4: BASELINE CNN MODEL WITH PARALLEL CONVOLUTIONAL LAYERS

An alternative model without any parallel layers is also considered for the comparison. This model architecture contains only one convolutional layer with max pooling followed by a fully connected dense layer with ReLU activation function and a softmax layer for the final class prediction values. Both the CNN models are used along with Multi-Layer Perceptron Neural Network with the same feature set as used in the proposed model. Additionally, the same rules are applied to the hybrid of CNN and Multi-Layer Perceptron to merge the output of both the neural network. Table VII demonstrates the result achieved by the CNN models and compares it with the proposed model.

TABLE VII: COMPARISON WITH CNN MODELS: ACCURACY AND MACRO AVERAGE F1-SCORE

	Accuracy	**Macro Average F1- Score**
CNN without parallel layers	68.4	0.26

CNN with parallel layers	50.5	0.21
Capsule Network	72.5	0.27
CNN without parallel layers with MLP	76.9	0.51
CNN with Parallel Layers with MLP	73.8	0.51
Proposed Model	**77.6**	**0.55**

CNN model without parallel layers outperforms the CNN model with parallel layers in terms of both accuracy and Macro-Average F1-Score. A parallel layer architecture makes it complex to implement and does not provide better result. Therefore, making a simple CNN architecture is a better choice. On the other hand, Capsule Network performance surpasses both the CNN models. Additionally, a hybrid of Capsule Network and Multi-Layer Perceptron provides better accuracy and F1-score than a hybrid of CNN and Multi-Layer Perceptron given the same set of features.

VII. Conclusion

This paper presents a technique to address the matter of Rumour Stance Classification by exploiting Capsule Network and Multilayer Perceptron Network. For the training of Multilayer Perceptron Network, we studied the impact of various features extracted from the dataset and some of the selected features proved to be effective, chiefly for the 'query' class. The final class label is assigned after combining the class probabilities of both the networks in a way that utilizes each network's strength. To improve the performance of the 'deny' class, few rules were selected after the careful examination of the training and development dataset and applied before merging the output from the two networks. For future work, more data can be added for the weaker classes such as the 'deny' class and some new features can be explored that focus on distinguishing the class better.

References

[1] Elena Kochkina, Maria Liakata, Isabelle Augenstein,, Turing at SemEval-2017 Task 8: Sequential Approach to Rumour Stance Classification with Branch-LSTM, Association for Computational Linguistics, Proceedings of the 11th International Workshop on Semantic Evaluation (2017) 475-480.

[2] A. Zubiaga, M. Liakata, R. Procter, G. W. S. Hoi, and P. Tolmie, Analysing how people orient to and spread rumours in social media by looking at conversational threads, arXiv:1511.07487v3 [cs.SI] (2016)

[3] Kaizhou Xuan, Rui Xia, Rumour Stance Classification via Machine Learning with Text, User and Propagation Features, International Conference on Data Mining Workshops (ICDMW) (2019)

[4] G. W. Allport and L. Postman. 1947, The psychology of rumor, Journal of Clinical Psychology (1947).

[5] Michal Lukasik, Kalina Bontcheva, Trevor Cohn, Arkaitz Zubiaga, Maria Liakata, Rob Procter, Using Gaussian Processes for Rumour Stance Classification in Social Media, arXiv: 1609.01962v1 [cs.CL] (2016).

[6] Bernard Guerin and Yoshihiko Miyazaki, Analyzing Rumors, Gossip, and Urban Legends Through Their Conversational Properties, The Psychological Record (v56) (2006).

[7] Ralph L. Rosnow, The psychology of rumor, American Psychologist, Vol 46(5) (1991), 484–496.

[8] T. Mikolov, K. Chen, G. Corrado, and J. Dean, Efficient estimation of word representations in vector space, arXiv preprint arXiv: 1301.3781, 2013

[9] Vahed Qazvinian, Emily Rosengren, Dragomir R. Radev, Qiaozhu Mei, Rumor has it: Identifying Misinformation in Microblogs, Association for Computational Linguistics, Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (D11-1147) (2011), 1589-1599.

[10] Xiaomo Lie, Armineh Nourbakhsh, Quanzhi Li, Rui Fang, Real-time Rumor Debunking on Twitter, The 24th ACM International Conference on Information and Knowledge Management (CIKM) (2015).

[11] Sardar Hamidian and Mona T Diab, Rumor Identification and Belief Investigation on Twitter, Association for Computational Linguistics, Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (W16-0403) (2016), 3-8.

[12] Michal Lukasik, P.K. Srijith, Duy Vu, Hawkes Processes for Continuous Time Sequence Classification: an Application to Rumour Stance Classification in Twitter, Association for Computational Linguistics, Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Pages) (2016), 393-398

[13] Arkaitz Zubiaga, Elena Kochkina, Mariaw Liakata, Rob Procter, Michal Lukasik, Stance Classification in Rumours as a Sequential Task Exploiting the Tree Structure of Social Media Conversations, The COLING 2016 Organizing Committee, Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers (C16-1230) (2016), 2438-2448.

[14] Zubiaga, Kochkina, Lukasik, Discourse-aware rumour stance classification in social media using sequential classifiers, arXvi: 1712.02223v1 [cs.CL] (2017).

[15] Marianela Garcia Lozano, Hanna Lilja, Edward and Maja Karasalo, Mama Edha at SemEval-2017 Task8: Stance classification with CNN and Rules, Association for Computational Linguistics, Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017) (2017), 481-485.

[16] Jing Ma, Wei Gao, Kam-Fai Wong, Detect Rumor Stance Jointly by Neural Network Multi-task Learning, WWW '18: Companion Proceedings of the The Web Conference (2018), 585-593.

[17] Martin Fajcik, Lukas Burget, Pavel Smrz, BUT-FIT at SemEval-2019 Task 7: Determing the Rumour Stance with Pre-Trained Deep Bidirectional Transformers, Association for Computational Linguistics, Proceedings of the 13th International Workshop on Semantic Evaluation (S19-2192) (2019), 1097-1104.

[18] Na Bai, Zhixiao Wang, Farrong Meng, A Stochastic Attention CNN Model for Rumor Stance Classification, IEEE Access (2020).

[19] Sara sabour, Nicholas Frost, Geoffrey E. Hinton, Dynamic Routing Between Capsules, 31st Conference on Neural Information Processing Systems (NIPS 2017), (2017).

[20] Mohammad Hadi Goldani, Saeedeh Momtazi, Reza Safabakhsh, Detecting Fake News with Capsule Neural Networks, arXvi: 2002:01030v1 [cs.CL] (2020).

[21] Wei Zhao, Jianbo Ye, Min Yang, Zeyang Lei, Soufei Zhang, Zhou Zhao, Investigating Capsule Networks with Dynamic Routing for Text Classification, Association for Computational Linguistics, Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (D18-1350) (2018), 3110-3119.

[22] Kim, J., Jang, S., Park, E., & Choi, S., Text classication using capsules, Neurocomputing, (Volume 376) (2020), 214-221.

[23] Jingjing Gong, Xipeng Qiu, Shaojing Wang, Xuanjing Huang, Information Aggregation via Dynamic Routing for Sequence Encoding, Association for Computational Linguitics, Proceedings of the 27th International Conference on Computational Linguistics (C18-1232) (2018), 2742-2752.

[24] Saurabh Srivastava, Prerna Khurana, Vartika Tewari, Identifying aggression and toxicity in comments using capsule network, Association for Computational Linguistics, Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018) (2018), 98-105.

[25] L. Derczynski, K. Bontcheva, M. Liakata, R. Procter, G. W. S. Hoi, and A. Zubiaga, Semeval-2017 task 8: Rumoureval: Determining rumour veracity and support for rumours, Association for Computer Linguistics, Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017) (S17-2006) (2017), 69-76.

International Journal of Mathematical, Engineering and Management Sciences
Vol. 6, No. 1, 207-222, 2021
https://doi.org/10.33889/IJMEMS.2021.6.1.013

SDE based Unified Scheme for Developing Entropy Prediction Models for OSS

Deepika
Department of Operational Research,
University of Delhi, Delhi, India.
E-mail: deepika.sre@gmail.com

Ompal Singh
Department of Operational Research,
University of Delhi, Delhi, India.
E-mail: drompalsingh1@gmail.com

Adarsh Anand
Department of Operational Research,
University of Delhi, Delhi, India.
Corresponding author: adarsh.anand86@gmail.com

Jagvinder Singh
University School of Management and Entrepreneurship,
Delhi Technological University, Delhi, India.
E-mail: jagvinder.singh@dtu.ac.in

(Received January 13, 2020; Accepted April 30, 2020)

Abstract
Today, so as to meet the user's requirement, modification of software is necessarily required. But at the same time, to incorporate these modifications and requirements there are enormous changes which are made to the coding of the software and over a period of time these changes make the software complex. Largely there are three types of code changes occur in the source code namely, bug repair, feature enhancement & addition of new features, but these changes bring the uncertainty in the bug removal rate. In this paper, these uncertainties have been explicitly modeled and using three-dimensional wiener processes that define the three types of fluctuation; we have come up with an entropy prediction modeling framework with a unified approach. The analytical solution of the equation is interpreted using Itô's process. The models are fitted on three real life projects namely Avro, Hive and Pig of Apache open source software (OSS) The experimental findings show that present models exhibit accurate estimation results and have strong prediction skills.

Keywords- Complexity of code changes (CCC), Differential equation (DE), Distribution, Entropy, Fluctuations, Open source software (OSS), Unification.

1. Introduction

Software's role in smooth functioning and productivity enhancement of almost every sector cannot be denied. "The advancements in computing capabilities of devices have proved to be one of the significant reasons for the introduction of software to time critical systems. OSS has brought major breakthroughs in development trends for large scale software systems. For such software, the prototype is developed and is evolved with voluntary contribution across the globe. The development pattern followed by OSS doesn't rigidly comply with standards described under the traditional software engineering lifecycle (Bhatt et al., 2017). With the enormous rise in

International Journal of Mathematical, Engineering and Management Sciences
Vol. 6, No. 1, 207-222, 2021
https://doi.org/10.33889/IJMEMS.2021.6.1.013

software demand, IT industry has become highly competitive. Quality is the most critical characteristic that ensures the sustainability of software in market and reliability is an important metric to quantify the system's quality". In past decades numerous researches have been carried out in the direction of reliability modeling and assessment (Kapur et al., 2011a).

Open Source Software (OSS) varies characteristically from closed source software. Features such as irregular volunteer participation, lack of hierarchical management, no delivery schedules, and lack of rigid SDLC principles etc. make it different from closed source systems. OSS development is initiated with an idea or concept and a working prototype is worked out by few developers or a small team. "The core system with limited functionality is then released over the internet along with source code for further enhancement and refinement with global volunteer participation. The most distinguishing feature over closed source software is that this is released with very little functionality testing and therefore it is the user population who adopts such systems become the deciding factor in reliability growth phenomenon for such software. The OSS keeps on evolving due to the contributions from volunteers across the world. The software is developed by the core team with few members and released over the internet".

"The OSS community generally follows a layered architecture which describes the users into various categories like core members, active developers, peripheral developers, bug fixer, bug reporter, passive users etc. from the core to the periphery. The users are assigned participation levels which may change with time and contributions. People start working on software and when they find some glitch in software workflow, they report it as a failure and the fault is identified". After fault identification developers try to remove the fault.

Software source codes are enhanced to satisfy the user's needs. "The huge changes in the code make the software complex over a period of time. These changes increase the complexity of the code which also leads to the introduction of the bugs. The changes are being made continuously in the software to remove the bugs, enhance the features and implement the new functionalities (Singh and Chaturvedi, 2013; Deepika et al., 2017). The open source software is being built through the contributors from diverse communities and keeps on moving rapidly. The code complexity is one of the major attributes which determines the quality and reliability of the software. Many contributors are making changes in the source code of the software to produce the quality of the software in a limited time. It is getting difficult to remember all those changes committed in the code which makes the source code complex. The contributors are interacting with each other through the discussion forum. Three types of code changes occur in the source code namely, bug repair, features enhancement and addition of new features. Bugs are generated in the software mainly due to miscommunication or no communication among active users, frequently changing requirements, early release pressures, the occurrence of programming errors, increase software complexity".

The remainder of the study is structured into the following sections: In Section 2, a detailed review of past literature is presented. Section 3 describes the modeling framework and also notations and assumptions are used for formulating the models. Following that, in Section 4, the parameter estimation and comparison criteria results of the proposed models are produced. Finally, in the next section, we summarize the key results and provide the concluding remarks which are then followed by the references.

International Journal of Mathematical, Engineering and Management Sciences
Vol. 6, No. 1, 207-222, 2021
https://doi.org/10.33889/IJMEMS.2021.6.1.013

2. Literature Review

Numerous studies have been carried out for assessing the reliability of OSS systems. Tamura and Yamada (2008) investigated reliability assessment combining it with neural networks. Some of the latest SRGMs proposed to assess the reliability growth of software systems incorporating practical phenomena are Chatterjee and Shukla (2016), Li and Pham (2017), Zhu and Pham (2017). Stochastic differential equation-based models have been proposed for OSS and further discussed its application in the optimal version update problem by Tamura et al. (2008). Li et al. (2011) presented another model to predict optimal version update time for OSS. Yang et al. (2016) included the process of fault detection and correction and performed reliability modeling for multi-release OSS. Many other studies tried to model reliability for OSS. Rahmani et al. (2009, 2010) and Zhou and Davis (2005) studied the OSS reliability model by experiments and also performed the comparative analysis.

The frequent change in the source code of open source software makes the software complex and more error-prone. "The CCC has been quantified and design using an information theory-based procedure called entropy. In literature, researchers have worked on entropy and it is widely being exercised in software reliability. It was Shannon (1948, 1951) who provided the foundation for the area of information theory and discussed measures for the efficiency of the communication channel and entropy. The CCC matrix is also used for the maintenance of software (Kafura and Reddy, 1987). In the earlier stage, an article was written by Fenton and Neil in 1999 where the detection rate was used to predict defects with an objective (Fenton and Neil, 1999). After Fenton and Neil (1999), Graves et al. (2000) gave the work related to product measures (Graves et al., 2000). Some researchers proposed in the literature to resolve the fault potential. (Khoshgoftaar et al., 1999; Leazak et al., 2002; Arisholm and Briand, 2006). Then some researchers have discussed the bugs prediction of various projects with the absence of single metrics (Nagappan and Ball, 2005). In 2008 Moser has come up with a comparative study of change and code metrics (Moser et al., 2008). Some researchers gave different approaches to compare the history of complexity metric for bug prediction". Then other researchers have given the comparison of complexity metric and product measures (Hassan, 2009; Kamei et al., 2010).

Later on, D'Ambros have come up with a study of decay based models (D'Ambros et al., 2010, 2012). Then a model was suggested by Singh and Chaturvedi (2012) on the basis of entropy for bug prediction utilizing support vector regression. Further, they have proposed a mathematical model to study the diffusion of the CCC. The release time of software has been predicted by Chaturvedi et al. (2013) using the CCC. Further, the researchers have extended their work and discussed about the prediction of subsequently year expected bugs based on the current year CCC (Singh and Chaturvedi, 2013).

Along with several other measures of entropy, Arora et al. (2014) considered bug prediction models using different types of measures. Chaturvedi et al. (2014) used measures of entropy in the prediction of CCC. A defect prediction model was studied to predict how many defects would arise throughout the improvement of software lines on the basis of entropy (Jeon et al., 2014). Singh et al. (2015) proposed three models- software reliability growth models, models based on CCC to predict bugs in software. Recently Anand et al. have characterized the CCC into two aspects: features improvement and a bug fix for OSS (Anand et al., 2019a).

In the present paper, we have come up with an entropy prediction modeling framework using various types of uncertainties with the help of a three-dimensional wiener process. This process

International Journal of Mathematical, Engineering and Management Sciences
Vol. 6, No. 1, 207-222, 2021
https://doi.org/10.33889/IJMEMS.2021.6.1.013

defines the three types of fluctuations namely; bug fix, improvement and addition of new characteristics. The proposed framework is capable of handling the distribution function and is thus an important step towards the unification of SDE based models, which depends on specific distribution functions.

3. Model Development

In this section, the emphasis is laid on describing the proposed entropy-based models. Section 3.1 lists the set of notations and assumptions that have been used.

3.1 Set of Notations

a	Potential CCC to be diffused over a period of time
$b(t)$	Rate at which CCC is diffused in the code
$H(t)$	CCC at any given time.
$E(H(t)) = G(t)$	Expected number of CCC
s_1	Positive constant irregular fluctuations due to bug fixes
s_2	Positive constant irregular fluctuations due to new features
s_3	Positive constant irregular fluctuations due to features enhancement
$\lambda(t)$	Standardized Gaussian white noise
k	Shape parameter for Weibull Distribution
μ	Mean parameter for Normal Distribution
σ	Standard Deviation parameter for Normal Distribution
$F(t)$	Cumulative Distribution Function
$f(t)$	Probability Density Function

3.2 Assumptions

(a) Potential Entropy is constant.
(b) At original time $t=0$, there is no change in file and CCC is zero.

Using the above set of assumptions to capture the uncertainty, we propose the unified models based on CCC to be diffused in software over a period of time. Stochastic differential equations have an inbuilt tendency to cater to irregular fluctuations that can be represented as an ordinary differential equation. This equation describes the Brownian motion, which can be solved by using the $\hat{I}to$ integral. In line with what is available in software reliability literature, the following linear differential equation contributed to model the diffusion rate of CCC (Kapur et al., 2011b):

$$\frac{dH(t)}{dt} = k(t)(a - H(t)) \tag{1}$$

where, $k(t)$ is the relative diffusion rate of errors being detected/removed. Under certain situations, $k(t)$ might not be known in its exact sense; also it may happen that the fluctuations are because of some environmental factor. In that case, it is important to account for such uncertainty that can be modeled by considering associated "noise" term as follows (Singhal et al., 2019):

International Journal of Mathematical, Engineering and Management Sciences
Vol. 6, No. 1, 207-222, 2021
https://doi.org/10.33889/IJMEMS.2021.6.1.013

$$\left.\begin{aligned} k(t) &= \frac{f(t)}{1-F(t)} + noise \\ &= h(t) + noise \end{aligned}\right\} \tag{2}$$

where, $h(t)$ is the hazard rate or random external factor. The exact behaviour of noise is difficult to understand and only the distribution function that it might follow can be identified in advance, the function $h(t) = \frac{f(t)}{1-F(t)}$ is time dependent and non-random in nature.

$$k(t) = \frac{f(t)}{1-F(t)} + s\lambda(t) \tag{3}$$

where, $\lambda(t)$ is the factor that portrays the Gaussian white noise and s accounts for the magnitude of irregular fluctuations, that is, $dZ(t)/dt = \lambda(t)$. Before we proceed, it is important to understand the basic definition and properties of the Wiener process (Brownian motion), which are as follows:

"The Wiener process $Z(t)$ is in essence a series of normally distributed random variables and some time points". "The variances of these normally distributed random variables increase to reflect that it is more uncertain to predict the value of the process after a longer period of time".

The properties of the Wiener process $\{ Z(t) \}$ for $t \geq 0$

- $Z(t) - Z(q) \sim N(0, t-s)$ {Normal increments}.
- $Z(t) - Z(q)$ and $Z(u)$ are independent, for $u \leq q < t$ {Independence of increments}.
- $Z(t)$ is a continuous function of t.

Eq^{n}. (3) when substitute in Eq^{n}. (1) results in uncertainty differential equation as follows:

$$\frac{dH(t)}{dt} = \left(\frac{f(t)}{1-F(t)} + s\lambda(t) \right)(a - H(t)) \tag{4}$$

Eq^{n}. (4) can be converted to the subsequent Ito's type DE (based on Taylor series) (Yamada et al., 2003; Oksendal, 2013):

$$dH(t) = \left(\frac{f(t)}{1-F(t)} - \frac{1}{2}s^2\right)(a - H(t))dt + s(a - H(t))dZ(t) \tag{5}$$

Further, solve the above DE with integration

$$\int dH(t) = \int \left(\frac{f(t)}{1-F(t)} - \frac{1}{2}s^2\right)(a - H(t))dt + s(a - H(t))dZ(t) \tag{6}$$

International Journal of Mathematical, Engineering and Management Sciences
Vol. 6, No. 1, 207-222, 2021
https://doi.org/10.33889/IJMEMS.2021.6.1.013

Random variable *H(t)* is presumed to be continuous and its expected value is given as *E(H(t))=G(t)*. Now taking expectation on equal sides

$$\int dG(t) = \int (\frac{f(t)}{1-F(t)} - \frac{1}{2}s^2)(a-G(t))dt + E\left[\int [s(a-H(t)dZ(t)]\right] \quad (7)$$

Using the ideology from Anand et al. (2018a, 2018b, 2019a, 2019b) of Ito Integral, the next component of the Eq^{n}. (7) has vanished, that is

$$E\left[\int s(a-H(t)dZ(t)\right] = 0$$

"which implies that the non-anticipating function will be statistically independent in the future time or mathematically we can say that if we take the expected value of any non-anticipating function then it vanishes the whole component" i.e.

$$E\left[\int_0^T J(t)dZ(t)\right] = 0, \text{ where, } J(t) = s(a-H(t)).$$

Therefore DE (7) can be modeled as:

$$G(t) = E(H(t)) = a\left[1-\exp\{-\int_0^t \frac{f(n)}{1-F(n)}dn - sZ(t)\}\right] \quad (8)$$

This is the distribution based mean value function using SDE. The complexity of the software system influenced the developer to deviate the testing phenomenon from the NHPP based modeling behaviour to a more realistic environment that accounts for various uncertain factors during the testing process. "The behaviour of the testing process, that is, majorly uncertain in nature is influenced by various factors like, testing efforts, testing skills, efficiency and methods. In order to capture these uncertain aspects regulating the testing process, various researchers have considered one-dimensional wiener process but changes in software source codes are unavoidable and the source code of the software is frequently modified to meet the user's huge requirements (Kapur et al., 2011a). Code change process means to study the patterns of source code Modifications. These changes are occurring due to bug repair (BR), features enhancement (FE) and addition of new features (NF)". In present work, these uncertainties have been represented by S_1 (BR), S_2 (FE) and S_3 (NF) respectively (Singh and Sharma, 2014).

Bug repair	BR is the changes or modifications made in the code due to fixing of a bug.
Features enhancement	FE is the changes made due to some cosmetic changes like formatting, alignment, justification, comments etc.
Addition of new features	NF is the changes made in the code due incorporation of new features or new components.

International Journal of Mathematical, Engineering and Management Sciences
Vol. 6, No. 1, 207-222, 2021
https://doi.org/10.33889/IJMEMS.2021.6.1.013

Hence, Eq^{n}. (9) can be modeled that reflect the three-dimensional Wiener process with three types of fluctuations; we get the mathematical structure as (Tamura and Yamada, 2014):

$$\left.\begin{aligned} dH(t) = \{&\frac{f(t)}{1-F(t)} - \frac{1}{2}(s_1^2+s_2^2+s_3^2)(a-H(t)\}dt \\ &+ s_1\{a-H(t)\}dZ_1(t)) \\ &+ s_2\{a-H(t)\}dZ_2(t)) \\ &+ s_3(a-H(t)\}dZ_3(t)) \end{aligned}\right\} \tag{9}$$

The solution of Eq^{n}. (9), under the seed value $H(0)=0$ as follow:

$$G(t) = E(H(t)) = a[1-\exp\{-\int_0^t \frac{f(n)}{1-F(n)}dn - s_1Z_1(t) - s_2Z_2(t) - s_3Z_3(t)\}] \tag{10}$$

$$G(t) = a\left[1-(1-F(t))\,e^{\frac{1}{2}(s_1^2+s_2^2+s_3^2)t}\right] \tag{11}$$

Now, using the solution process $G(t)$ in Eq^{n}. (10), several entropy predictions models can be derived for different distributions. It is to note that as per our knowledge, this is the first attempt to model the complexity of code changes with irregularity with a unified approach. There have been proposals in the past that have studied about these two aspects separately and in different fields, but in unification; they have been studied for the first time (which is obtained by unified approach)

3.3 Proposed Model I

In this proposed model, it is assumed the distribution of diffusion for entropy follows Weibull distribution. It is much used in reliability engineering. Due to its versatile nature, it can take the characteristics of other types of distributions. Thus, we can say that generalization of the exponential distribution is Weibull distribution because of its flexible nature. Hence, in expression (11), *F(t)* follows the Weibull distribution

$$F(t) \sim (1-e^{-bt^{\gamma}}) \quad ;\gamma > 0$$

where γ is shape parameter (or slope) and *b* is the rate at which the complexity of code changes is diffused in the code. Now putting the value of *F(t)* in Eq^{n}. (11), thus the proposed model for entropy prediction is given as

$$G_1(t) = E(H_1(t)) = a(1-\exp(-bt^{\gamma} + \frac{1}{2}(s_1^2+s_2^2+s_3^2)t)) \tag{12}$$

The preceding equation represents the expected number of CCC at any given time *t*.

International Journal of Mathematical, Engineering and Management Sciences
Vol. 6, No. 1, 207-222, 2021
https://doi.org/10.33889/IJMEMS.2021.6.1.013

3.4 Proposed Model II

It is considered that *F(t)* follows the Rayleigh distribution in the given Eq^{n}. (11). This distribution is the measure of a two-dimensional random vector whose coordinates are distributed identically.

$$F(t) \sim (1-e^{-bt^2}).$$

Now substituting the value of *F(t)* in Eq^{n}. (11), so the entropy prediction can be modeled as

$$G_2(t) = E(H_2(t)) = a(1-\exp(-bt^2 + \frac{1}{2}(s_1^2 + s_2^2 + s_3^2)t)) \quad (13)$$

The above Eq^{n}. (13) describes the amount of entropy prediction based on the unified approach.

3.5 Proposed Model III

In proposed model III, we have assumed that the cumulative distribution function follows the normal distribution. The normal distribution is a type of continuous probability distribution for a real value random variable. It is a bell-shaped curve and also called Gaussian distribution. The general form of its probability density function

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}.$$

The cumulative distribution function (CDF) of the normal distribution, usually denoted with the Greek letter ϕ, is the integral

$$\phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx.$$

The related error function *erf (t)* gives the probability of a random variable with the normal distribution of mean 0 and variance ½ falling in the range [-*t*,*t*]; that is

$$erf(t) = \frac{2}{\sqrt{\pi}} \int_{0}^{t} e^{-x^2} dx.$$

The above two functions closely related, namely

$$\phi(t) = \frac{1}{2}\left[1 + erf\left(\frac{t}{\sqrt{2}}\right)\right].$$

For a generic normal distribution with density *f*, mean μ and deviation σ, the cumulative distribution function is

International Journal of Mathematical, Engineering and Management Sciences
Vol. 6, No. 1, 207-222, 2021
https://doi.org/10.33889/IJMEMS.2021.6.1.013

$$F(t)=\phi\left(\frac{t-\mu}{\sigma}\right)=\frac{1}{2}\left(1+erf\left(\frac{t-\mu}{\sigma\sqrt{2}}\right)\right).$$

The parameter μ is the mean or expectation of the distribution (and also its median and mode); and σ is its standard deviation. A variance of the distribution is σ^2.

Now substitute the value of *F(t)* in Eq^{n}. (11), we get,

$$G_3(t)=E(H_3(t))=a\left[1-(1-\frac{1}{2}\{1+erf(\frac{t-\mu}{\sigma\sqrt{2}})\}\mathrm{e}^{\frac{1}{2}(s_1^2+s_2^2+s_3^2)t}\right] \quad (14)$$

In the above Eq^{n}. (14), we have described the expected number of entropy predictions for normal distribution.

4. Result and Discussion

4.1 Data Description and Analysis

For validation of the above proposal, we have carried out the data analysis on Apache open source software data sets. Data Set- 1(Abbreviated DS-1) consists of 62.68 entropy that have been predicted in 17 months from Avro Project. Similarly, Data Set-2 (Abbreviated DS-2) and Data Set-3 (Abbreviated DS-3) consist of 68.58 entropy and 56.79 entropy that has been predicted in 18 months and 15 months from Hive and Pig projects respectively. All three data sets can be seen in Table 1. The parameters of the models have been estimated using Statistical Analytical Software (SAS)/ETS user's guide 9.1 (2004) based on the nonlinear regression method. "The performance of models is judged by their capability to fit the data (goodness of fit) and predicting the future performance of the entropy (as shown in Figures 1, 2 and 3). The parameter estimation and comparison criteria result for DS-1, DS-2 and DS-3 of all models under consideration can be viewed through Table 2, Table 3 and Table 4 respectively.".

Table 1. Monthly data of apache open source software.

Avro Data Set		Hive Data Set		Pig Data Set	
Time (Monthly)	Non-Cumulative Entropy	Time (Monthly)	Non-Cumulative Entropy	Time (Monthly)	Non-Cumulative Entropy
1	0.988798	1	3.918735	1	1.9815
2	3.877407	2	3.876313	2	3.937522
3	3.855519	3	3.85041	3	3.88785
4	3.823077	4	3.853447	4	3.918368
5	3.858738	5	3.892308	5	3.901588
6	3.939776	6	3.882447	6	3.886225
7	3.896395	7	3.84032	7	3.876901
8	3.940347	8	3.810116	8	3.90468
9	3.941543	9	3.872064	9	3.941262
10	3.963296	10	3.885398	10	3.927332
11	3.926444	11	3.84444	11	3.921351
12	3.948949	12	3.874096	12	3.941777
13	3.923868	13	3.814615	13	3.912891
14	3.915305	14	3.897128	14	3.936509
15	3.97836	15	3.878384	15	3.923166
16	3.906676	16	3.843008		
17	3	17	3.843319		
		18	2.910245		

Source: DS-I, DS-II and DS-III: (Anand et al. 2019a).

International Journal of Mathematical, Engineering and Management Sciences
Vol. 6, No. 1, 207-222, 2021
https://doi.org/10.33889/IJMEMS.2021.6.1.013

Table 2. Estimation of model parameters for DS -1.

Parameter	Model-I	Model-II	Model-III
a	96.43	64.90	78.98
b	0.018	0.0087	-
γ	1.427	-	-
s_1	0.0032	0.0021	0.030
s_2	0.00052	0.0012	0.215
s_3	0.0042	0.001	0.109
μ	-	-	15.90
σ	-	-	11.29

Table 3. Estimation of model parameters for DS –II.

Parameter	Model-I	Model-II	Model-III
a	79.06	67.46	86.43
b	0.024	0.0093	-
γ	1.470	-	-
s_1	0.012	0.011	0.0031
s_2	0.0032	0.0033	0.113
s_3	0.00013	0.00012	0.110
μ	-	-	13.42
σ	-	-	9.98

Table 4. Estimation of model parameters for DS –III.

Parameter	Model-I	Model-II	Model-III
a	87.02	57.71	61.23
b	0.024	0.0146	-
γ	1.39	-	-
s_1	0.0112	0.011	0.055
s_2	0.0032	0.0032	0.215
s_3	0.0124	0.0124	0.109
μ	-	-	11.88
σ	-	-	9.01

4.2 Comparison Criteria for Proposed Models

The following table gives the results of comparison criteria for developed models which are calculated by Sum of Square errors, Mean Squared Error, Root Mean Square Errors and Coefficient of Determination.

International Journal of Mathematical, Engineering and Management Sciences
Vol. 6, No. 1, 207-222, 2021
https://doi.org/10.33889/IJMEMS.2021.6.1.013

Table 5. Comparison criteria for all the proposed models.

Model →	Model-I	Model-II	Model-III
DS-1			
SSE	6.592	96.49	241.1
MSE	0.439	6.433	18.54
Root MSE	0.662	2.536	4.306
R-Square	0.99	0.98	0.96
DS-2			
SSE	67.083	240.7	124.1
MSE	4.19	15.04	8.274
Root MSE	2.047	3.885	2.87
R-Square	0.99	0.97	0.98
DS-3			
SSE	7.647	91.08	117.4
MSE	0.588	7.006	10.67
Root MSE	0.767	2.647	3.266
R-Square	0.99	0.98	0.97

On the basis of analysis performed in above Table 5, it can be clearly seen that the values of comparison criteria are quite significant, that is the obtained values are in acceptable range as the value of *R*-square obtained in approaching 1 which means the proposed Model-I are able to cater the entropy prediction in the good sense. Figures 1-3 also provide a way to graphically analyze the predictive capability of models which show that computed values are close correspondence to open source software data.

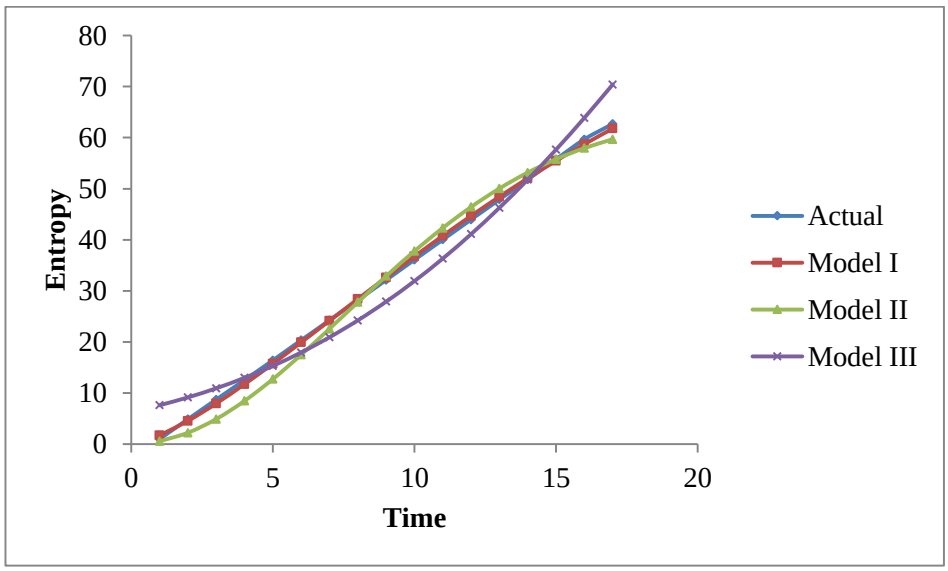


Figure 1. Goodness fit curve for Data Set-I.

International Journal of Mathematical, Engineering and Management Sciences
Vol. 6, No. 1, 207-222, 2021
https://doi.org/10.33889/IJMEMS.2021.6.1.013

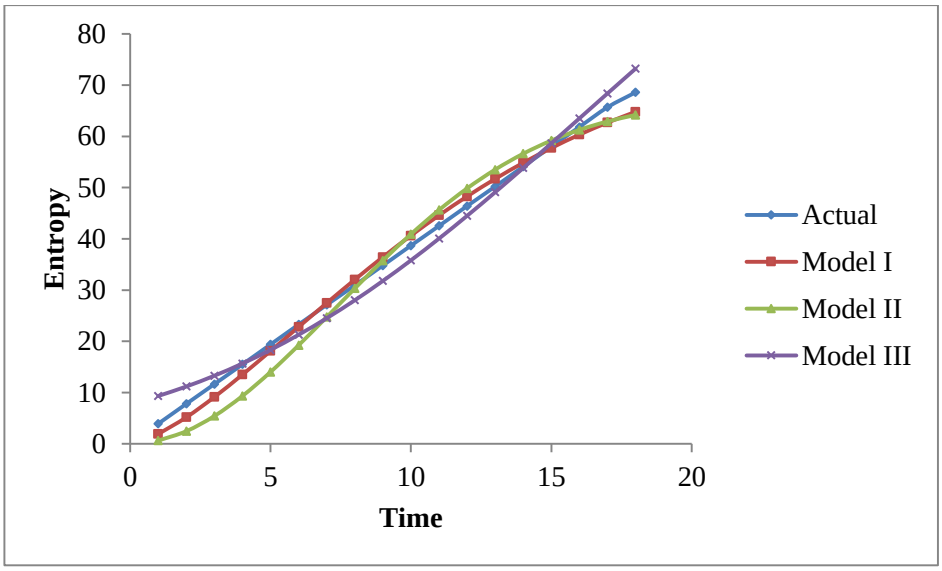


Figure 2. Goodness fit curve for Data Set-II.

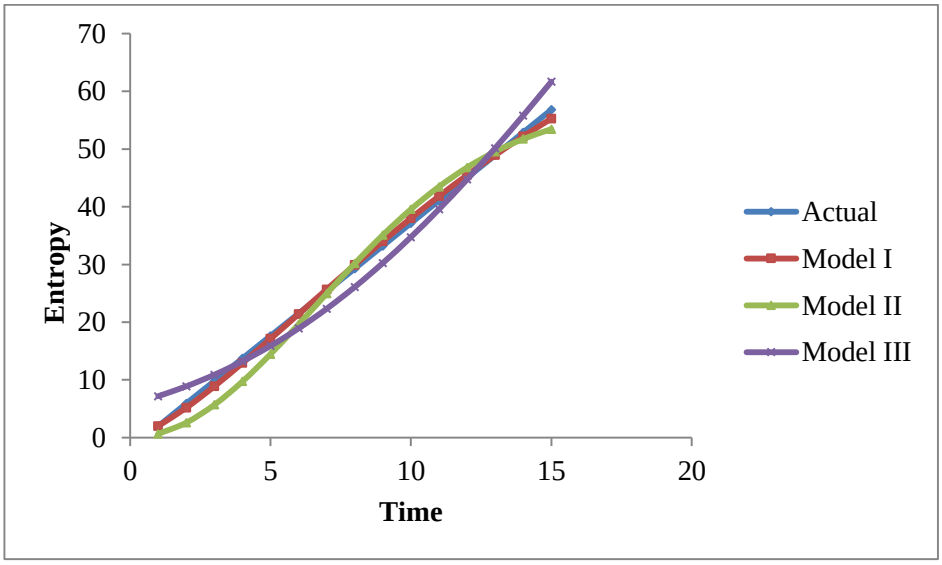


Figure 3. Goodness fit curve for Data Set-III.

5. Conclusion

In the present work, a generalized framework for deriving several entropy prediction models based on SDE of Ito's type has been presented. Three basic reasons that; bug fixing, features improvement and new feature introduction have been studied for the diffusion process which can bring out changes in the code structure. The proposed framework is capable of handling any general distribution function and takes care of three-dimensional Wiener processes for inculcating

the irregularity. The proposed set of models has been validated on three open source data sets and Model- I turns out to be the best fit in all data sets under consideration.

Conflict of Interest
The authors confirm that there is no conflict of interest to declare for the publication.

Acknowledgment
The authors would like to thank the editor in chief and reviewers for giving valuable inputs in improvising the overall quality of the paper.

References

Anand, A., Bharmoria, S., & Ram, M. (2019a). Characterizing the complexity of code changes in open source software. In: Anand, A., Ram, M. (eds) *Recent Advancements in Software Reliability Assurance*. pp. 1-14, CRC Press- A Taylor & France Group, Boca Raton.

Anand, A., Deepika, & Singh, O. (2019b). Formulation of error generation-based srgms under the influence of irregular fluctuations. In: Kapur, P., Klochkov, Y., Verma, A., Singh, G. (eds) *System Performance and Management Analytics*. Asset Analytics (Performance and Safety Management). pp. 103-117, Springer, Singapore.

Anand, A., Deepika, Verma, A.K., & Ram, M. (2018a). Revisiting error generation and stochastic differential equation-based software reliability growth models. In: Anand, A., Ram, M. (eds) *System Reliability Management: Solutions and Technologies*. pp. 65-78, CRC Press- A Taylor & France Group, Boca Raton.

Anand, A., Singhal, S., & Singh, O. (2018b). Revisiting dynamic potential adopter diffusion models under the influence of irregular fluctuations in adoption rate. In *Handbook of Research on Promoting Business Process Improvement Through Inventory Control Techniques*. pp. 499-519, IGI Global, USA.

Arisholm, E., & Briand, L.C. (2006, September). Predicting fault-prone components in a java legacy system. In *Proceedings of the 2006 ACM/IEEE International Symposium on Empirical Software Engineering* (pp. 8-17). Rio de Janeiro, Brazil. DOI: 10.1145/1159733.1159738.

Arora, H.D., Kumar, V., & Sahni, R. (2014, October). Study of bug prediction modeling using various entropy measures-a theoretical approach. In *Proceedings of 3rd International Conference on Reliability, Infocom Technologies and Optimization* (pp. 1-5). IEEE. Noida, India.

Bhatt, N., Anand, A., Yadavalli, V.S.S., & Kumar, V. (2017). Modeling and characterizing software vulnerabilities. *International Journal of Mathematical, Engineering and Management Sciences, 2*(4), 288-299.

Chatterjee, S., & Shukla, A. (2016). Modeling and analysis of software fault detection and correction process through Weibull-type fault reduction factor, change point and imperfect debugging. *Arabian Journal for Science and Engineering, 41*(12), 5009-5025.

Chaturvedi, K.K., Bedi, P., Misra, S., & Singh, V.B. (2013, December). An empirical validation of the complexity of code changes and bugs in predicting the release time of open source software. In *2013 IEEE 16th International Conference on Computational Science and Engineering* (pp. 1201-1206). IEEE. Sydney, Australia.

Chaturvedi, K.K., Kapur, P.K., Anand, S., & Singh, V.B. (2014). Predicting the complexity of code changes using entropy based measures. *International Journal of System Assurance Engineering and Management, 5*(2), 155-164.

D'Ambros, M., Lanza, M., & Robbes, R. (2012). Evaluating defect prediction approaches: a benchmark and an extensive comparison. *Empirical Software Engineering*, *17*(4-5), 531-577.

D'Ambros, M., Lanza, M., & Robbes, R. (2010, May). An extensive comparison of bug prediction approaches. In *2010 7th IEEE Working Conference on Mining Software Repositories (MSR 2010)* (pp. 31-41). IEEE. Cape Town, South Africa.

Deepika, Singh, O., Anand, A., & Singh, J.N.P. (2017). Testing domain dependent software reliability growth models. *International Journal of Mathematical, Engineering and Management Sciences*, *2*(3), 140-149.

Fenton, N.E., & Neil, M. (1999). A critique of software defect prediction models. *IEEE Transactions on Software Engineering*, *25*(5), 675-689.

Graves, T.L., Karr, A.F., Marron, J.S., & Siy, H. (2000). Predicting fault incidence using software change history. *IEEE Transactions on Software Engineering*, *26*(7), 653-661.

Hassan, A.E. (2009, May). Predicting faults using the complexity of code changes. In *2009 IEEE 31st International Conference on Software Engineering* (pp. 78-88). IEEE. Vancouver, Canada.

Jeon, C.K., Byun, C., Kim, N.H., & In, H. (2014). An entropy-based method for defect prediction in software product lines. *International Journal of Multimedia and Ubiquitous Engineering*, *9*(3), 375-377.

Kafura, D., & Reddy, G.R. (1987). The use of software complexity metrics in software maintenance. *IEEE Transactions on Software Engineering*, *13* (3), 335-343.

Kamei, Y., Matsumoto, S., Monden, A., Matsumoto, K.I., Adams, B., & Hassan, A.E. (2010, September). Revisiting common bug prediction findings using effort-aware models. In *2010 IEEE International Conference on Software Maintenance* (pp. 1-10). IEEE. Timisoara, Romania.

Kapur, P.K., Pham, H., Gupta, A., & Jha, P.C. (2011a). *Software reliability assessment with OR applications* (p. 364). Springer, London.

Kapur, P.K., Singh, O., & Singh, J. (2011b). Stochastic differential equation-based software reliability growth modeling with change point and two types of imperfect debugging. *In Proceedings of 5th National Conference on Computing for Nation Development,* (pp. 605-12). INDIACom, New Delhi.

Khoshgoftaar, T.M., Allen, E.B., Jones, W.D., & Hudepohl, J.P. (1999). Data mining for predictors of software quality. *International Journal of Software Engineering and Knowledge Engineering*, *9*(05), 547-563.

Leszak, M., Perry, D.E., & Stoll, D. (2002). Classification and evaluation of defects in a project retrospective. *Journal of Systems and Software*, *61*(3), 173-187.

Li, Q., & Pham, H. (2017). A testing-coverage software reliability model considering fault removal efficiency and error generation. *PloS one*, *12*(7), e0181524-e0181524.

Li, X., Li, Y.F., Xie, M., & Ng, S.H. (2011). Reliability analysis and optimal version-updating for open source software. *Information and Software Technology*, *53*(9), 929-936.

Moser, R., Pedrycz, W., & Succi, G. (2008, May). A comparative analysis of the efficiency of change metrics and static code attributes for defect prediction. In *Proceedings of the 30th International Conference on Software Engineering* (pp. 181-190). ACM. DOI: 10.1145/1368088.1368114.

Nagappan, N., & Ball, T. (2005, May). Use of relative code churn measures to predict system defect density. In *Proceedings of the 27th International Conference on Software Engineering* (pp. 284-292). ACM. Doi:10.1145/1062455.1062514.

Oksendal, B. (2013). *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media. London.

Rahmani, C., Azadmanesh, A.H., & Najjar, L. (2010). A Comparative analysis of open source software reliability. *Journal of Software*, *5*(12), 1384-1394.

Rahmani, C., Siy, H., & Azadmanesh, A. (2009). An experimental analysis of open source software reliability. *Department of Defense/Air Force Office of Scientific Research*.

SAS Institute Inc. (2004). SAS/ETS User's Guide Version 9.1. Cary, NC: SAS Institute Inc.

Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379-423.

Shannon, C.E. (1951). Prediction and entropy of printed English. *Bell System Technical Journal*, *30*(1), 50-64.

Singh, V.B., Chaturvedi, K.K., Khatri, S.K., & Kumar, V. (2015). Bug prediction modeling using complexity of code changes. *International Journal of System Assurance Engineering and Management*, *6*(1), 44-60.

Singh, V.B., & Sharma, M. (2014, November). Prediction of the complexity of code changes based on number of open bugs, new feature and feature improvement. In *2014 IEEE International Symposium on Software Reliability Engineering Workshops* (pp. 478-483). IEEE. Naples, Italy.

Singh, V.B., & Chaturvedi, K.K. (2013, June). Improving the quality of software by quantifying the code change metric and predicting the bugs. In *International Conference on Computational Science and Its Applications* (pp. 408-426). Springer, Berlin, Heidelberg.

Singh, V.B., & Chaturvedi, K.K. (2012, November). Entropy based bug prediction using support vector regression. In *2012 12th International Conference on Intelligent Systems Design and Applications (ISDA)* (pp. 746-751). IEEE. Kochi, India.

Singhal, S., Anand, A., & Singh, O. (2019). SDE based generalized innovation diffusion modeling. *International Journal of Mathematical, Engineering and Management Sciences*, *4*(3), 697-707.

Tamura, Y., & Yamada, S. (2014, October). 3D software tool for reliability assessment based on three dimensional wiener process model considering big data on cloud computing. In *Proceedings of 3rd International Conference on Reliability, Infocom Technologies and Optimization* (pp. 33-37). IEEE. Noida, India.

Tamura, Y., & Yamada, S. (2008). A component-oriented reliability assessment method for open source software. *International Journal of Reliability, Quality and Safety Engineering*, *15*(01), 33-53.

Yamada, S., Nishigaki, A., & Kimura, M. (2003). A stochastic differential equation model for software reliability assessment and its goodness-of-fit. *International Journal of Reliability and Applications*, *4*(1), 1-12.

Yang, J., Liu, Y., Xie, M., & Zhao, M. (2016). Modeling and analysis of reliability of multi-release open source software incorporating both fault detection and correction processes. *Journal of Systems and Software*, *115*, 102-110.

Zhou, Y., & Davis, J. (2005). Open source software reliability model: an empirical approach. *ACM SIGSOFT Software Engineering Notes*, *30*(4), 1-6.

Zhu, M., & Pham, H. (2018). A multi-release software reliability modeling for open source software incorporating dependent fault detection process. *Annals of Operations Research*, *269*(1-2), 773-790.

Original content of this work is copyright © International Journal of Mathematical, Engineering and Management Sciences. Uses under the Creative Commons Attribution 4.0 International (CC BY 4.0) license at https://creativecommons.org/licenses/by/4.0/

Sentiment Analysis of Twitter Corpus related to COVID-19 induced Lockdown

Kritika Saini
Department of Electronics and Communication
Delhi Technological University
Delhi, India
sainikritika46@gmail.com

Dinesh Kumar Vishwakarma
Department of Information Technology
Delhi Technological University
Delhi,India
dinesh@dtu.ac.in

Chhavi Dhiman
Department of Electronics and Communication
Delhi Technological University
Delhi, India
chhavi.dhiman@dtu.ac.in

2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC) | 978-1-6654-4415-6/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICSCCC51823.2021.9478112

***Abstract*— During the global COVID-19 outbreak, social media has been the biggest platform for the rapid spread of both accurate and inaccurate information. This infodemic has given rise to a plethora of misinformation manifesting in the form of mass fear and anxiety. According to a study conducted by the Oxford Researchers, 59 percent of the inaccurate information associated with Covid pandemic that has been invalidated by the fact checker remain on Twitter without any prior warning[1]. This paper analyzes the psychometric impact of COVID-19 infodemic using Stacked LSTM model and Stanford's pretrained GloVe embeddings. In this, sentiment score ranging from -1 to +1 was computed for more than 6.7 tweets, where -1 denotes purely negative sentiment and +1 represents positive sentiment. This research provides insights into Corona virus psychometric progression, and outlines the proposed methodology, its implications, limitations and opportunities.**

Keywords*—sentiment analysis; twitter sentiment; COVID-19; pandemic; stacked LSTM*

I. INTRODUCTION

Social media platforms not only provide a virtual space to communicate with peers, they are easily accessible and have a global reach. On these platforms, enormous amount of data is made publicly available, which comprises of both beneficial and misleading information[1][2]. Nowadays, people are extensively using these platforms to talk about some taboo topics, like sexual orientation, assaults, mental wellbeing which they cannot discuss in their immediate circles. Since social media is becoming such a crucial part of the day-to-day life of masses, delving into enormous pool of social data to understand people's psychological and mental state is significant.

The unforeseeable outbreak of Corona Virus pandemic has grabbed a lot of attention on social media. People are using social media to express their thoughts, fears, opinions and to disseminate important information and health advisories. Twitter itself has seen a sharp increase in COVID-19 tweets starting from only 63 tweets on January 1,2020 and peaking up to 19.8Million tweets per day on March 13,2020 [2]. There are almost 40,000 tweets about Covid19 which have received more than 1000 RTs in a tweet.[2] With these exponentially escalating tweets, misinformation regarding Corona Virus is also increasing, in form of doctored videos, fake images and messages with different attributes. According to an analysis by BBC, false accusations targeting Muslims for deliberately spreading virus peaked in beginning of April in India after the Tablighi Jamaat case.[3] In addition, unproven solutions are being proposed such as, [3] as a cure to Corona Virus. Some false/unproven allegations on China have resulted in a rise of Sinophobic behaviour on the social media. These false claims manifest in the form of mass fear, anxiety, social riots and hysteria, targeting meat traders, poultry workers, Muslims, Chinese and other sections of society.

The sentiment analysis uses deep learning based approaches[4] where data is available in form of multimedia[5][6].

This paper presents textual analysis of twitter data to understand public sentiments associated with Corona Virus infodemic. It uses Stacked LSTM model by incorporating Stanford's GloVe embeddings, consisting of 6 Billion words of 50 dimension each as transfer learning. This research outlines a methodological approach to analysing tweets, especially for calculating sentiment score associated with them.

A. Organisation

The paper is organised as follows. Section II briefly reviews the related work on sentiment analysis algorithms for tweets associated with COVID-19. Section III presents the data collection, pre-processing steps and model framework used in this research. Section IV depicts the experimental results and finally Section V concludes this paper.

II. RELATED WORK

Copious amount of extensive research work has been previously done on social data to understand the mental and psychological state of public. In 2002, Pang et al. [7] used unigram models for sentiment classification on movies reviews and analyzed the performance of different classifiers. His work has served as a foundation for many researchers across different domains. Alec Go et al. [8] first classified messages on microblogging services like twitter using distant supervision learning, and added emoticons along with the microblogs bringing a drastic increase in the overall accuracy.

With the recent outbreak of corona virus, World's researchers and scientists have come together to accelerate the research and development process in the field of textual analysis. Lamsal et al. [9] provided a global level twitter database by collecting real-time twitter microblogs related to COVID-19. This is an ongoing database collection commencing from March 20,2020. Bani Leen Kaur Jolly et al. [10] introduced the first social media based Corona Virus dataset at national and state levels from India. They modified Empath to study the psychometric impact and time-series analysis of COVID-19 infodemic at national and state level[10].

978-1-6654-4415-6/21/$31.00 ©2021 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:23:38 UTC from IEEE Xplore. Restrictions apply.

Fig 1. An instance of Word Cloud in Twitter data

Wang et al. [11] analyzed negative sentiments on Sina Weibo, a popular Chinese microblogging site using unsupervised BERT model and conducted sentiment classification with considerable accuracy. R.F Sear et al. [3] analyzed the growing public health debates associated with vaccination on in-built online community spaces, reporting that the vaccine supporter group has more detailed and focused discussion around COVID-19 than the anti-vaccination community.

Study regarding the generic attitude toward social distancing in Australia was conducted by Kayes et al. [12], describing that 82.5% of tweets supported social distancing. P. Ghasiya and K. Okamura[14] investigated the COVID-19 news across 4 different nations and the sentiment associated with the pandemic in those countries. Other researchers have focused on the human psychological needs and sentiment classification in the challenging times of COVID-19.

III. PROPOSED METHDOLOGY

This section describes the methodological approach used to analysing twitter data, which proposes the use of deep learning based Stacked LSTM model to calculate the sentiment score associated with COVID-19 tweets.

A. DATA COLLECTION AND EXPLORATORY ANALYSIS

The dataset used in this research was curated by Rabindra Lamsal [9] by monitoring the twitter feed for world-wide Covid tweets in real-time. The dataset is acquired using the streaming API of twitter incorporated with more than 90 different keywords and hashtags. The streaming API of twitter permits the use of 400 different keywords and 500 location options to filter the real-time twitter feed. It matches these keywords against the tokenised tweets. Till August 2020, 54 dissimilar keywords and hashtags generally used while referring to this global health crisis were used to filter out the required tweets. The hashtags used to mine Covid tweets were related to corona virus, quarantine, social distancing, working at home, vaccines, masks, hand sanitation etc. This dataset is a large collection starting from March 20, 2020 publicly available at IEEE Dataport. It consists of more than 900 million Twitter IDs with their associated sentiment score.

In this research, only 6.7 Lakh tweets from this dataset have been used to build the model. Fig 1 shows the graphical representation of word frequency in the dataset in form of Word Cloud. The word cloud of the twitter data also contains many blasphemous, vulgar and obscene phrases which clear show the resentment of public against COVID-19. Early exploratory analysis of the dataset revealed that majority of the tweets had sentiment score close to 0(i.e., neutral sentiment), the tweet frequency decreases as the sentiment becomes positive or negative. Fig 2 represents the tweet frequency associated with sentiment score in the form of a histogram for the first 2 csv files.

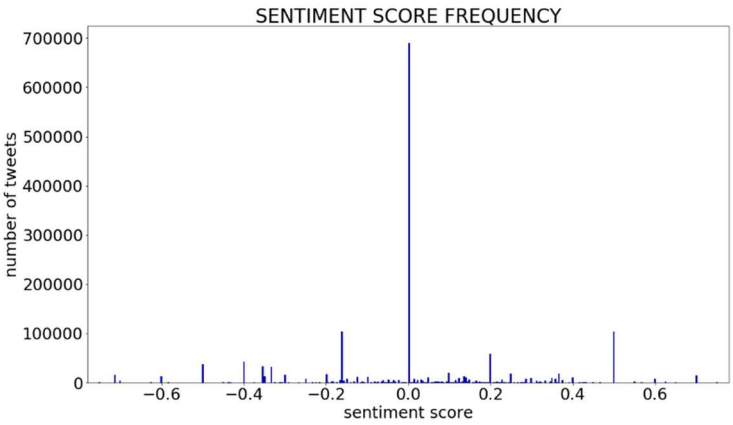


Fig 2. Tweet frequency associated with sentiment score

Fig 3 depicts the scatter plot between Twitter IDs and the sentiment score. It has been found that there is no linear correlation between the Twitter IDs and its associated sentiment, having correlation coefficient of 0.052. Moreover, Fig 3 shows that the density of datapoints is maximum when sentiment score is close to 0(neutral sentiment). This infers that the majority tweets have either neutral sentiment or slightly positive or negative sentiment.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:23:38 UTC from IEEE Xplore. Restrictions apply.

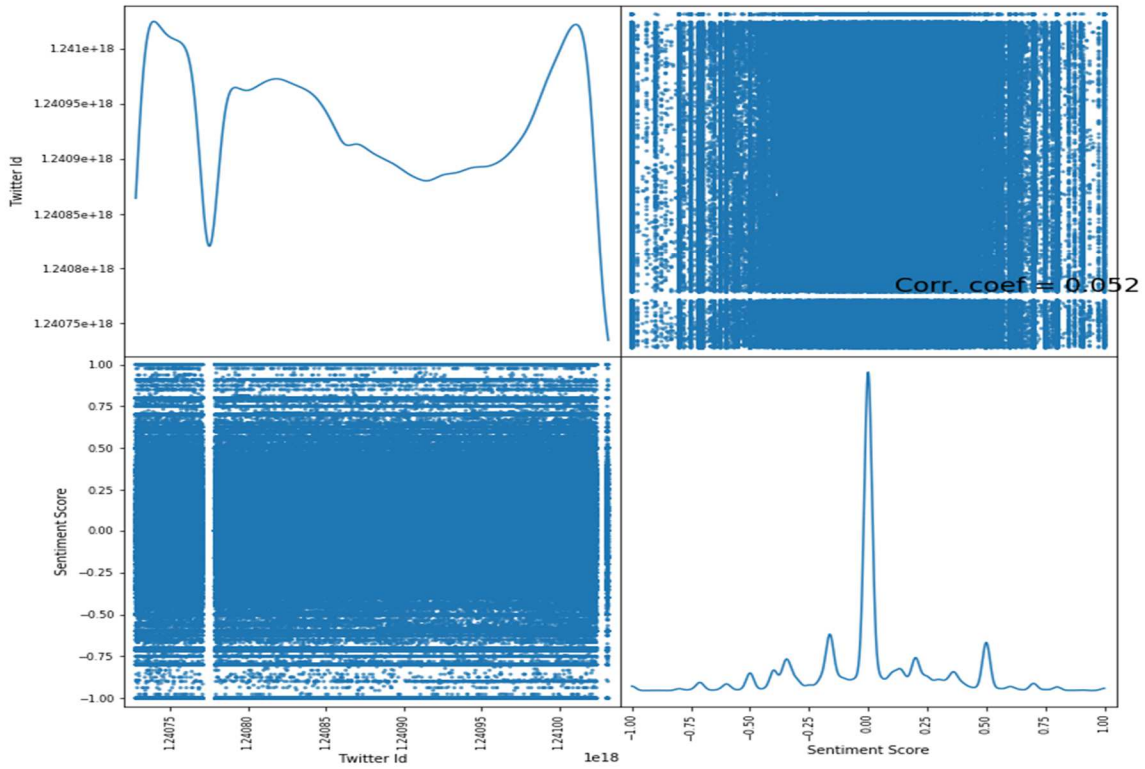


Fig 3. Scatter and Density plot between Twitter IDs and sentiment score

B. TWEET HYDRATION AND PREPROCESSING

As the original database consists of only the tweet identifiers, hydration of the tweets IDs is required to extract the original tweet. In order to hydrate the Twitter identifiers from the database, Twarc was used, which is a python library for archiving Twitter JSON data developed by Documenting the Now. Twarc uses the Twitter API to retrieve the twitter JSON data from the identifiers.

The following steps were used to pre-process the tweets after their hydration:

1) Lowercasing – the entire text was converted into lower case characters.
2) Removal of Retweets|RT|via – Any reference to the retweet was removed as it didn't provide any relevant information.
3) Removal of mentions and links – Mentions and links used in the twitter data were taken out.
4) Negation Handling – all the negative helping verbs, including don't, didn't, wasn't, isn't etc were replaced by "not" as it represents the same semantic. This helped in decreasing the dimensionality of feature space.
5) Tokenisation of the twitter data was done to capture words with only standard English Alphabets.
6) Removal of stopwords- Useless words or phrases which do not provide any relevant information regarding the sentiment of the text were removed. These include pronouns, conjunctions, linking verbs etc.
7) Lemmatization – Wordnet Lemmatizer was used to remove inflectional endings to produce the dictionary form of words, referred as lemma.

Table 1 provides a comparative representation of the original and processed text along with its associated sentiment score.

TABLE 1 EXAMPLES OF COVID-19 TWEETS IN THE DATABASE

ORIGINAL TWEET	PROCESSED TWEET	SENTIMENT SCORE
I'm really proud of how @MayorOfLA and his staff have been handling the #COVID19 situation. He's a poet who is passionate about his city and the people in it, and has displayed cool, calm and compassionate leadership #SaferAtHome https://t.co/TtrHlRMxJp\n	really proud staff handling covid situation poet passionate city people displayed cool calm compassionate leadership saferathome	0.35 (positive)
RT @yennikwok: The Indonesian state-owned company that bought coronavirus test kits from China will sell the kits to hospitals	indonesian state owned company bought coronavirus test kit china sell kit hospital	0.00 (neutral)
Corona's epidemic is a very vague secret that only China and America know. There is a strange matter related to biological weapons between America and China. What is the purpose of it, how it is used, and why??? We will discover that soon #COVID	corona epidemic vague secret china america know strange matter related biological weapon america china purpose used discover soon covid	-0.222 (negative)

C. MODEL FRAMEWORK

Long-Short Term Memory (LSTM) networks are distinct type of Recurrent Neural Networks which are efficient in

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:23:38 UTC from IEEE Xplore. Restrictions apply.

handling long-term dependencies in the data. LSTMs use state vector to store the long-term dependencies and thereby associate it with the present task. They are highly efficient in language modelling due to their capability of preserving the sequential order of text. The translational equations of an LSTM cell are as follow:

$$i_t = \sigma_{si}(\omega_i x_t + U_i h_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma_{si}(\omega_f x_t + U_f h_{t-1} + b_f) \quad (2)$$

$$o_t = \sigma_{si}(\omega_o x_t + U_o h_{t-1} + b_o) \quad (3)$$

where i_t, f_t, o_t symbolise the activation functions for input gate, forgot gate and output gate respectively. x_t is the input word vector for the LSTM. ω, U and b represent the weight matrices and the bias parameter respectively. The sigmoid function is denoted by σ_{si}. The equations of cell memory and input block for the LSTM are stated below:

$$\tilde{c}_t = \tanh(\omega_c x_t + U_c h_{t-1} + b_c) \quad (4)$$

$$c_t = h_t \oplus c_{t-1} + i_t \oplus \tilde{c}_t \quad (5)$$

$$h_t = o_t \oplus \tanh(c_t) \quad (6)$$

where c_t is the cell state, h_t is the hidden output, x_t is the input vector. $\oplus$ denotes element-wise multiplication.

Deep LSTM or Stacked LSTM networks were first introduced by Alex Graves et al. [9], in his paper regarding speech recognition in deep neural networks, reporting that the deepness of the network had a greater impact on accuracy than the number of memory cells in a particular layer. Adding additional layers results in greater level of comprehension as it processes the learned representations from prior layers. This allows the hidden/context vector at every level to act at different time periods, thus enhancing the performance of the network.

In this research, a Stacked LSTM based network has been proposed with two vertical layers to predict the sentiment intensity of the Corona Virus microblogs. Fig 4 represents the overview of the architecture of the model used in this research.

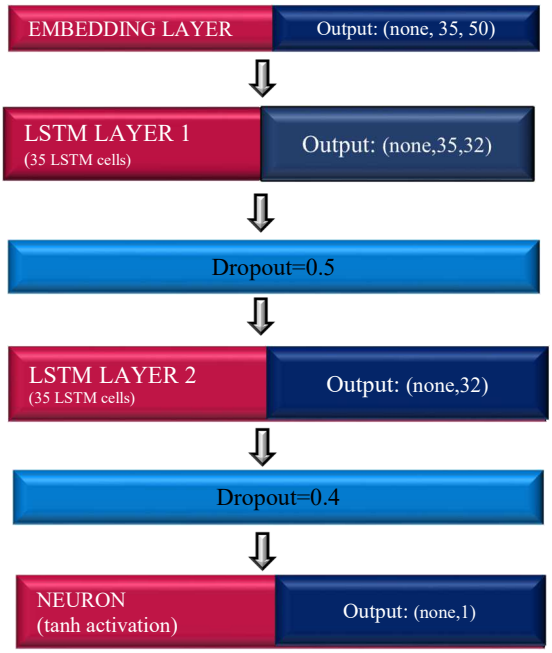


Fig 4. Overview of the model architecture

The input to the LSTM network is the embedding output in which each word or phrase in the tweet is portrayed by its corresponding 50-dimensional word-vector. In this model, each word from the lexicon has been mapped into a vector of real numbers using pretrained Global Vector (GloVe) embeddings. Word embeddings have proved to enhance the performance of the model in the past few years in various domains of natural language processing. The variable length text was first converted into a feature vector consisting of only 35 words. Each word was then mapped to its corresponding vector of 50dimension, forming the embedding output.

The output of the embedding layer was then fed into the first LSTM layer, to execute sequential computation of the tweet and pass the output vector to the next layer. dropout layer with 50percent dropout was added in order to minimize overfitting and generalization error. The second layer in the LSTM network recombined the learned representations from the first layer to produce the final context vector. A neuron with Hyperbolic Tangent as activation was added at last, to recognize the sentiment score elicited from the tweet.

IV. EXPERIMENTS AND RESULT

In order to assess the performance of the proposed work, an experiment was conducted on the covid database, followed by a comprehensive discussion of the outcome. The model was trained on 538,215 processed tweets having sentiment scores in the range of -1 to +1. Out of these tweets more than 1.6 Lakh were taken for validation.

It was found that optimization of the mean absolute error by using adam as an optimizer, gave remarkable results. Adam is a combination of RMSprop and Adagrad optimization algorithms and it is able to handle sparse gradients on noisy problems. The training was done in batch size of 512 for 200 epochs. The evaluation of the mean absolute error for training and validation samples with the number of epochs is shown in Fig 5.

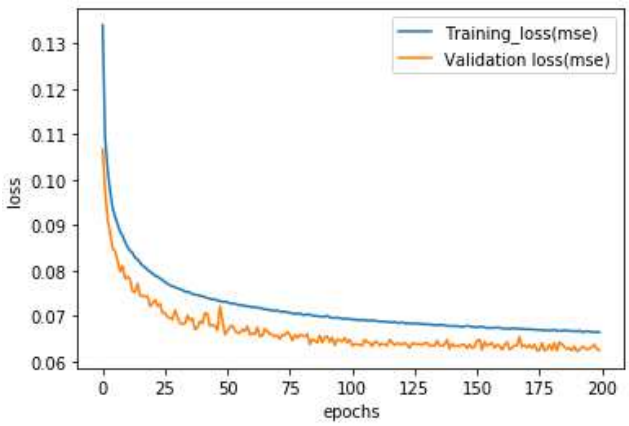


Fig 5. Evaluation of training loss (mean absolute error) and validation loss with the number of epochs

It can be seen that the mean absolute error for both training and validation samples decreases exponentially with the number of epochs. The validation error outperforms the training error, signifying a constructive generalization by the model. Fig 6 represents the time analysis of R-squared score for training and validation set.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:23:38 UTC from IEEE Xplore. Restrictions apply.

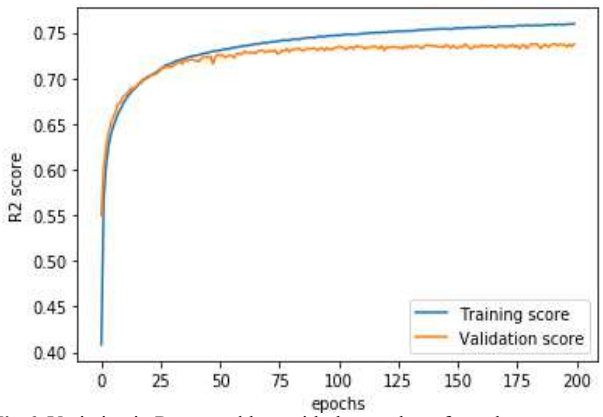


Fig 6. Variation in R-squared loss with the number of epochs

Fig 6 reveals that as the number of epochs increases, the comprehension of the model to replicate the observed outcomes also increases. The training R-squared score continues to increase while the validation score becomes static, indicating optimal generalization point for goodness-of-fit.

The performance of the model was evaluated against various metrics including R-squared score (R2 score), Mean absolute error (MAE), Root mean squared error (RMSE), Mean Squared error (MSE) and Mean Squared Logarithmic error (MSLE). The result of the model for the training data and the testing data are displayed in Table 2.

TABLE 2. RESULT OF STACKED LSTM MODEL ON TRAINING AND TESTING DATASET

Metric	Training set	Testing set
R-squared score	0.7478	0.7301
Mean absolute error (MAE)	0.0604	0.0636
Root mean squared error (RMSE)	0.0604	0.0636
Mean squared error (MSE)	0.0211	0.0225
Mean squared logarithmic error	0.007155	0.007573

A comparative analysis of the proposed method was performed with the results obtained using TextBlob. TextBlob is an efficient open-source library for Natural language Processing and textual analysis. Table 3 gives the comparison between the positive, negative, and neutral tweets computed using TextBlob and the proposed method on the same dataset.

TABLE 3. COMPARATIVE ANALYSIS OF POSITIVE, NEGATIVE AND NEUTRAL TWEETS LABELED BY TEXT BLOB AND THE PROPOSED METHOD

METHOD USED	TextBlob	Proposed Method
Number of positive tweets	211039	186269
Number of neutral tweets	312986	367599
Number of negative tweets	148744	118901
Total number of tweets analyzed	672769	672769

V. CONCLUSION

This paper introduces the textual analysis of tweets associated with Corona Virus using Deep LSTM model incorporated with GloVe embeddings. The goal of the paper is to present a novel application to detect the latent sentiment in microblogs. The results show while majority of the tweets had a neutral sentiment score, about 17.67% of the total tweets analyzed corresponding to negative sentiments. These negative sentiments in the tweets are due to the rising fear, frustration and hatred due to the pandemic. Some of these negative sentiments also arise due to the spread of misinformation on the social media giving a rise to Global Sinophobic behavior. The contents of this paper will aid in deeper levels of understanding of the mental and psychological state of public with respect to COVID-19 related issues. This may also provide an initial step to detect and gradually diminish the growing rate of misinformation on the social networking sites.

However, this sentiment analysis model can only predict the sentiment intensity of only textual data. Moreover, the tweets used in this research comprised of English language microblogs excluding the emoticons. The model won't be able to classify multi-lingual or non-English content. This might serve as a limitation in some cases. For the future works, this research on COVID-19 will be extended to deducing multilingual tweets consisting of graphic images and video content also.

REFERENCES

[1] P. Meel and D. K. Vishwakarma, "Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities," *Expert Systems with Applications*, vol. 153. Elsevier Ltd, Sep. 01, 2020, doi: 10.1016/j.eswa.2019.112986.

[2] D. Varshney and D. K. Vishwakarma, "A review on rumour prediction and veracity assessment in online social network," *Expert Syst. Appl.*, vol. 168, no. October 2020, p. 114208, 2020, doi: 10.1016/j.eswa.2020.114208.

[3] R. F. Sear *et al.*, "Quantifying COVID-19 Content in the Online Health Opinion War Using Machine Learning," *IEEE Access*, vol. 8, pp. 91886–91893, 2020, doi: 10.1109/ACCESS.2020.2993967.

[4] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: a review," *Artif. Intell. Rev.*, vol. 53, no. 6, pp. 4335–4385, 2020, doi: 10.1007/s10462-019-09794-5.

[5] A. Yadav and D. K. Vishwakarma, "A deep learning architecture of RA-DLNet for visual sentiment analysis," *Multimed. Syst.*, vol. 26, no. 4, pp. 431–451, Aug. 2020, doi: 10.1007/s00530-020-00656-7.

[6] A. Yadav and D. K. Vishwakarma, "A unified framework of deep networks for genre classification using movie trailer," *Appl. Soft Comput. J.*, vol. 96, Nov. 2020, doi: 10.1016/j.asoc.2020.106624.

[7] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in *Empirical Methods in Natural Language Processing (EMNLP)*, 2002, vol. 10, no. July, pp. 79–86, doi: 10.3115/1118693.1118704.

[8] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification using Distant Supervision," *CS224N Proj. Report, Stanford*, vol. 1, no. 12, pp. 1–6, 2009, doi: 10.1016/j.sedgeo.2006.07.004.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:23:38 UTC from IEEE Xplore. Restrictions apply.

[9] R. Lamsal, “Coronavirus (COVID-19) Tweets Dataset | IEEE DataPort.” .

[10] B. L. K. Jolly, P. Aggrawal, A. Gulati, A. S. Sethi, P. Kumaraguru, and T. Sethi, “Psychometric Analysis and Coupling of Emotions Between State Bulletins and Twitter in India during COVID-19 Infodemic,” no. July 2017, 2020.

[11] B. Model, “COVID-19 Sensing : Negative Sentiment Analysis on Social Media in China via,” vol. 4, 2020, doi: 10.1109/ACCESS.2020.3012595.

[12] Kays, “Automated Measurement of Attitudes Towards Social Distancing Using Social Media: A COVID-19 Case Study,” no. April, 2020, doi: 10.20944/preprints202004.0057.v1.

[13] A. Graves, A. Mohamed, and G. Hinton, “SPEECH RECOGNITION WITH DEEP RECURRENT NEURAL NETWORKS Alex Graves, Abdel-rahman Mohamed and Geoffrey Hinton Department of Computer Science, University of Toronto,” no. 3.

[14] P. Ghasiya and K. Okamura, "Investigating COVID-19 News Across Four Nations: A Topic Modeling and Sentiment Analysis Approach," in *IEEE Access*, vol. 9, pp. 36645-36656,2021, doi: 10.1109/ACCESS.2021.3062875.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 06:23:38 UTC from IEEE Xplore. Restrictions apply.

Sentiment Analysis Using XLM-R Transformer and Zero-shot Transfer Learning on Resource-poor Indian Language

AKSHI KUMAR, Department of Computer Science & Engineering, Delhi Technological University, New Delhi, India
VICTOR HUGO C. ALBUQUERQUE, Laboratory of Industrial Informatics, Electronics and Health, University of Fortaleza (UNIFOR), Ceará, Brazil

Sentiment analysis on social media relies on comprehending the natural language and using a robust machine learning technique that learns multiple layers of representations or features of the data and produces state-of-the-art prediction results. The cultural miscellanies, geographically limited trending topic hash-tags, access to aboriginal language keyboards, and conversational comfort in native language compound the linguistic challenges of sentiment analysis. This research evaluates the performance of cross-lingual contextual word embeddings and zero-shot transfer learning in projecting predictions from resource-rich English to resource-poor Hindi language. The cross-lingual XLM-RoBERTa classification model is trained and fine-tuned using the English language Benchmark SemEval 2017 dataset Task 4 A and subsequently zero-shot transfer learning is used to evaluate the classification model on two Hindi sentence-level sentiment analysis datasets, namely, IITP-Movie and IITP-Product review datasets. The proposed model compares favorably to state-of-the-art approaches and gives an effective solution to sentence-level (tweet-level) analysis of sentiments in a resource-poor scenario. The proposed model compares favorably to state-of-the-art approaches and achieves an average performance accuracy of 60.93 on both the Hindi datasets.

CCS Concepts: • **Information systems** → **Information retrieval**; **Retrieval tasks and goals**; **Sentiment analysis**; • **Computing methodologies** → **Artificial intelligence**; **Natural language processing**; **Information extraction**;

Additional Key Words and Phrases: Sentiment analysis, resource-poor language, transformer, deep learning

ACM Reference format:
Akshi Kumar and Victor Hugo C. Albuquerque. 2021. Sentiment Analysis Using XLM-R Transformer and Zero-shot Transfer Learning on Resource-poor Indian Language. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 20, 5, Article 90 (June 2021), 13 pages.
https://doi.org/10.1145/3461764

Authors' addresses: A. Kumar, Department of Computer Science & Engineering, Delhi Technological University, New Delhi, India; email: akshikumar@dce.ac.in; V. H. C. de Albuquerque, Laboratory of Industrial Informatics, Electronics and Health, University of Fortaleza (UNIFOR), Ceará, Brazil; email: victor.albuquerque@ieee.org.
Author's Current Address: Victor Hugo C. de Albuquerque, Graduate Program on Tele-Informatics Engineering, Federal University of Ceará, Fortaleza, Fortaleza/CE, Brazil. Graduate Program on Telecommunication Engineering, Federal Institute of Education, Science and Technology of Ceará, Fortaleza/CE, Brazil.
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 ACM 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.
© 2021 Association for Computing Machinery.
2375-4699/2021/06-ART90 $15.00
https://doi.org/10.1145/3461764

1 INTRODUCTION

Social media has become extremely popular in today's time as people often view it as a platform to voice their opinions about various organizations, people, companies, products, or events. There is endless possibility to use this wealth of information for getting to know the sentiments of people, their attitudes, and their emotions. The quantum of opinionated data on social media is so vast that it now finds use in tracking customer reactions, monitoring competitions, anticipating election outcomes, and predicting investment trends and box office revenues. **Sentiment analysis (SA)** [1, 2] is the field of study that automatically ascertains the polarity of opinions expressed by users toward entities in a chunk of text or review. In general, polarity is categorized into one of three possible classes, namely, positive, negative, or neutral. But determining real-time sentiments is challenging owing to volume of available reviews and variety in review expression. Biased and fake/spam reviews further complicate predictive analytics. The studies within the domain of SA are primarily divided into six types, which include studies based on granularity, language, modality, techniques, application areas, and sub-tasks. On the basis of granularity of analysis, SA are either categorized as coarse-grained (document level and sentence level) or fine-grained (phrase level or aspect level). On the basis of language analyzed, SA can either be monolingual (one language like English, Hindi, Chinese, Arabic, etc.) or Multilingual (combination of two or more languages (code-mix and code-switch). With the recent proliferated use of multimedia content in social media, the studies on SA can either be categorized into mono-modal (single modality, primarily text only) or Multimodal (combination of two or more media types, for example, text +image). The techniques used for SA broadly fall into four categories, that is, lexicon-based, machine learning-based, deep learning-based, and hybrid techniques. There are various application areas where the use of SA can assist intelligence and decision making. For example, business, marketing, governance and politics, and so on. Finally, the study on SA can further be done based on the related sub-tasks such as subjectivity analysis, sarcasm/irony/pun/humour detection, and emotion analysis, to name a few.

Automatic detection of sentiments in social media platforms is typically a natural language understanding-based classification task [3]. The data scientists have accomplished a lot at creating more accurate sentiment classifiers, but there's still a lot to do. Few key challenges of automated sentiment analysis include identifying subjectivity and tone; context and polarity; irony and sarcasm; negations; comparisons; emojis or emoticons and defining neutral. The cultural miscellanies, geographically limited trending topic hash-tags, access to aboriginal language keyboards and conversational comfort in native language exemplify the variability and size of user-generated content, thereby compounding the linguistic challenges of SA. Hindi is the official language of India and a sizeable population speaks/writes Hindi, while the rest are comfortable in their regional language. The availability of keyboards with "Devanagari" scripts on mobile phones has made it a popular language choice. But like other Indian languages, Hindi is also a resource-poor language [4]. Sentiment analysis in Indian languages is largely unexplored due to the scarcity of various resources and tools such as annotated corpora, lexicons, dependency parser, **Part-of-Speech (PoS)** tagger and benchmark datasets, and so on. Also, like most of the Indian languages, Hindi too has free-word order. For example, गाने अच्छे हैं इस फिल्म के, इस फिल्म के गाने अच्छे हैं, अच्छे गाने हैं इस फिल्म के, all three statements convey the same meaning with different word order.

There have been several recent studies published on sentiment analysis focusing on different aspects of affect recognition such as emotion [5, 6], aspect [7, 8], and mood [9], to name a few. While there are a few studies published on languages such as Arabic [10, 11] and Urdu [12], most studies and datasets created primarily contain English data. The use of monolingual word embeddings is pervasive in **natural language processing (NLP)**. Data augmentation [13, 14] and multilingual word embeddings [15, 16] have also been applied to take advantage of existing English datasets to improve the performance in systems dealing with languages other than

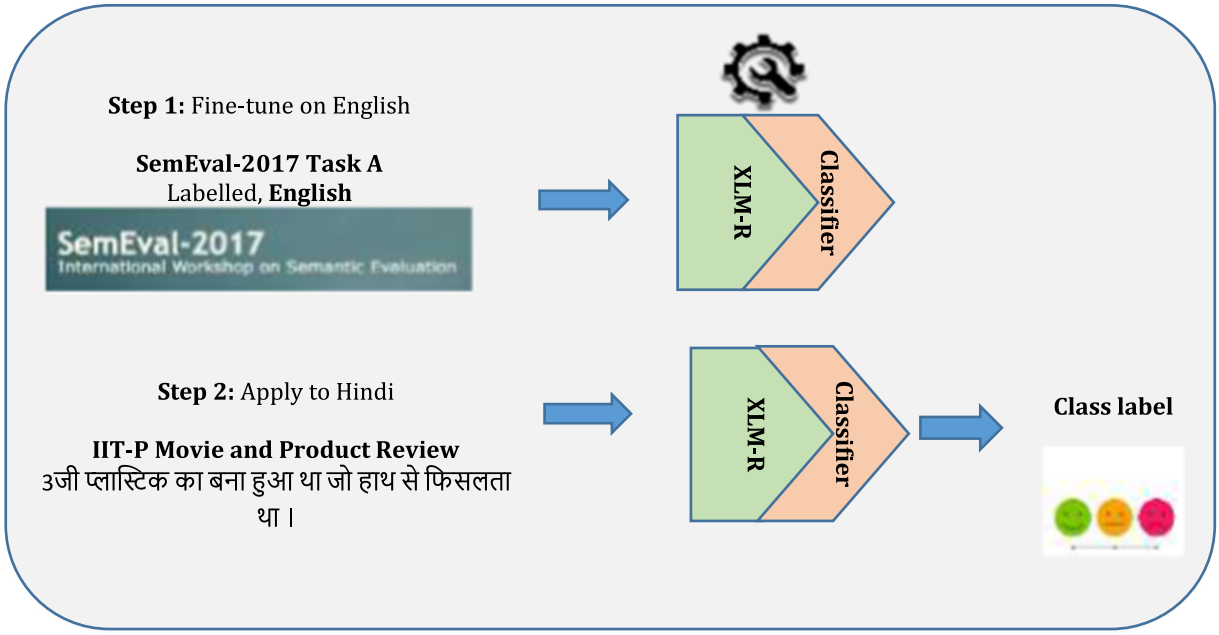


Fig. 1. Cross-lingual XLM-R transformer model.

English. Even though, multilingual model like BERT-m [13] show some cross-lingual characteristics it has not been trained on cross-lingual data [17]. To represent meaning and transfer knowledge across different languages, cross-lingual word embeddings can be used. These are embeddings that are obtained from a dataset in any available language and then used to train a model that will be used to predict polarity of texts in resource-poor language. To the best of our knowledge, the state-of-the-art cross-lingual contextual embeddings such as **XLM-RoBERTa (XLM-R)** [18] have not been applied yet to sentiment analysis in resource-poor Hindi language. To address this gap, we evaluate the performance of cross-lingual contextual word embeddings and transfer learning in projecting predictions from English to Hindi language. The classification model is trained on a resource rich language, typically English, using a cross-lingual transformer model and zero-shot transfer learning (while training, the model is not seeing even a single example of test language, but it can still make predictions in test language) is performed on a resource-poor language, i.e., Hindi (Figure 1)

The proposed model compares favorably to state-of-the-art approaches and gives an effective solution to sentence-level (tweet-level) analysis of sentiments in a resource-poor scenario. Thus, the primary contribution of this work is:

- Use of cross-lingual XLM-R transformer model for tweet-level sentiment classification for resource-poor Hindi language
- Apply zero-shot transfer learning for target resource-poor languages that lack extensive labelled and unlabelled data sets

Organization of the article is as follows: Section 2 briefs about the related work within the domain of Hindi sentiment analysis, followed by the discussion on methodology used in Section 3. Section 4 presents the details of model performance, and the conclusion is given in Section 5.

2 RELATED WORK

Sentiment analysis is the detection of attitudes, defined using a 4-tuple {Holder, Target, Type, Text}, where the holder is the source of attitude, target is the aspect of attitude, type defines the

commonly weighted polarity type such as positive, negative or neutral and text is the piece of information containing the attitude (sentence or entire document).

Having started with simple polarity detection, contemporary SA has advanced to a more nuanced analysis of context, affect and emotion sensing. But detecting fine-grained sentiment in natural language is tricky even for humans, making its automated detection more complicated. Moreover, online opinions can be put forth in the form of text reviews or ratings, for a product as a whole, or each of its individual aspects. Multiple and lengthy reviews, usage of casual dialect with micro-text (wordplay, neologism and slang), use of figurative language (sarcasm, irony), multilingual content (code-mixed and code-switched), and opinion spamming add challenges to the task of extracting opinions. Recently, memes, GIFs, typo-graphic (artistic way of text representation), info-graphic (text embedded along with an image) visual content and edited videos dominate the social feeds. Hence the task of SA can range from a simplistic classification of text attitude as positive or negative to a more advanced fine-grain emotion analysis and complex attitude type detection. In recent times the "text" component has compounded as multimedia text owing to the extensive use of various semiotic modalities (aural, visual, and textual) on social media. At the same time, multilingual sentiment analysis has also emerged as a prominent research problem as most of the existent work has been done on English only text [19]. Figure 2 depicts the various sentiment analysis tasks and the approaches used.

Various studies on diverse resource-poor languages have been reported in literature. Delbrouck et al. [20] proffered a **transformer-based joint-encoding (TBJE)** for the task of emotion recognition and SA in CMU-MOSEI dataset. González et al. [21] gave a transformer encoder model for Twitter SA in Spanish language. Sultan et al. [22] used XLM-R and transfer learning on both the monolingual English and Spanish data and Spanish-English code-mixed data. Kuratov and Arkhipov [23] used transfer learning from a multilingual BERT model to monolingual model for SA in Russian language. Sarkar et al. [24] proposed a Hierarchical Attentive Network using BERT for multilingual document-level SA. Recent studies report SA in code-mix Hindi-English text [25, 26], which was provided as the SemEval 2020 Task 9 [27]. Literature also reports emotion analysis in Hindi text. In 2019, Kumar et al. [28] introduced a Hindi emotion annotated text corpus, BHAAV. In 2020, Garg and Lobiyal [29] propose Hindi EmotionNet for SA in Hindi.

Studies related to SA on Hindi have been reported since the start of decade. Benchmark studies have been reported on three datasets, namely, the Twitter (SAIL 2015) dataset,[1] IIT-Patna Product/Service reviews dataset (sentence-level and aspect-level), and IIT-Patna Movie review dataset (sentence-level).[2] In 2012, Bakliwal et al. [30] provided Hindi subjective lexicon for SA. In the same year, Balamurali et al. [31] linked WordNets of Hindi and Marathi for cross-lingual SA. In 2015, various studies have been reported on the shared task SAIL dataset [32-34], which contained Hindi, Bengali, and Tamil tweets. Akhtar et al. [35] presented work on aspect-based SA (ABSA) in Hindi have also been reported. Cross-lingual and multi-lingual ABSA has also been reported [36, 37]. Pertinent studies also report deep learning architectures for SA in Hindi. In 2016, Akhtar et al. [38] proposed a hybrid deep learning model and evaluated its performance on four Hindi language datasets. The model achieved an average accuracy of 51.11% on the IITP-Movie and IIIT-Product review datasets. Most recently, in 2020 Kunchukuttan et al. [16] presented the IndicNLP word embeddings and achieved state-of-the-art results for the IITP-Movie and IIIT-Product review datasets.

[1] http://amitavadas.com/SAIL/data.html.
[2] https://www.iitp.ac.in/~ai-nlp-ml/resources.html.

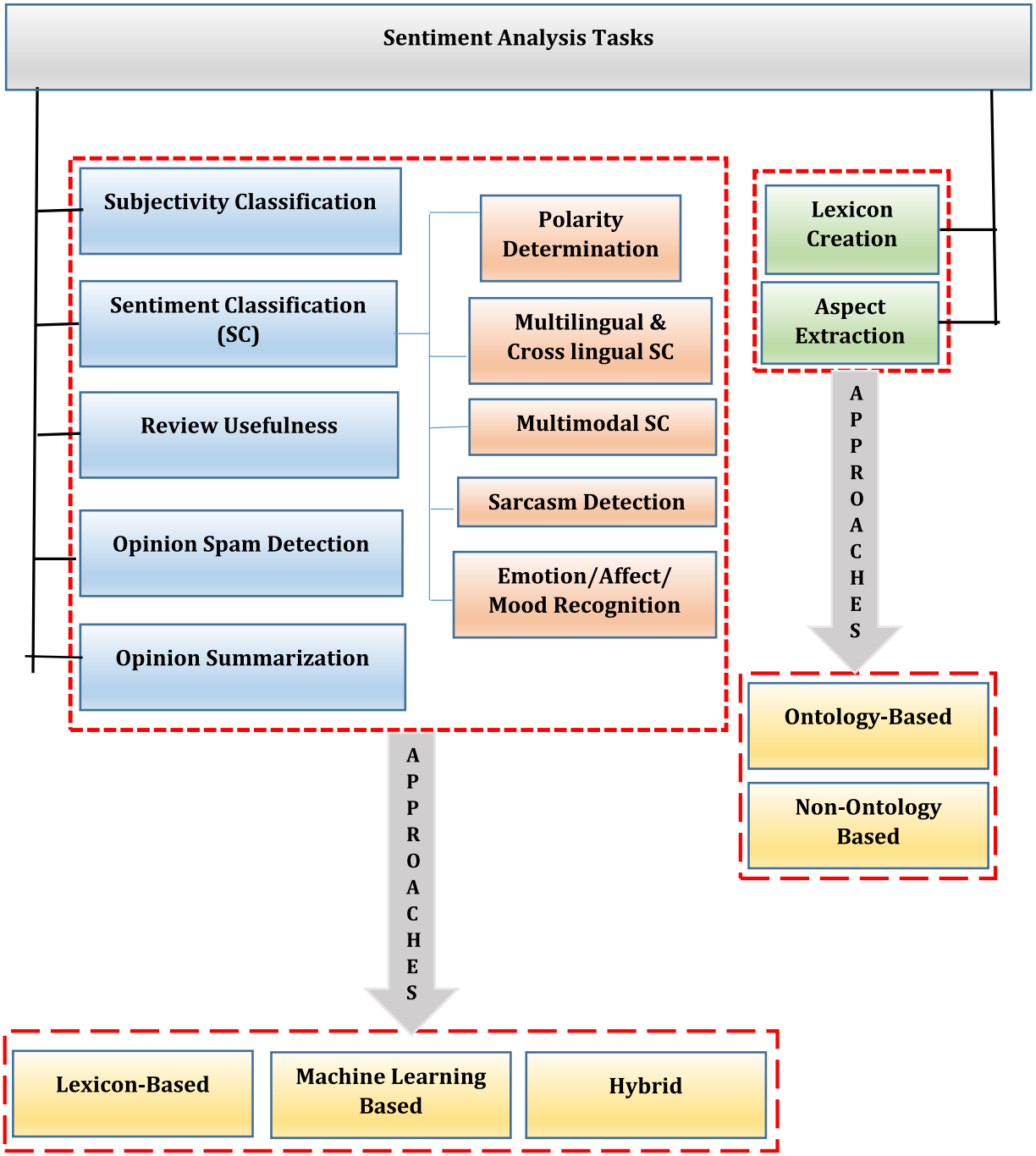


Fig. 2. Sentiment analysis tasks.

3 METHODOLOGY

3.1 Datasets

The SemEval 2017 Task 4A dataset,[3] which consists of English tweets annotated for positive, negative, or neutral categories, is used to train the language model on the sentiment analysis task on English, i.e., the training dataset. The IIT-Patna Movie and Product review datasets, which contains Hindi texts, are used to experiment transfer learning and accomplish sentence-level

[3]https://alt.qcri.org/semeval2017/task4/.

Table 1. Dataset Statistics

Dataset	Language	Positive	Negative	Neutral	Conflict	Total
SemEval-2017 Task 4A*	English	2,352	3,811	5,742	F	11,905
Movie Review	Hindi	823	530	598	201	2,152
Product Review	Hindi	2,290	712	2,226	189	5,417

*Only Tweet IDs provided by organizers and therefore some tweets were not available for download due to edited privacy or removed.

sentiment analysis task on Hindi language. Though the IIT-Patna Movie review set has been annotated with four labels (positive, negative, neutral, or conflict), we choose to ignore the conflict class to ensure the same number of classes in the target dataset. The dataset statistics are given in Table 1.

3.2 Pre-processing

Deep learning models accept certain kinds of inputs, which is vectors of integers, each value representing a token. Each string of text must first be converted to a list of indices to be fed to the model. Also, the data is converted into BERT specific format. That is, before performing tokenization, the text is converted to lower case. This conversion to lower-case is performed through the BERT tokenizer and the TFIDF vectorizer. Long sentences, if any, were split into multiple samples.

3.3 XLM-R Transformer Architecture for Sentiment Classification

XLM-R is a transformer-based multilingual masked language model released by the Facebook AI team in November 2019 as an update of its original XLM-100 model. It is a scaled cross-lingual sentence encoder trained on 2.5 Tb of data across 100 languages data filtered from CommonCrawl texts. Compared to the original version, the biggest update of XLM-Roberta is a significant increase in the amount of training data [18].

XLM-R achieves state-of-the-arts results on multiple cross-lingual benchmarks. Based on masked language model, it works as a successful alternative for non-English NLP. The remarkable fact about XLM-R is that it is compatible with both monolingual as well as cross-lingual benchmarks and alleviates the curse of multilinguality. In this research, we first fine-tune pre-trained XLM-RoBERTa (XML-R) with the SemEval-2017 Task 4 A dataset and then use zero-shot cross lingual transfer learning to test Hindi sentiment dataset. That is, the classification model is trained on a resource rich English language using XML-R cross-lingual transformer model and transfer learning is done on the resource-poor Hindi language. The model consists of two components: (1) Sentiment classification component for both languages and (2) cross-lingual transfer learning component that uses English sentiment analysis data to predict positive, negative and neutral sentiment class in resource-poor Hindi.

We use the XLM-R_{large} model. The model contains approximately 355M parameters with 24-layers, 1,027 hidden-states, 4,096 feed-forward hidden-states and 16-heads. The maximum sequence length can be set to default, 512, that is, it takes an input of a sequence of no more than 512 tokens and outputs the representation of the sequence. The first token of the sequence is always [CLS], which contains the special classification embedding [39]. The first token of every sequence is always a special classification token ([CLS]). The final hidden state, h, corresponding to this token is used as the aggregate sequence representation for classification tasks. A simple softmax classifier is added to the top of XLM-R to predict the probability of label c, as shown in Equation (1), where W is the task-specific parameter matrix:

$$(c|h) = softmax(W\mathrm{h}). \tag{1}$$

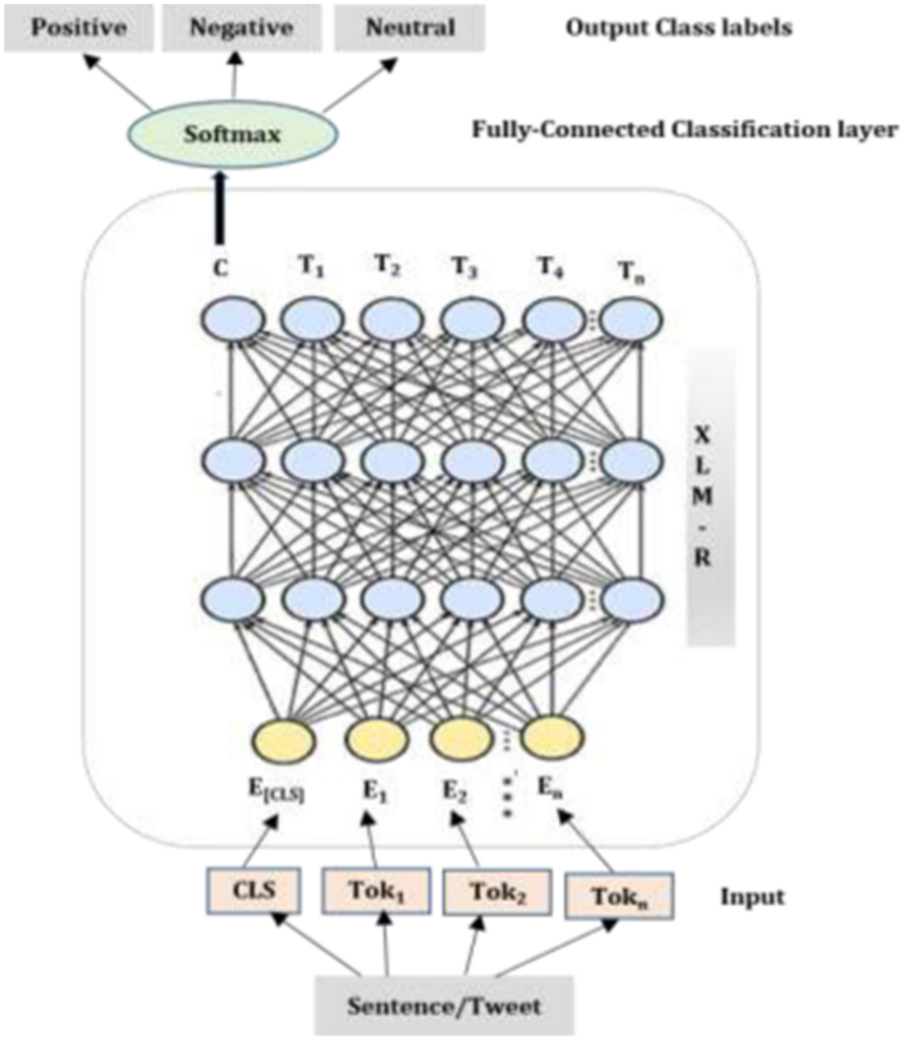


Fig. 3. Sentiment classification architecture using XLM-R.

We fine-tune all the parameters from XLM-R as well as W jointly by maximizing the log-probability of the correct label. The architecture of the sentiment classification is shown in Figure 3.

3.4 Zero-shot Cross-lingual Transfer Learning

From a psychological point of view, transfer of learning is the "study of dependency of human conduct, learning, or performance on prior experience." Analogously, in machine learning, transfer learning is the ability of a system to recognize and apply knowledge and skills learned in previous domains/tasks to novel tasks/domains, which share some commonality. Building every model from scratch is time consuming and expensive in terms of data collection/labeling, privacy and training time whereas reusing common knowledge extracted from existing systems is a viable solution. Transfer learning differs from traditional machine learning, because it involves using a pre-trained model as a springboard to start a secondary task. A pre-trained model is a saved network that was previously trained on a large dataset, typically on a large-scale sentiment classification task. Advantages of using transfer learning include less training data with better generalization capabilities make deep learning more accessible, faster convergence and higher asymptomatic accuracy. In transfer learning, we reuse by freezing or fine tuning the layers of a model. Domain adaptation, domain confusion, multi-task learning, one-shot learning and zero-shot learning are various

types of transfer learning. For training and testing a learning model, we need data and the number of samples of a class required in data for learning are shots for that class. In zero-shot the machine is capable of describing what class an unlabeled sample belongs to when it does not fall into the category of any of the trained categories, that is, zero shots for the data point. In this work, we utilize zero-shot cross-lingual learning [40] to solve a sentiment analysis task in Hindi despite not having received any training examples of the task. Zero-shot means that the cross-lingual XLM-R is fine-tuned on one language and then evaluated on the foreign language test. In this case, machine translation was not involved at all in either the pre-training or fine-tuning.

The concept of transfer learning allows a novel algorithm to inherit the knowledge (features, weights) of the existing algorithm [41, 42]. That is, we use what well-trained, well-constructed networks have learned over large sets, and apply them to boost the performance of a detector on a smaller data set. Deep neural models are centered on network weights that come from training them on huge amount of data. That is, if the weights in the existing model are obtained and transferred to the new neural model, then it implies the "transfer" of learned features without building the model from scratch and retraining it. The object in the transfer method has both the network structure and weights. For adopting transfer learning to cross-lingual sentiment analysis two methods describing what to transfer can be used:

- *Transfer learning method 1 (TLM1):* Network is pre-trained on source task and the full structure (all layers), and all weights of network are transferred to the second network. That is, it starts with all the previously trained, fine-tuned weights from source, and then test it for the target. (Freeze the base.)
- *Transfer learning method 2 (TLM 2):* Network is pre-trained on source task and the partial structure (some layers), and their weights are transferred to the second network. The transferred weights are frozen, whereas the non-transferred weights are randomly initialized for a second training phase for the target task. Fine-tuning is done. (Train some layers, leave some frozen, i.e., selectively unfreeze.)

In this work, we use TLM1, where the XLM-R classification model is trained using the English language Benchmark SemEval 2017 dataset Task 4A. The complete network structure including the fully-connected softmax layer is adopted along with all weights to initialize the weights for the Hindi language sentiment classification model. Basically, we freeze the weights of pre-trained model in one language and use them for extracting the features in the other. In Keras, each layer has a parameter called "trainable". For freezing the weights of a particular layer, we should set this parameter to False, indicating that this layer should not be trained. To examine TLM1, we experiment on both the IIT-Patna Movie and Product review datasets, which has four sentiment classes, namely, positive, negative, and neutral or conflict, but as discussed, we choose to ignore the conflict class to ensure the same number of classes in the target dataset. If we consider different number of output classes to predict in the target language, then TLM2 is a better strategy where the last softmax layer is detached from the pre-trained base XLM-R. Therefore, in this case new fully-connected layers according to the number of classes in the target dataset are added with randomized weights (with frozen weights of re-trained model) and updated by training the classifier layers on training data available for task. It involves hyperparameter tuning and may require unfreezing more layers as needed.

4 MODEL PERFORMANCE

4.1 Experimental Setup

The pre-trained model that is used were made available online using the transformers library (Wolf et al., 2019) and the same have been fine-tuned for the training of the model. To examine

Table 2. Hyperparameters

Hyperparameter	Value
Epochs	3
Batch Size	32
Learning Rate	9e-06
Max Sequence Length	64
Warmup Proportion	0.1
Optimizer	Adam

Table 3. Performance of XLM-R_{large} on SemEval2017-Task4A

Model	AvgRec	$F1^{PN}$	Acc
BB_twtr [44] (Leader SemEval-2017, TaskA)	68.1	68.5	65.8
DataStories [45] (Leader SemEval-2017, TaskA)	68.1	67.7	65.1
State-of-the-art [43]	73.2	72.8	71.7
XLM-R_{large}	**73.5**	72.3	**71.8**

the effectiveness of our classification model, we tested it on two datasets, namely, the IIT-Patna Movie and Product review datasets. The training dataset was SemEval-2017 Task A. We divided the SemEval 2017 Task A English dataset into a training set and a validation set using 80:20 split and primarily fine-tuned the learning rate and number of epochs of the classification model manually to obtain the best results for the validation set. Training XLM-R on the SemEval-17 data required nearly 30 min. Nvidia Tesla K80 GPU to train the models. Early stopping was done if the evaluation loss did not improve over ten evaluation rounds. Table 2 depicts the hyperparameter setting.

4.2 Baselines and Evaluation

To ensure that the XLM-R achieves state-of-the-art performance, we compared its effectiveness to the existing best reported performance on both the Hindi language datasets. The metrics used for evaluation were average recall, $F1^{PN}$ and Macro F-score. The best performing system on the SemEval 2017 task (English language) dataset is the one described in Reference [43], which achieved a macro-average recall of 73.2. There were two winners [44, 45] of the SemEval 2017 Task A with a macro-average recall of 68.1. The focus of this research was to implement XML-R for sentiment analysis in Hindi language (non-English data) and not to outperform the SemEval-2017 state-of-the-art (SOTA). Interestingly, we achieved superlative results than the winners of task and also better results to the SOTA. Table 3 depicts the comparison of SemEval-2017 sentiment analysis task on English language.

Akhtar et al. [38] reported an average performance accuracy of 51.55 on both the Hindi datasets. They proposed a hybrid deep learning CNN-SVM (W+X) model trained and evaluated using word embeddings and extra hand-crafted features. The SOTA was recently given by Kunchukuttan et al. [16], reporting a 54.64 average accuracy on both datasets. The transfer learning with XLM-R cross-lingual embeddings provided the best results achieving an average performance accuracy of 60.93 on both the Hindi datasets. Table 4 depicts the comparison of IIT-P: Movie and product sentiment analysis datasets on Hindi language.

Figures 4 and 5 depict the precision-recall bar graph for both the Hindi datasets.

We take a closer look at the predictions made by XLM-R with zero-shot TL with the help of heatmap visualization for both the Hindi datasets as shown in Figures 6 and 7. The rows show the

Table 4. Accuracy of XLM-R_{large} with Zero-shot TL on IIT-P, Movie and Product Hindi Dataset

Model	IITP-Movie$_{Acc}$	IITP-Product$_{Acc}$
Pre-trained word embeddings using the IndicNLP corpus (SOTA) [16]	45.81	63.48
Review$_{SH}$ and Movie$_{H}$ [38]	44.88	57.34
XLM-R_{large} with Zero-shot TL	**51.74**	**70.12**

Fig. 4. Precision-recall of IITP-movie dataset.

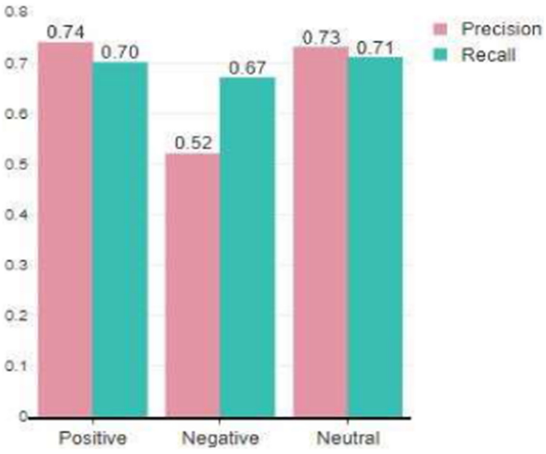


Fig. 5. Precision-recall of IITP-product dataset.

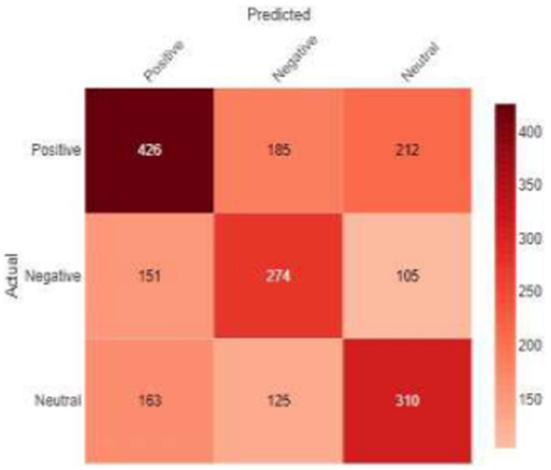


Fig. 6. HeatMap confusion matrix of IITP-movie dataset.

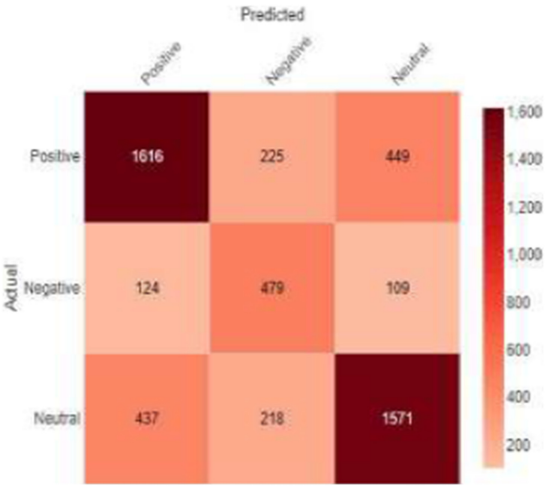


Fig. 7. HeatMap confusion matrix of IITP-product dataset.

actual class of a repetition and columns show the classifier's prediction of the respective confusion matrix.

5 CONCLUSION

Huge volumes of unstructured, multi-lingual, and multi-modal data is created on social media every day. Undeniably, social intelligence on these platforms is continuously increasing but the heterogeneity of content makes it hard to analyze and understand sentiments. This work uses the

existing cross-lingual transformer model, XLM-R as the pre-training model to provide an effective solution to sentence-level (tweet-level) analysis of sentiments in a resource-poor scenario with the help of zero-shot transfer learning. Both the network structure and weights are fine-tuned on benchmark English dataset and tested on two benchmark Hindi datasets. Interestingly, XLM-R achieves an average recall of 73.5 and accuracy of 71.8 for the English benchmark dataset, which is superlative to the best performing models. XLR-M outperforms the SOTA on Hindi datasets too. Future work included considering more Hindi test datasets for classification. Also, currently the training and test dataset are for sentence-level sentiment analysis, evaluation of XLM-R for aspect-based sentiment analysis is also a prominent direction of future research. Last, the model can be tested on other regional Indian languages such as Kannada, Marathi, or Sindhi.

REFERENCES

[1] Bing Liu. 2012. Sentiment analysis and opinion mining. *Synth. Lect. Hum. Lang. Technol.* 5, 1 (2012), 1–167.

[2] Akshi Kumar, Kathiravan Srinivasan, Wen-Huang Cheng, and Albert Y. Zomaya. 2020. Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data. *Info. Process. Manage.* 57, 1 (2020), 102141.

[3] Akshi Kumar and Arunima Jaiswal. 2020. A deep swarm-optimized model for leveraging industrial data analytics in cognitive manufacturing. *IEEE Trans. Industr. Info.* 17, 4 (2020), 2938–2946. doi: 10.1109/TII.2020.3005532

[4] Santwana Chimalamarri, Dinkar Sitaram, and Ashritha Jain. 2020. Morphological segmentation to improve crosslingual word embeddings for low resource languages. *ACM Trans. Asian Low-Resource Lang. Info. Process.* 19, 5 (2020), 1–15. https://doi.org/10.1145/3390298

[5] Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2018. Meld: A multimodal multi-party dataset for emotion recognition in conversations. Retrieved from https://arXiv:1810.02508.

[6] Md Shad Akhtar, Asif Ekbal, and Erik Cambria. 2020. How intense are you? predicting intensities of emotions and sentiments using stacked ensemble. *IEEE Comput. Intell. Mag.* 15, 1 (2020), 64–75.

[7] Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub et al. 2016. Semeval-2016 task 5: Aspect-based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval'16)*. 2016.

[8] Ning Liu and Bo Shen. 2020. Aspect-based sentiment analysis with gated alternate neural network. *Knowl.-Based Syst.* 188 (2020), 105010.

[9] Fazel Keshtkar and Diana Inkpen. 2009. Using sentiment orientation features for mood classification in blogs. In *Proceedings of the International Conference on Natural Language Processing and Knowledge Engineering*. IEEE, 1–6.

[10] Mahmoud Al-Ayyoub, Abed Allah Khamaiseh, Yaser Jararweh, and Mohammed N. Al-Kabi. 2019. A comprehensive survey of arabic sentiment analysis. *Info. Process. Manage.* 56, 2 (2019), 320–342.

[11] Majdi Beseiso and Haytham Elmousalami. 2020. Subword attentive model for Arabic sentiment analysis: A deep learning approach. *ACM Trans. Asian Low-Resource Lang. Info. Process.* 19, 2 (2020), 1–17.

[12] Asad Khattak, Muhammad Zubair Asghar, Anam Saeed, Ibrahim A. Hameed, Syed Asif Hassan, and Shakeel Ahmad. 2021. A survey on sentiment analysis in Urdu: A resource-poor language. *Egypt. Info. J.* 22, 1 (2021), 53–74.

[13] Valentin Barriere and Alexandra Balahur. 2020. Improving sentiment analysis over non-english tweets using multilingual transformers and automatic translation for data-augmentation. Retrieved from https://arXiv:2010.03486.

[14] Wenhuan Wang, Bohan Li, Ding Feng, Anman Zhang, and Shuo Wan. 2020. The OL-DAWE Model: Tweet polarity sentiment analysis with data augmentation. *IEEE Access* 8 (2020), 40118–40128.

[15] De Leon, Frances Adriana Laureano, Florimond Guéniat, and Harish Tayyar Madabushi. 2020. CS-embed-francesita at semeval-2020 Task 9: The effectiveness of code-switched word embeddings for sentiment analysis. Retrieved from https://arXiv:2006.04597.

[16] Anoop Kunchukuttan, Divyanshu Kakwani, Satish Golla, Avik Bhattacharyya, Mitesh M. Khapra, and Pratyush Kumar. 2020. AI4Bharat-IndicNLP Corpus: Monolingual corpora and word embeddings for indic languages. Retrieved from https://arXiv:2005.00085.

[17] K Karthikeyan, Zihan Wang, Stephen Mayhew, and Dan Roth. 2020. Cross-lingual ability of multilingual BERT: An empirical study. In *Proceedings of the International Conference on Learning Representations (ICLR'20)*.

[18] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. Retrieved from https://arXiv:1911.02116.

[19] Kumar Akshi and Geetanjali Garg. 2019. Systematic literature review on context-based sentiment analysis in social multimedia. *Multimedia Tools Appl.* 79, 21 (2019), 15349–15380.

[20] Jean-Benoit Delbrouck, Noé Tits, Mathilde Brousmiche, and Stéphane Dupont. 2020. A transformer-based joint-encoding for emotion recognition and sentiment analysis. Retrieved from https://arXiv:2006.15955.

[21] José Ángel González, Lluís-F. Hurtado, and Ferran Pla 2020. Self-attention for Twitter sentiment analysis in Spanish. *J. Intell. Fuzzy Systems* 39, 2 (2020), 2165–2175.

[22] Ahmed Sultan, Mahmoud Salim, Amina Gaber, and Islam El Hosary. 2020. WESSA at SemEval-2020 Task 9: Code-mixed sentiment analysis using transformers. Retrieved from https://arXiv:2009.09879.

[23] Y Kuratov, M. Arkhipov. 2019. Adaptation of deep bidirectional multilingual transformers for Russian language. Retrieved from https://arXiv:1905.07213.

[24] Anindya Sarkar, Sujeeth Reddy, and Raghu Sesha Iyengar. 2019. Zero-shot multilingual sentiment analysis using hierarchical attentive network and BERT. In *Proceedings of the 3rd International Conference on Natural Language Processing and Information Retrieval (NLPIR'19)*. Association for Computing Machinery, New York, NY, 49–56. DOI : https://doi.org/10.1145/3342827.3342850

[25] Avishek Garain, Sainik Kumar Mahata, and Dipankar Das. 2020. JUNLP@ SemEval-2020 Task 9: Sentiment analysis of Hindi-English code mixed data using grid search cross validation. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*. 1276–1280. https://arxiv.org/abs/2007.12561.

[26] Somnath Banerjee, Sahar Ghannay, Sophie Rosset, Anne Vilnat, and Paolo Rosso. 2020. LIMSI_UPV at SemEval-2020 Task 9: Recurrent convolutional neural network for code-mixed sentiment analysis. Retrieved from https://arXiv:2008.13173.

[27] Parth Patwa, Gustavo Aguilar, Sudipta Kar, Suraj Pandey, Srinivas PYKL, Björn Gambäck, Tanmoy Chakraborty, Thamar Solorio, and Amitava Das. 2008. Semeval-2020 task 9: Overview of sentiment analysis of code-mixed tweets. Retrieved from https://arxiv.org/abs/2008.04277.

[28] Yaman Kumar, Debanjan Mahata, Sagar Aggarwal, Anmol Chugh, Rajat Maheshwari, Rajiv Ratn Shah. 2019. BHAAV—A text corpus for emotion analysis from Hindi stories. Retrieved from https://arXiv:1910.04073.

[29] Kanika Garg and D. K. Lobiyal. 2020. Hindi EmotionNet: A scalable emotion lexicon for sentiment classification of Hindi text. *ACM Trans. Asian Low-Resource Lang. Info. Process.* 19, 4 (2020), 1–35.

[30] A. Bakliwal, P. Arora, and V. Varma. 2012. Hindi subjective lexicon: A lexical resource for Hindi polarity classification. *Int. J. Comput. Linguist. Appl. (IJCLA)* 2012

[31] A Balamurali, R. Joshi, A, and P. Bhattacharyya. 2012. Cross-lingual sentiment analysis for Indian languages using linked wordnets. In *Proceedings of the International Conference on Computational Linguistics (COLING'12)*.

[32] Braja Gopal Patra, Dipankar Das, Amitava Das, and Rajendra Prasath. 2015. Shared task on sentiment analysis in Indian languages (SAIL) tweets—An overview. In *Proceedings of the International Conference on Mining Intelligence and Knowledge Exploration (MIKE'15)*. Springer.

[33] Se Shriya, R. Vinaya Kumar, M. Anand Kumar, and K. P. Soman. 2015. AMRITA-CEN@SAIL2015: Sentiment analysis in Indian languages. In *Proceedings of the International Conference on Mining Intelligence and Knowledge Exploration (MIKE'15)*. Springer.

[34] A. Kumar, S. Kohail, A. Ekbal, and C. Biemann. 2015. IIT-TUDA: System for sentiment analysis in Indian languages using lexical acquisition. In *Proceedings of the International Conference on Mining Intelligence and Knowledge Exploration (MIKE'15)*. Springer.

[35] M. S. Akhtar, A. Ekbal, and P. Bhattacharyya. 2016. Aspect-based sentiment analysis in Hindi: Resource creation and sentiment classification. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC'16)*.

[36] Shad Akhtar, Palaash Sawant, Sukanta Sen, Asif Ekbal, and Pushpak Bhattacharyya. 2018. Solving data sparsity for aspect-based sentiment analysis using cross-linguality and multi-linguality. In *Proceedings of the 16th Annual Conference of the NAACL on Human Language Technologies (HLT'18)*. 572–582.

[37] Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. 2015, Bilingual word representations with monolingual quality in mind. In *Proceedings of the NAACL Workshop on Vector Space Modeling*.

[38] M. S. Akhtar, A. Kumar, A. Ekbal, and P. Bhattacharyya. 2016. A hybrid deep learning architecture for sentiment analysis. In *Proceedings of the 26th International Conference on Computational Linguistics (COLING'16)*. 482–493.

[39] Chi Sun, Xipeng Qiu, Yige Xu, and Xuanjing Huang. 2019. How to fine-tune BERT for text classification? In *Chinese Computational Linguistics, Maosong Sun*, Xuanjing Huang, Heng Ji, Zhiyuan Liu, and Yang Liu (Eds.). 194–206

[40] Anne Lauscher, Vinit Ravishankar, Ivan Vulić, and Goran Glavaš. 2020. From zero to hero: On the limitations of zero-shot cross-lingual transfer with multilingual transformers. Retrieved from https://arXiv:2005.00633.

[41] Jie, Tao and Xing Fang. 2020. Toward multi-label sentiment analysis: A transfer learning-based approach. *J. Big Data* 7, 1 (2020), 1–26.

[42] Sultan Ahmed, Mahmoud Salim, Amina Gaber, and Islam El Hosary. 2020. WESSA at SemEval-2020 Task 9: Code-mixed sentiment analysis using transformers. Retrieved from https://arXiv:2009.09879.

[43] Dat Quoc Nguyen, Thanh Vu, and Anh Tuan Nguyen. 2020. BERTweet: A pre-trained language model for English tweets. Retrieved from https://arXiv:2005.10200.

[44] Mathieu Cliche. 2017. BB twtr at SemEval-2017 Task 4: Twitter sentiment analysis with CNNs and LSTMs. *Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval'17)*. 573–580.

[45] C. Baziotis, N. Pelekis, and C. Doulkeridis. 2017. DataStories at SemEval-2017 Task 4: Deep LSTM with attention for message-level and topic-based sentiment analysis. *Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval'17)*. 747–754.

Received October 2020; revised March 2021; accepted April 2021

Journal of Applied Mathematics and Computing
https://doi.org/10.1007/s12190-021-01601-1

ORIGINAL RESEARCH

Stability analysis of a logistic growth epidemic model with two explicit time-delays, the nonlinear incidence and treatment rates

Kanica Goel[1] · Abhishek Kumar[2] · Nilam[1]

Received: 4 April 2021 / Revised: 28 June 2021 / Accepted: 13 July 2021
© Korean Society for Informatics and Computational Applied Mathematics 2021

Abstract

In the present study, a time-delayed SIR epidemic model with a logistic growth of susceptibles, Crowley–Martin type incidence, and Holling type III treatment rates is proposed and analyzed mathematically. We consider two explicit time-delays: one in the incidence rate of new infection to measuring the impact of the latent period, and another in the treatment rate of infectives to analyzing the effect of late treatment availability. The stability behavior of the model is analyzed for two equilibria: the disease-free equilibrium (DFE) and the endemic equilibrium (EE). We derive the threshold quantity, the basic reproduction number R_0, which determines the eradication or persistence of infectious diseases in the host population. Using the basic reproduction number, we show that the DFE is locally asymptotically stable when $R_0 < 1$, linearly neutrally stable when $R_0 = 1$, and unstable when $R_0 > 1$ for the time-delayed system. We analyze the system without a latent period, revealing the forward bifurcation at $R_0 = 1$, which implies that keeping R_0 below unity can diminish the disease. Further, the stability behavior for the EE is investigated, demonstrating the occurrence of oscillatory and periodic solutions through Hopf bifurcation concerning every possible grouping of two time-delays as the bifurcation parameter. To conclude, the numerical simulations in support of the theoretical findings are carried out.

✉ Abhishek Kumar
abhishek.dtu14@gmail.com

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

Kanica Goel
kanica.dtu2016@gmail.com

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

[2] Department of Mathematics, School of Basic Sciences and Research, Sharda University, Greater Noida 201310, India

Published online: 21 July 2021

Keywords Logistic growth rate · Two explicit time delays · Nonlinearity · Stability · Bifurcations · Numerical simulations

Mathematics Subject Classification 34D20 · 92B05 · 37M05

1 Introduction

Infectious diseases are significant reasons for affliction and mortality in developing nations. Therefore, understanding the transmission attributes of irresistible ailments and the surveillance and continuation of control interventions are necessary to maintain infectious disease control achievements. Mathematical modeling is one of the powerful tools to provide valuable insight in applied areas of epidemiology and ecology [1–8]. The compartmental mathematical modeling of infectious diseases is a powerful tool that focuses on predicting, assessing, and controlling potential outbreaks. Kermack and McKendrick [9] proposed a compartmental SIR (susceptible–infected–recovered) epidemic model, and following up the work of Kermack and McKendrick, many researchers studied epidemic models with different modifications (for instance, SIS [10], SIR [11–16], SIRS [17], SVEIR [18], SEIR [19], SVIRS [20], SAAIR [21], SFIR [22], etc.).

An incidence rate is a valuable tool that determines the number of new infection cases per unit time. A significant assumption in disease transmission models is how individuals make contact with each other. The classical Kermack and McKendrick epidemic model [9] addresses the mass-action incidence rate in the disease transmission cycle, suggesting that the probability of infection of a susceptible is analogous to the number of contacts with infected individuals, which is not realistic for a large population. Therefore, to model the realistic scenario, several authors [16,19,22–25] focused on considering nonlinear incidence rates such as Holling type II, Monod–Haldane, Beddington–DeAngelis, etc. Crowley and Martin [26] introduced an incidence rate of the form $\frac{\beta SI}{(1+\alpha S)(1+\gamma I)}$, which includes the forms of other incidences also, for instance, bilinear incidence rate if $\alpha = 0, \gamma = 0$; saturated incidence rate with susceptibles if $\gamma = 0$; Holling-II incidence rate if $\alpha = 0$. This incidence rate has the important property of including preventive anti-epidemic measures adopted by susceptible and infected populations (such as disease surveillance and hygienic measures, social distancing, travel restrictions, quarantine, case isolation, etc.). It normalizes the influence of inhibition on infectives even in the high density of susceptible populations [15,25]. When an outbreak happens, then the human behavior towards that disease is an important factor that influences disease emergence events. If individuals adopt preventive measures against infectious disease, it can change the transmission pattern of the disease's spread. Preventive anti-epidemic measures describe human behavior towards the disease, aiming to prevent the occurrence and slow down the spread of infections in a population. Moreover, in the case of newly emerging diseases, it will often be impractical to provide treatment to all the infected individuals as the accessibility of the resources such as the hospital's condition, medicines, doctor's availability, etc. may be restricted. Thus, adopting inhibition measures is a long-established public health strategy for controlling the spread of diseases. When a disease emerges, there

is a time lag between the exposure and identification of the symptoms. That time lag can be modeled by including the latency period in the incidence function. It has a significant role in formulating the epidemic model. The higher the time delay in recognizing the disease, the higher the deferral in executing preventive measures. So, the consideration of the latent period leads to a more natural phenomenon of disease dynamics. Therefore, some authors have introduced the latent period and studied the disease dynamics [10,20,21,27,28].

Treatment is the most important medical intervention for reducing disease spread and deaths during an outbreak. The adequacy and effectiveness of the therapy may influence the recovery pace of infectives [29]. In classical epidemic models, the treatment rate of infectives is assumed to be proportional to the number of the infectives, which is not suitable for a large population of infectives as the resources for treatment should be quite large. Thus, it is vital to determine a reasonable capacity for the treatment of a disease. Therefore, to simulate a limited availability of treatment, Wang and Ruan [30] adopt a constant treatment. Wang and Ruan [31] further modified the constant treatment rate and considered an epidemic model with a treatment rate of the form

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

This treatment rate is proportional to the number of the infectives when the treatment capacity is not reached and otherwise takes the maximal capacity. Zhou and Fan [32] studied a SIR epidemic model with an improved treatment rate. They considered a continuous and differentiable functional response of the form $h(I) = \frac{aI}{1+bI}$ $(I \geq 0,\ a \geq 0,\ b \geq 0)$, called Holling type II treatment rate, which saturates at its maximum value. During the disease outbreak, the hospital facilities, including the number of health care workers, medical equipment, the number of hospital beds, and medicines available to the public, maybe inadequate. Therefore, to capture these factors, Dubey et al. [25] investigated the disease dynamics of an epidemic model by considering Holling type III treatment rate, which is a continuously differentiable increasing function of I, having the form $\frac{aI^2}{(1+bI^2)}$, approaches to its maximum value (a/b) due to the limited availability of resources when I is large.

In the beginning phase of an outbreak, it follows an exponential or generalized - exponential growth. However, it slows down and reaches its maximum, as the increasing number of cases reaches its inflection point, and the daily incidence curve reaches its most extreme. Thus, the growth pattern departs from the (sub-)exponential path and follows a logistic growth rate [33]. To capture this situation, some authors studied the effect of the logistic growth rate in their epidemic models [12–14,34]. Wang et al. [13] assumed that the logistic growth governed the population growth in susceptible individuals and proposed the epidemic model with bilinear incidence rate, including the incubation period. They investigated the stability behavior and showed the existence of Hopf bifurcation in their model. Zhang et al. [12] also studied the logistic growth SIR epidemic model with the saturated incidence rate, including the incubation time and the inhibitory effects of infectives, and studied the dynamical behavior of their model. Xue et al. [14] extended the work of Zhang et al. [12] by incorporating the impact

of the latent period in the saturated incidence rate and obtained sufficient conditions for the Hopf bifurcations. Further, Li et al. [34] considered the logistic growth SIR epidemic model with bilinear incidence and saturated treatment rates and studied the stability analysis and various bifurcations. They showed that to eradicate the disease, the government should raise the efficiency and enlarge treatment capacity.

Although many authors have studied models with time delay or saturated treatment, to the best of our knowledge, there are not many models that incorporate time delays in both nonlinear incidence and saturated treatment functions. Motivated by Eric et al. [35], who incorporated a nonlinear incidence rate and a logistic growth rate into a SIR epidemic model and other subsequent works, such as [15,21,35], in which authors studied delayed epidemic models with different functional responses, in the present article, we study a susceptible-infected-recovered (SIR) epidemic model with two explicit time-delays, the logistic growth of susceptible individuals, Crowley–Martin incidence rate, and the Holling type III treatment rate. The logistic growth of susceptibles comprises the increasing growth of the epidemic in the beginning period, but a decreasing development at a later stage as the number of infections approaches its maximum. To measure the transmission of infection in susceptibles individuals with the condition of taking protective anti-epidemic measures by both susceptibles and infected individuals, and the time-lag between being exposed to a disease and having its symptoms, we have incorporated the Crowley–Martin incidence rate with the latent period. To study the impact of treatment on infected individuals with limited medical resources and delay in providing the appropriate therapy to infectives, we have considered the Holling type III treatment rate with time-delay. The two explicit time delays have a significant role in the present article. A delay in identifying the disease leads to a delay in applying the right protection measures, which allow the virus to spread unseen. If the infection remains undiagnosed after the recognition, this time lag enables the virus to reach capital cities, where the outbreak grows into large epidemics. Therefore, the latent period and the delay in providing medical services to infectives let the disease stay for an extended period and spread faster at a higher rate. By incorporating the above-enlightened facts, we formate and analyze a mathematical epidemic model. To explore the disease dynamics through equilibrium analysis, we investigate the stability behavior of the disease-free and endemic equilibria by deriving the basic reproduction number R_0. This threshold quantity determines the eradication or persistence of the disease infection in a host population. The results indicate that for $R_0 < 1$, there is a unique disease-free equilibrium, which is locally asymptotically stable; this equilibrium becomes unstable when $R_0 > 1$. Meanwhile, if R_0 is greater than one, then there exists an endemic equilibrium. The stability switches may happen near this equilibrium for different ranges of explicit time delays, and the oscillatory and periodic solutions may appear via Hopf bifurcation. We discuss the significance of two explicit time delays and other parameters by simulating the proposed model numerically, revealing the disease's endemic behavior, and suggesting its control strategies.

The rest of the paper is structured as follows: Sect. 2 is devoted to the model formulation and its basic properties. Section 3 obtains the disease-free and endemic equilibria and performs the stability analysis for both the equilibria using the derived basic reproduction number. Further, the stability changes of endemic equilibrium are

shown via Hopf bifurcation regarding every possible combination of the two time-delays as a bifurcation parameter. In Sect. 4, numerical simulations are performed to justify the analytical results. In the end, a brief discussion and conclusion are given in Sect. 5.

2 Model description and basic properties

We assume that the total population $N(t)$ at time t is divided into three catogories: susceptible $S(t)$, infected $I(t)$ and $R(t)$, respectively. The description of the model's parameters is as follow: The logistic growth $rS(t)\left(1-\frac{S(t)}{K}\right)$ governs the susceptible population with a carrying capacity K and a specific growth rate r. The term $\frac{\beta S(t-\varrho_1)I(t-\varrho_1)}{(1+\alpha S(t-\varrho_1))(1+\gamma I(t-\varrho_1))}$ represents the Crowley–Martin incidence rate with latent period ϱ_1, where β denotes the transmission rate of infection, and α and γ denote the saturating factors representing the inhibitory effects adopted by susceptibles and infectives, respectively. ϑ, d, and θ are the natural mortality, disease-induced mortality, and recovery rate of infectives, respectively. The term $\frac{\sigma I^2(t-\rho_2)}{1+\xi I^2(t-\rho_2)}$ represents the Holling type III treatment rate, where σ is the cure rate (or the maximum treatment rate), ξ is the saturating rate that measures the effect of limited availability of treatment resources, and ρ_2 measures the time delay in giving treatment to infectives. Due to epidemiological reasons, we assume that σ is non-negative, and all other parameters are positive.

Thus, the model under the assumptions and descriptions mentioned above follows the following system of delay differential equations:

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= rS\left(1-\frac{S}{K}\right)-\frac{\beta S(t-\rho_1)I(t-\rho_1)}{(1+\alpha S(t-\rho_1))(1+\gamma I(t-\rho_1))},\\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \frac{\beta S(t-\rho_1)I(t-\rho_1)}{(1+\alpha S(t-\rho_1))(1+\gamma I(t-\rho_1))}-(\vartheta+\theta+d)I-\frac{\sigma I^2(t-\rho_2)}{1+\xi I^2(t-\rho_2)},\\
\frac{\mathrm{d}R}{\mathrm{d}t} &= \theta I+\frac{\sigma I^2(t-\rho_2)}{1+\xi I^2(t-\rho_2)}-\vartheta R
\end{aligned} \tag{1}$$

Let $h=\max\{\rho_1,\rho_2\}$. The initial conditions of the model (1) are given by

$$S(\Theta)=\Phi_1(\Theta),\ I(\Theta)=\Phi_2(\Theta),\ R(\Theta)=\Phi_3(\Theta),\ \text{for } \Theta\in[-h,0], \tag{2}$$

where $\Phi_i\in C\left([-h,0],\mathbb{R}^3_+\right)$, $i=1,2,3$, the space of continuous functions from $[-h,0]$ to $\mathbb{R}^3_+$, equipped with the sup-norm: $\|\Phi\|=\sup_{\Theta\in[-h,0]}\Phi(\Theta)$.

By the theory of functional differential equations [36], it follows that all the solutions of model (1) with initial condition (2) exists and are differentiable for all $t\geq 0$.

It is noticed that the variable $R(t)$ is not present in the first two equations of the model (1); therefore, without loss of generality, for the analysis purpose, we can omit the third equation and study the following reduced system:

$$\begin{aligned}\frac{\mathrm{d}S}{\mathrm{d}t} &= rS\left(1-\frac{S}{K}\right)-\frac{\beta S(t-\rho_1)I(t-\rho_1)}{(1+\alpha S(t-\rho_1))(1+\gamma I(t-\rho_1))},\\ \frac{\mathrm{d}I}{\mathrm{d}t} &= \frac{\beta S(t-\rho_1)I(t-\rho_1)}{(1+\alpha S(t-\rho_1))(1+\gamma I(t-\rho_1))}-(\vartheta+\theta+d)I-\frac{\sigma I^2(t-\rho_2)}{1+\xi I^2(t-\rho_2)}.\end{aligned} \tag{3}$$

Lemma 1 *Let* $\nu_m = \min\{1, \vartheta+\theta+d\}$ *and* $\Omega = \max\{S(0), K\}$. *Then the compact set*

$$D = \left\{(S, I) \in \mathbb{R}^2_+ : S \le \Omega,\ S+I \le \frac{(r+1)\Omega}{\nu_m}\right\}$$

is a positively invariant set for the system (3).

Proof From the first equation of the system (3), we have

$$\frac{\mathrm{d}S}{\mathrm{d}t} \le rS\left(1-\frac{S}{K}\right).$$

Thus, a standard comparison argument gives that

$$\limsup_{t\to\infty} S(t) \le K.$$

On adding first and second equations of the system (1), we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(S+I) &= rS\left(1-\frac{S}{K}\right)-(\vartheta+\theta+d)I-\frac{\sigma I^2(t-\rho_2)}{1+\xi I^2(t-\rho_2)}\\ &\le (r+1)\,S-S-(\vartheta+\theta+d)I\\ &\le (r+1)\Omega-\nu_m\,(S+I)\,.\end{aligned}$$

Thus, we get

$$0 \le (S+I) \le \frac{(r+1)\,\Omega}{\nu_m} \text{ for sufficiently large } t.$$

Hence, all the solutions of the system (3) are closed and bounded and enter the region $D = \left\{(S, I) \in \mathbb{R}^2_+ : S \le \Omega,\ S+I \le \frac{(r+1)\Omega}{\nu_m}\right\}$.

Hence, the system (3) is well-posed mathematically and epidemiologically. □

3 Mathematical analysis of the model

In this section, the mathematical analysis of the model is established with the following steps: the existence of equilibria, the derivation of basic reproduction number, stability behavior of obtained equilibria, and the presence of Hopf bifurcation near-endemic equilibrium regarding the possible combinations of two time-delays.

3.1 Existence of model's equilibria

Substituting the right-hand side of the system (3) to zero reveals that it has two types of equilibrium points, namely:

(i) The disease-free equilibrium (DFE): $E_0\,(S_0, I_0) = (K, 0)$,

(ii) The endemic (positive) equilibrium (EE): $E_e\,(S^*, I^*)$, where S^* and I^* are evaluated in Sect. 3.3.

3.2 Disease-free equilibrium and its stability analysis

The disease-free equilibrium (DFE) is defined as the point at which no disease is present. The DFE of the system (3) lies at $E_0(S_0, I_0) = (K, 0)$.

The Jacobian matrix J of the system (3) at DFE E_0 is given as

$$J(E_0) = \begin{bmatrix} -r & -\frac{\beta K \mathrm{e}^{-\lambda \varrho_1}}{1+\alpha K} \\ 0 & \frac{\beta K \mathrm{e}^{-\lambda \varrho_1}}{1+\alpha K} - (\vartheta + d + \theta) \end{bmatrix}. \tag{4}$$

The characteristic equation associated with $J(E_0)$ is

$$(\lambda + r)\left(\lambda - \frac{\beta K \mathrm{e}^{-\lambda \varrho_1}}{1+\alpha K} + \vartheta + d + \theta\right) = 0. \tag{5}$$

The roots of Eq. (5) are $\lambda_1 = -r$ and λ_2, where λ_2 is a solution of the following transcendental equation:

$$T(\lambda) := \lambda - \frac{\beta K \mathrm{e}^{-\lambda \varrho_1}}{1+\alpha K} + \vartheta + d + \theta = 0. \tag{6}$$

The term $\frac{\beta K \mathrm{e}^{-\lambda \varrho_1}}{(1+\alpha K)(\vartheta+d+\theta)}$ at $\varrho_1 = 0$ is defined as the basic reproduction number (R_0) of the system (3), which is defined as the average number of secondary infections caused by a single infected agent, during his/her entire infectious period, in a completely susceptible population [37]. Thus, R_0 of the system (3) is

$$R_0 = \frac{\beta K}{(1+\alpha K)(\vartheta + d + \theta)}.$$

Now, we state and prove the following results using the obtained basic reproduction number R_0 for the local stability behavior of disease-free equilibrium $E_0(S_0, I_0)$.

Theorem 1 *The disease-free equilibrium $E_0(S_0, I_0)$ of the system (3) has the following properties:*

1. *When $R_0 > 1$, then $E_0(S_0, I_0)$ is unstable for $\varrho_1 > 0$.*

2. *When* $R_0 < 1$, *then* $E_0(S_0, I_0)$ *is locally asymptotically stable for* $\varrho_1 > 0$.
3. *When* $R_0 = 1$, *then* $E_0(S_0, I_0)$ *is linearly neutrally stable for* $\varrho_1 > 0$.
4. *When* $R_0 = 1$ *and* $\varrho_1 = 0$, *then the system* (3) *exhibits a forward bifurcation at* $E_0(S_0, I_0)$.

Proof We have

$$T(\lambda) := \lambda - \frac{\beta K \mathrm{e}^{-\lambda \varrho_1}}{1 + \alpha K} + \vartheta + d + \theta = 0.$$

The stability of the disease-free equilibrium E_0 is investigated for a different ranges of R_0 under the following cases:

1. Assume that $R_0 > 1$. Then, we have

$$\begin{aligned} T(0) &= -\frac{\beta K}{1 + \alpha K} + \vartheta + d + \theta \\ &= (\vartheta + d + \theta)(1 - R_0) \\ &< 0. \end{aligned}$$

   Also, $\lim_{\lambda \to \infty} T(\lambda) = +\infty$.

   Note that $T(\lambda) = 0$, $T(0) < 0$, $\lim_{\lambda \to \infty} T(\lambda) = +\infty$ and $T'(\lambda) = 1 + \frac{\beta K \rho_1}{1+\alpha K} \mathrm{e}^{-\lambda \rho_1} > 0$ imply that $T(\lambda)$ is a increasing function for real λ. Therefore, $T(\lambda) = 0$ has a real positive root when $R_0 > 1$ and hence, E_0 is unstable.
2. Assume that $R_0 < 1$. On the contrary, let $T(\lambda)$ has a root $\lambda_* \in \mathrm{C}$ with $Re(\lambda_*) \geq 0$. Then, we have $\lambda_* = \frac{\beta K \mathrm{e}^{-\lambda_* \varrho_1}}{1+\alpha K} - (\vartheta + d + \theta)$.
   So, $Re(\lambda_*) \leq \frac{\beta K}{1+\alpha K} - (\vartheta + d + \theta) = (\vartheta + d + \theta)(R_0 - 1) < 0$, which is a contradiction to our assumption. Thus, all the roots of $T(\lambda)$ has a negative real part, which proves that E_0 is locally asymptotically stable.
3. Let $R_0 = 1$. When $R_0 = 1$, then we have $\beta K = (1 + \alpha K)(\vartheta + d + \theta)$ and one root of $T(\lambda) = 0$ vanishes.
   Let $\lambda = m + \mathrm{i}n$ be any other solution of $T(\lambda) = 0$, Then, we have

$$m + \mathrm{i}n - \frac{\beta K \mathrm{e}^{-(m+\mathrm{i}n)\varrho_1}}{1 + \alpha K} + \vartheta + d + \theta = 0. \tag{7}$$

   Splitting real and imaginary parts of Eq. (7) and using $R_0 = 1$, we obtain:

$$(\vartheta + d + \theta)\mathrm{e}^{-m\varrho_1} \cos n\varrho_1 = m + \vartheta + d + \theta, \tag{8}$$

$$(\vartheta + d + \theta)\mathrm{e}^{-m\varrho_1} \sin n\varrho_1 = -n. \tag{9}$$

   Squaring Eqs. (8) and (9) and then summing up the resultant yields:

$$(\vartheta + d + \theta)^2 \mathrm{e}^{-2m\varrho_1} = (m + \vartheta + d + \theta)^2 + n^2. \tag{10}$$

   If there exists a root satisfying the Eqs. (8) and (9), then that root will satisfy Eq. (10) too. For Eq. (10) to be verified, we must have $m \leq 0$. Therefore, E_0 is linearly neutrally stable.

4. In case when $R_0 = 1$ and $\varrho_1 = 0$, then one of the eigenvalue of Jacobian matrix $J(E_0)$, given in Eq. (4), vanishes. So, the linearization method is not applicable, and we can resort to center manifold theory [38].
If we consider β as the bifurcation parameter, then the critical value β^* can be calculated as $\beta^* = \frac{(1+\alpha K)(\vartheta+d+\theta)}{K}$. We see that $J(E_0)$ has a simple zero eigenvalue $\lambda = 0$.
The components of right eigenvector $V = (V_1, V_2)^T$ and the left eigenvector of $U = (U_1, U_2)$ of $J(E_0)$ corresponding to $\lambda = 0$ are

$$V_1 = -\frac{\beta k}{r(\alpha k + 1)}, \quad V_2 = 1; \; U_1 = 0, \; U_2 = 1.$$

Let $g = (g_1, g_2)$ represents the right hand side of the system (3). Using Theorem 4.1 of [39], we calculate the bifurcation constants a_1 and b_1 as follows:

$$a_1 = \sum_{k,i,j=1}^{2} U_k V_i V_j \left(\frac{\partial^2 g_k}{\partial x_i \partial x_j}\right)_{E_0},$$
$$b_1 = \sum_{k,i=1}^{2} U_k V_i \left(\frac{\partial^2 g_k}{\partial x_i \partial \beta^*}\right)_{E_0}$$

where x_1 and x_2 represent the susceptible individuals S and the infected individuals I.
The non-zero partial derivatives associated with a_1 and b_1 are:

$$\frac{\partial^2 g_2(E_0, \beta^*)}{\partial S \partial I} = \frac{\partial^2 g_2(E_0, \beta^*)}{\partial I \partial S} = \frac{\beta^*}{(\alpha K + 1)^2}, \; \frac{\partial^2 g_2(E_0, \beta^*)}{\partial I^2} = -\frac{2\beta^* \gamma K}{\alpha K + 1} - 2\sigma, \text{ and}$$
$$\frac{\partial^2 g_2(E_0, \beta^*)}{\partial I \partial \beta^*} = \frac{K}{\alpha K + 1}.$$

So, a_1 and b_1 take the form

$$\begin{aligned} a_1 &= V_1 V_2 \left(\frac{\partial^2 g_2(E_0, \beta^*)}{\partial S \partial I}\right) + V_2 V_1 \left(\frac{\partial^2 g_2(E_0, \beta^*)}{\partial I \partial S}\right) + V_2^2 \left(\frac{\partial^2 g_2(E_0, \beta^*)}{\partial I^2}\right) \\ &= -\frac{2\beta^* k \left(\gamma r(\alpha k + 1)^2 + \beta^*\right)}{r(\alpha k + 1)^3} - 2\sigma, \\ b_1 &= U_2 V_2 \left(\frac{\partial^2 g_2(E_0, \beta^*)}{\partial I \partial \beta^*}\right), \\ &= \frac{K}{\alpha K + 1}. \end{aligned}$$

It can be seen that $a_1 < 0$ and $b_1 > 0$. Thus, using Theorem 4.1(iv) of [39], there exists a forward bifurcation at $R_0 = 1$. Forward bifurcation means that if

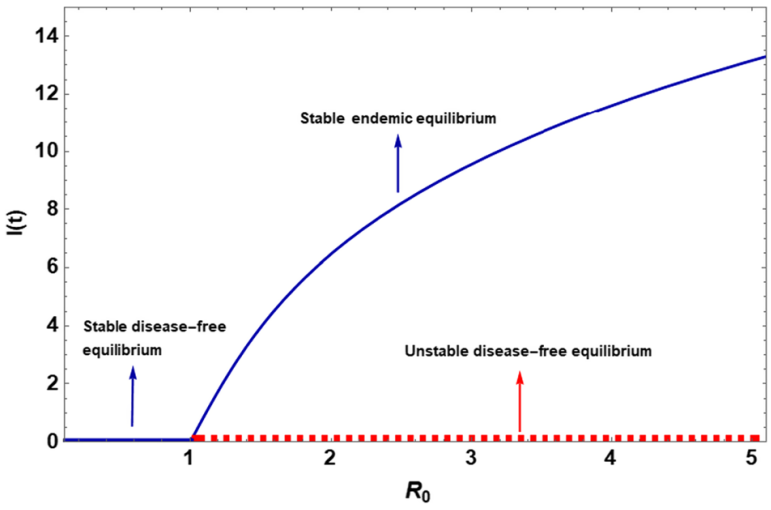


Fig. 1 Forward bifurcation

some infectives are introduced into the population then the system will return to the disease-free equilibrium $I = 0$. That is, disease will die out if $R_0 < 1$.

□

The graphical representation of the forward bifurcation is illustrated in the Example 1 below:

Example 1 Forward Bifurcation: We consider the following set of parameters: $r = 0.08$, $K = 100$, $\alpha = 0.002$, $\gamma = 0.001$, $\vartheta = 0.01$, $d = 0.008$, $\theta = 0.02$, $\sigma = 0.004$, $\xi = 0.009$. At these parameters' values, the bifurcation parameter is obtained as $\beta = \beta^* = 0.000456$. Figure 1 shows that when $R_0 < 1$ then, there is a stable disease free equilbrium, and when $R_0 > 1$, then the disease-free equilibrium becomes unstable and there exists a stable endemic equilibrium. It guarantees the eradication of disease when $R_0 < 1$, whereas the disease will invade when $R_0 > 1$.

3.3 Endemic equilibrium and stability analysis

To establish the existence of endemic equilibrium $E_e(S^*.I^*)$, first, we put the right-hand side term of the second equation of the system (3) to zero. We obtain:

$$S^* = \frac{(\gamma I^* + 1)\left(\left(\xi I^{*2} + 1\right)(d + \theta + \vartheta) + \sigma I^*\right)}{\beta\left(\xi I^{*2} + 1\right) - (\gamma I^* + 1)\left(\alpha\left(\xi I^{*2} + 1\right)(d + \theta + \vartheta) + \alpha\sigma I^*\right)}. \tag{11}$$

$S^* > 0$ if $\beta\left(\xi I^{*2} + 1\right) > (\gamma I^* + 1)\left(\alpha\left(\xi I^{*2} + 1\right)(d + \theta + \vartheta) + \alpha\sigma I^*\right)$.

On setting the right-hand side term of the first equation of the system (3) to zero and putting the value of S^* using Eq. (11), we obtain that I^* is a solution of the following seventh-degree equation:

$$P(I^*) := A_0 + A_1 I^* + A_2 I^{*2} + A_3 I^{*3} + A_4 I^{*4} + A_5 I^{*5} + A_6 I^{*6} + A_7 I^{*7} = 0, \tag{12}$$

where, the coefficients $A_i's$, $i = 0, 1, 2, \ldots, 7$ are given below:

$$
\begin{aligned}
A_0 &= r(\alpha k + 1)(d + \theta + \vartheta)(1 - R_0),\\
A_1 &= \alpha^2 d^2 k + 2d(\alpha k(\alpha(\theta + \vartheta) - \beta + \gamma r) + \gamma r) + k(\beta^2 - \beta(2\alpha(\theta + \vartheta) + \gamma r)\\
&\quad + \alpha(\theta + \vartheta)(\alpha(\theta + \vartheta) + 2\gamma r) + \alpha r\sigma) + r(2\gamma(\theta + \vartheta) + \sigma),\\
A_3 &= \alpha^2 d^2 k\left(\gamma^2 + 2\xi\right) + 2d\left(\alpha k\left(\alpha\left(\gamma^2 + 2\xi\right)(\theta + \vartheta) + 2\alpha\gamma\sigma - 2\beta\xi + 2\gamma\xi r\right) + 2\gamma\xi r\right)\\
&\quad + k(\alpha^2\left(\left(\gamma^2 + 2\xi\right)(\theta + \vartheta)^2 + 4\gamma\sigma(\theta + \vartheta) + \sigma^2\right) + \alpha\sigma\left(r\left(\gamma^2 + \xi\right) - 2\beta\gamma\right)\\
&\quad - 4\alpha\xi(\theta + \vartheta)(\beta - \gamma r) + 2\beta\xi(\beta - \gamma r)) + r\sigma\left(\gamma^2 + \xi\right) + 4\gamma\xi r(\theta + \vartheta),\\
A_4 &= 4\alpha^2\gamma d^2 k\xi + 2\alpha^2 dk\sigma\left(\gamma^2 + \xi\right) + d\xi(2\gamma(\alpha k(4\alpha(\theta + \vartheta) - 2\beta + \gamma r) + \gamma r) + \xi r(\alpha k + 1))\\
&\quad + 2\alpha^2\gamma k\sigma^2 + 2\sigma\left(\alpha k\left(\alpha\left(\gamma^2 + \xi\right)(\theta + \vartheta) - \beta\xi + \gamma\xi r\right) + \gamma\xi r\right) + \xi(2\gamma(\theta + \vartheta)\\
&\quad \times (\alpha k(2\alpha(\theta + \vartheta) - 2\beta + \gamma r) + \gamma r) + \xi r(\theta + k(\alpha(\theta + \vartheta) - \beta) + \vartheta)),\\
A_5 &= \alpha^2 d^2 k\xi\left(2\gamma^2 + \xi\right) + 2d\xi\left(\alpha k\left(\alpha\left(2\gamma^2 + \xi\right)(\theta + \vartheta) + 2\alpha\gamma\sigma - \beta\xi + \gamma\xi r\right) + \gamma\xi r\right)\\
&\quad + k(\alpha^2\left(\xi\left(2\gamma^2 + \xi\right)(\theta + \vartheta)^2 + \gamma^2\sigma^2 + 4\gamma\xi\sigma(\theta + \vartheta)\right) + \alpha\xi(\gamma\sigma(\gamma r - 2\beta) - 2\xi(\theta + \vartheta)(\beta - \gamma r))\\
&\quad + \beta\xi^2(\beta - \gamma r)) + \gamma\xi r(\gamma\sigma + 2\xi(\theta + \vartheta)),\\
A_6 &= \gamma\xi(d + \theta + \vartheta)\left(2\alpha^2 dk\xi + \alpha k(2\alpha(\gamma\sigma + \xi(\theta + \vartheta)) - 2\beta\xi + \gamma\xi r) + \gamma\xi r\right),\\
A_7 &= \alpha^2\gamma^2 k\xi^2(d + \theta + \vartheta)^2.
\end{aligned}
$$

On analyzing Eq. (12), we propose the following theorem:

Theorem 2 *When $R_0 > 1$, then there exists either a unique, three, five, or seven solutions of the polynomial $P(I^*)$.*

Proof Let $R_0 > 1$. Note that $P(0) = A_0 < 0$ when $R_0 > 1$ and the coefficient $A_7 = \alpha^2\gamma^2 k\xi^2(d + \theta + \vartheta)^2$ is always positive. Thus, we have

$$\lim_{I^* \to \infty} P(I^*) = +\infty.$$

The polynomial $P(I^*)$ is a seventh-degree polynomial in I^*, and it is a continuous function of I^*. Thus, by the fundamental theorem of algebra, $P(I^*)$ can have at most seven roots. □

In the present study, we analyze the system for the existence of a unique endemic equilibrium only. Using Descartes' rules of signs [40], the polynomial $P(I^*)$ can have a unique root when $R_0 > 1$, if any of the following conditions is satisfied:

$$
\begin{aligned}
&A_i > 0 \ (i = 1 \text{ to } 6),\\
&A_1 < 0, \quad A_i > 0 \ (i = 2 \text{ to } 6),\\
&A_1 < 0, \ A_2 < 0, \ A_i > 0 \ (i = 3 \text{ to } 6),\\
&A_i < 0 \ (i = 1 \text{ to } 3), \ A_j > 0 \ (j = 4 \text{ to } 6),\\
&A_i < 0 \ (i = 1 \text{ to } 4), \ A_j > 0 \ (j = 5 \text{ to } 6),\\
&A_i < 0 \ (i = 1 \text{ to } 5), \ A_6 > 0.
\end{aligned}
\tag{13}
$$

After obtaining the value of I^*, the value of S^* can be obtained from Eq. (11). Thus, there exists a unique endemic equilibrium $E_e(S^*, I^*)$ if one of the conditions (13) holds.

To determine the stability behavior of the endemic equilibrium $E_e(S^*, I^*)$, the characteristic equation of the system (3) at E_e is obtained as:

$$\lambda^2 + m_1\lambda + m_2 + (n_1\lambda + n_2)\, e^{-\lambda\rho_1} + (l_1\lambda + l_2)\, e^{-\lambda\rho_2} + o_1 e^{-\lambda(\rho_1+\rho_2)} = 0, \quad (14)$$

where,

$$\begin{aligned}
m_1 &= d + \theta + r\left(\frac{2S^*}{K} - 1\right) + \vartheta,\\
m_2 &= -\frac{r(K - 2S^*)(d + \theta + \vartheta)}{K},\\
n_1 &= \frac{\beta(I^*(\gamma I^* + 1) - S^*(\alpha S^* + 1))}{(\alpha S^* + 1)^2(\gamma I^* + 1)^2},\\
n_2 &= \frac{\beta\left(K(I^*(\gamma I^* + 1)(d + \theta + \vartheta) + rS^*(\alpha S + 1)) - 2rS^{*2}(\alpha S^* + 1)\right)}{K(\alpha S^* + 1)^2(\gamma I^* + 1)^2},\\
l_1 &= \frac{2\sigma I^*}{\left(\xi I^{*2} + 1\right)^2},\\
l_2 &= -\frac{2r\sigma I^*(K - 2S^*)}{K\left(\xi I^{*2} + 1\right)^2},\\
o_1 &= \frac{2\beta\sigma I^{*2}}{(\alpha S^* + 1)^2(\gamma I^* + 1)\left(\xi I^{*2} + 1\right)^2}.
\end{aligned} \quad (15)$$

Now, we investigate the stability of E_e for the following possible cases of ϱ_1 and ϱ_2:

Case (i) $\rho_1 = \rho_2 = 0$.

In this case, the characteristic Eq. (14) becomes:

$$\lambda^2 + m_{01}\lambda + m_{11} = 0, \quad (16)$$

where

$$\begin{aligned}
m_{01} &= (m_1 + n_1 + l_1),\\
m_{11} &= (m_2 + n_2 + l_2 + o_1).
\end{aligned} \quad (17)$$

Using Routh–Hurwitz criteria, it follows that the roots of Eq. (16) must have negative real parts if the condition (**H1:**) $m_{01} > 0$ and $m_{11} > 0$. Thus, we state the following theorem:

Theorem 3 *The endemic equilibrium $E_e(S^*, I^*)$ is locally asymptotically stable for $\rho_1 = \rho_2 = 0$, if the condition (**H1**) holds.*

Case (ii) $\rho_1 > 0,\ \rho_2 = 0$.

The characteristic Eq. (14) at E_e when $\rho_1 > 0,\ \rho_2 = 0$ becomes

$$\lambda^2 + l_{01}\lambda + l_{11} + (n_1\lambda + n_2 + o_1)\, e^{-\lambda\rho_1} = 0, \tag{18}$$

where

$$\begin{aligned} l_{01} &= (m_1 + l_1)\,, \\ l_{11} &= (m_2 + l_2)\,. \end{aligned} \tag{19}$$

where $m_1, m_2, n_1, n_2, l_1, l_2$ and o_1 are given in Eq. (15).

From Theorem 3, we know that all the solutions of Eq. (14) have negative real parts when $\varrho_1 = \varrho_2 = 0$ if (H1) holds. If for a particular value of $\varrho_1 > 0$ and fixed $\varrho_2 = 0$, a characteristic root of Eq. (18) must pass through the imaginary axis if it lies on the right half-plane. Hence, on the contrary, suppose that $\lambda = \mathrm{i}\kappa$ $(\kappa > 0)$ is a root of the Eq. (18).

Substituting $\lambda = \mathrm{i}\kappa$ into Eq. (18) and separating the real and imaginary parts, we obtain:

$$\kappa^2 - l_{11} = n_1\kappa \sin\kappa\varrho_1 + (n_2 + o_1)\cos\kappa\varrho_1, \tag{20}$$

$$-l_{01}\kappa = n_1\kappa\cos\kappa\varrho_1 - (n_2 + o_1)\sin\kappa\varrho_1. \tag{21}$$

On squaring and adding Eqs. (20) and (21), we get

$$\kappa^4 + A_{11}\kappa^2 + A_{12} = 0, \tag{22}$$

where

$$\begin{aligned} A_{11} &= -2l_{11} - n_1^2 + l_{01}^2, \\ A_{12} &= l_{11}^2 - (n_2 + o_1)^2\,. \end{aligned}$$

Let $\kappa^2 = \zeta$, then Eq. (22) becomes

$$H(\zeta) := \zeta^2 + A_{11}\zeta + A_{12} = 0. \tag{23}$$

Let $\mathbf{H2_a}$: $A_{11} > 0$, and $A_{12} > 0$. Using Routh Hurwitz criterion, we have that Eq. (23) cannot have a positive root, which is a contraction to the fact that $\zeta = \kappa^2$ is such a root. Thus, we state the following theorem:

Theorem 4 *The endemic equilibrium $E_e(S^*, I^*)$ is locally asymptotically stable for $\rho_1 > 0$ and $\rho_2 = 0$, if $\mathbf{H2_a}$ holds.*

From Theorem 4, it is inferred that there exists a purely imaginary root $\mathrm{i}\kappa$ of the characteristic equation at E_e if and only if Eq. (18) has a positive real root ζ. On analyzing the quadratic polynomial $H(\zeta)$ graphically, we see that Eq. (23) has at least one positive root if any of the following conditions hold:

$\mathbf{H2_b}$. $A_{12} = l_{11}^2 - (n_2 + o_1)^2 < 0$,

H2$_\mathbf{c}$. $A_{11} < 0$, and $A_{11}^2 - 4A_{12} > 0$.

Without loss of generality, we assume that Eq. (23) has two positive roots ζ_1 and ζ_2, and let $\kappa_i = \sqrt{\zeta_i}$, $i = 1, 2$.

From Eqs. (20) and (21), ϱ_{1_j} corresponding to κ_i can be obtained as

$$\varrho_{1_i}^{(j)} = \frac{1}{\kappa_i} \arccos\left(\frac{\kappa_i^2((n_2+o_1) - l_{01}n_1) - l_{11}(n_2+o_1)}{n_1^2\kappa_i^2 + (n_2+o_1)^2}\right) + \frac{2j\pi}{\kappa_i},$$
$$i = 1, 2, \quad j = 0, 1, 2, \ldots. \tag{24}$$

Let $\varrho_1^* = \varrho_{1_{i_0}}^{(j_0)} = \min\left\{\varrho_{1_i}^{(j)} : i = 1, 2, \quad j = 0, 1, 2, \ldots.\right\}$.

Lemma 2 *Suppose that **H1** holds.*

1. *If either* **H2$_\mathbf{b}$** *or* **H2$_\mathbf{c}$** *satisfies, then Eq.* (18) *has all the roots with negative real parts for* $\varrho_1 \in [0, \varrho_1^*)$.
2. *If neither* **H2$_\mathbf{b}$** *nor* **H2$_\mathbf{c}$** *satisfies, then Eq.* (18) *has all the roots with negative real parts for* $\varrho_1 \geq 0$.

let $\Gamma(\kappa) = \gamma(\varrho_1) + \mathrm{i}\kappa(\varrho_1)$ be the root of Eq. (18) such that $\gamma(\varrho_1^*) = 0$ and $\kappa(\varrho_1^*) = \kappa_{i_0} := \kappa_0$.

To estabilish the Hopf bifurcation, we show the transversality condition $\left[\frac{\mathrm{d}(Re\lambda)}{\mathrm{d}\varrho_1}\right]\Big|_{\lambda=\mathrm{i}\kappa_0} \neq 0$.

Lemma 3 *Suppose that* **H2$_\mathbf{d}$** : $H'(\kappa_0^2) \neq 0$, *then the transversality condition* $\left[\frac{\mathrm{d}(Re\lambda)}{\mathrm{d}\varrho_1}\right]\Big|_{\lambda=\mathrm{i}\kappa_0} \neq 0$ *holds.*

Proof On differentiating Eq. (18) with respect to $\lambda(\varrho_1)$ and then simplifying, we obtain

$$\left[\frac{\mathrm{d}\lambda}{\mathrm{d}\varrho_1}\right]^{-1} = \frac{2\lambda + l_{01}}{-\lambda(\lambda^2 + l_{01}\lambda + l_{11})} + \frac{n_1}{\lambda(n_1\lambda + (n_2+o_1))} - \frac{\varrho_1}{\lambda}.$$

$$\begin{aligned}
\mathrm{Re}\left(\frac{\mathrm{d}\lambda}{\mathrm{d}\varrho_1}\right)^{-1}\Big|_{\lambda=\mathrm{i}\kappa_0} &= \mathrm{Re}\left(\frac{2\mathrm{i}\kappa_0 + l_{01}}{-\mathrm{i}\kappa_0(-\kappa_0^2 + \mathrm{i}l_{01}\kappa_0 + l_{11})} + \frac{n_1}{-n_1\kappa_0{}^2 + \mathrm{i}(n_2+o_1)\kappa_0} + \frac{\mathrm{i}\varrho_1}{\kappa_0}\right) \\
&= \frac{1}{\kappa_0}\left(\frac{2\kappa_0(\kappa_0^2 - l_{11}) + l_{01}^2\kappa_0}{(l_{01}\kappa_0)^2 + (\kappa_0^2 - l_{11})^2} - \frac{n_1^2\kappa_0}{(n_1\kappa_0)^2 + (n_2+o_1)^2}\right) \\
&= \frac{2(\kappa_0^2 - l_{11}) + l_{01}^2}{(l_{01}\kappa_0)^2 + (\kappa_0^2 - l_{11})^2} - \frac{n_1^2}{(n_1\kappa_0)^2 + (n_2+o_1)^2}.
\end{aligned}$$

From Eqs. (20), and (21), we get

$$(n_1\kappa_0)^2 + (n_2+o_1)^2 = (l_{01}\kappa_0)^2 + (\kappa_0^2 - l_{11})^2. \tag{25}$$

Thus, we obtain

$$\begin{aligned}\operatorname{sign}\left\{\frac{\mathrm{d}}{\mathrm{d}\varrho_1}(\operatorname{Re}\lambda)\Big|_{\lambda=\mathrm{i}\kappa_0}\right\} &= \operatorname{sign}\left\{\operatorname{Re}\left(\frac{\mathrm{d}\lambda}{\mathrm{d}\varrho_1}\right)^{-1}\Big|_{\lambda=\mathrm{i}\kappa_0}\right\}\\ &= \operatorname{sign}\left\{\frac{2\kappa_0^2+(l_{01}^2-2l_{11}-n_1^2)}{n_1^2\kappa_0^2+(n_2+o_1)^2}\right\}\\ &= \operatorname{sign}\left\{\frac{H'(\kappa_0^2)}{n_1^2\kappa_0^2+(n_2+o_1)^2}\right\}\\ &\neq 0.\end{aligned} \tag{26}$$

Thus, it follows that $\left[\frac{\mathrm{d}(Re\lambda)}{\mathrm{d}\varrho_1}\right]\Big|_{\lambda=\mathrm{i}\kappa_0} \neq 0$ and hence the proof. □

From the above analysis, we have the following theorem:

Theorem 5 *Suppose that* **H1** *holds.*

1. *If neither* $\mathbf{H2}_b$ *nor* $\mathbf{H2}_c$ *holds, then* E_e *is locally asymptotically stable for all* $\varrho_1 \geq 0$.
2. *If either* $\mathbf{H2}_b$ *or* $\mathbf{H2}_c$ *holds, then* E_e *is locally asymptotically stable for* $\varrho_1 \in [0, \varrho_1^*)$.
3. *If either* $\mathbf{H2}_b$ *or* $\mathbf{H2}_c$ *holds, and* $H'(\kappa_0^2) \neq 0$, *then the system* (3) *undergoes a Hopf bifurcation at* E_e *when* $\varrho_1 = \varrho_1^*$, *and a family of periodic solutions bifurcate from* E_e *when* ϱ_1 *crosses* ϱ_1^*.

Case (iii) $\rho_1 = 0,\ \rho_2 > 0$.
The characteristic equation (14) at E_e when $\rho_1 = 0,\ \rho_2 > 0$ becomes

$$\lambda^2 + m_{01}\lambda + m_{02} + (m_{11}\lambda + m_{12})\, e^{-\lambda\rho_2} = 0, \tag{27}$$

where,

$$\begin{aligned} m_{01} &= m_1 + n_1,\\ m_{02} &= m_2 + n_2,\\ m_{11} &= l_1,\\ m_{12} &= l_2 + o_1. \end{aligned} \tag{28}$$

Let $\lambda = \mathrm{i}\kappa$ be root of Eq. (27). On putting $\lambda = \mathrm{i}\kappa$ in Eq. (27) and simplifying, we obtain

$$\kappa^2 - m_{02} = m_{12}\cos\kappa\varrho_2 + m_{11}\kappa\sin\kappa\varrho_2, \tag{29}$$

$$m_{01}\kappa = m_{12}\sin\kappa\varrho_2 - m_{11}\kappa\cos\kappa\varrho_2. \tag{30}$$

From Eqs. (29) and (30), we obtain

$$\kappa^4 + B_{11}\kappa^2 + B_{12} = 0, \tag{31}$$

where,

$$\begin{aligned} B_{11} &= -2m_{02} + m_{01}^2 - m_{11}^2, \\ B_{12} &= m_{02}^2 - m_{12}^2. \end{aligned} \tag{32}$$

Let $\kappa^2 = \eta$. Then, Eq. (31) becomes

$$G(\eta) := \eta^2 + B_{11}\eta + B_{12} = 0. \tag{33}$$

In the same way, as explained in Case (ii), Eq. (33) has atleast one postive root when either of the following conditions holds:

H3$_\mathbf{a}$. $B_{12} < 0$,
H3$_\mathbf{b}$. $B_{11} < 0$, and $B_{11}^2 - 4B_{12} > 0$.

Without loss of generality, we assume that Eq. (32) has two positive roots η_1 and η_2, and let $\kappa_i = \sqrt{\eta_i}$, $i = 1, 2$.

From Eqs. (29) and (30), ϱ_{2_j} corresponding to κ_i can be obtained as

$$\begin{aligned} \varrho_{2_i}^{(j)} &= \frac{1}{\kappa_i} \arccos\left(\frac{\kappa_i^2(m_{12} - m_{01}m_{11}) - m_{02}m_{12}}{m_{11}^2\kappa_i^2 + m_{12}^2}\right) + \frac{2j\pi}{\kappa_i}, \\ & i = 1, 2, \quad j = 0, 1, 2, \ldots. \end{aligned} \tag{34}$$

Let $\varrho_2^* = \varrho_{2_{i_0}}^{(j_0)} = \min\left\{\varrho_{2_i}^{(j)} : i = 1, 2, \quad j = 0, 1, 2, \ldots.\right\}$.

As shown in Case(ii), **H3$_\mathbf{c}$** : $G'(\kappa_0^2) \neq 0$, then the transversality condition $\left[\frac{\mathrm{d}\,(Re\lambda)}{\mathrm{d}\varrho_2}\right]\Big|_{\lambda=\mathrm{i}\kappa_0} \neq 0$ holds, where $\kappa_0 = \kappa(\varrho_2^*)$. Thus, the following theorem is stated:

Theorem 6 *Suppose that* **H1** *holds.*

1. *If neither* **H3$_\mathbf{a}$** *nor* **H3$_\mathbf{b}$** *satisfies, then E_e is locally asymptotically stable for all $\varrho_2 \geq 0$.*
2. *If either* **H3$_\mathbf{a}$** *or* **H3$_\mathbf{b}$** *satisfies, then E_e is locally asymptotically stable for $\varrho_2 \in [0, \varrho_2^*)$.*
3. *If either* **H3$_\mathbf{a}$** *or* **H3$_\mathbf{b}$** *satisfies, and $G'(\kappa_0^2) \neq 0$, then the system (3) undergoes a Hopf bifurcation at E_e when $\varrho_2 = \varrho_2^*$ and a family of periodic solutions bifurcate from E_e when ϱ_2 crosses ϱ_2^*.*

Case (iv) $\rho_1 = \rho_2 = \varrho$.

In this case, the characteristic equation (14) becomes

$$\lambda^2 + m_1\lambda + m_2 + (q_1\lambda + q_2)\,\mathrm{e}^{-\lambda\varrho} + o_1\mathrm{e}^{-2\lambda\varrho} = 0, \tag{35}$$

where,

$$n_1 + l_1 = q_1, \quad \text{and} \quad n_2 + l_2 = q_2.$$

On multiplying Eq. (35) by $e^{\lambda\varrho}$, we get

$$e^{\lambda\varrho}\left(\lambda^2 + m_1\lambda + m_2\right) + (q_1\lambda + q_2) + o_1 e^{-\lambda\varrho} = 0. \tag{36}$$

Let $\lambda = i\kappa$ be root of Eq. (36). Then, Eq. (36) becomes:

$$\begin{aligned}&\left(-\kappa^2 + m_2 + o_1\right)\cos\kappa\varrho - m_1\kappa\sin\kappa\varrho + q_2\\ &+i\left(\left(-\kappa^2 + m_2 - o_1\right)\sin\kappa\varrho + m_1\kappa\cos\kappa\varrho + q_1\kappa\right) = 0.\end{aligned} \tag{37}$$

On separating real and imaginary parts of Eq. (37), we obtain

$$J_1(\kappa)\cos\kappa\varrho - J_2(\kappa)\sin\kappa\varrho = J_3(\kappa), \tag{38}$$

$$J_4(\kappa)\sin\kappa\varrho + J_5(\kappa)\cos\kappa\varrho = J_6(\kappa), \tag{39}$$

where,

$$\begin{aligned}J_1(\kappa) &= -\kappa^2 + m_2 + o_1, \quad J_2(\kappa) = m_1\kappa, \quad J_3(\kappa) = -q_2,\\ J_4(\kappa) &= -\kappa^2 + m_2 - o_1, \quad J_5(\kappa) = m_1\kappa, \quad J_6(\kappa) = -q_1\kappa.\end{aligned}$$

From Eqs. (38), and (39), we obtain

$$\cos\kappa\varrho = \frac{J_{01(\kappa)}}{J_{00}(\kappa)}, \tag{40}$$

$$\sin\kappa\varrho = \frac{J_{02(\kappa)}}{J_{00}(\kappa)}, \tag{41}$$

where,

$$\begin{aligned}J_{01}(\kappa) &= J_3(\kappa)J_4(\kappa) + J_2(\kappa)J_6(\kappa),\\ J_{02}(\kappa) &= J_1(\kappa)J_6(\kappa) - J_3(\kappa)J_5(\kappa),\\ J_{00}(\kappa) &= J_1(\kappa)J_4(\kappa) + J_2(\kappa)J_5(\kappa).\end{aligned}$$

Equations (40) and (41) implies that

$$J_{01}^2 + J_{02}^2 = J_{00}^2. \tag{42}$$

Now, assume that, (**H4$_\mathbf{a}$**): Eq. (42) has one positive root κ_0. Then, Eq. (36) has a pair of purely imaginary roots $\pm i\kappa_0$. For κ_0, the corresponding critical value of the time delay ϱ is obtained as

$$\varrho_j = \frac{1}{\kappa_0}\arccos\frac{J_{01}(\kappa_0)}{J_{00}(\kappa_0)} + \frac{2\pi j}{\kappa_0}, \quad j = 0, 1, 2, \ldots. \tag{43}$$

Let $\varrho_0 = \min\left\{\varrho_j, \quad j = 0, 1, 2, \ldots\right\}$.

To estabilish Hopf bifurcation, we show that

$$\left[\frac{\mathrm{d}\,(Re\lambda)}{\mathrm{d}\varrho}\right]\Big|_{\lambda=\mathrm{i}\kappa_0} \neq 0.$$

On differentiating Eq. (36) with respect to $\lambda(\varrho)$, we obtain

$$\begin{aligned}&\frac{\mathrm{d}\lambda}{\mathrm{d}\varrho}\left((2\lambda+m_1)\mathrm{e}^{\lambda\varrho}+\varrho(\lambda^2+m_1\lambda+m_2)\mathrm{e}^{\lambda\varrho}+q_1-o_1\varrho\mathrm{e}^{-\lambda\varrho}\right)\\&+\lambda(\lambda^2+m_1\lambda+m_2)\mathrm{e}^{\lambda\varrho}-\lambda o_1\mathrm{e}^{-\lambda\varrho}=0.\end{aligned}\tag{44}$$

$$\begin{aligned}\left[\frac{\mathrm{d}\lambda}{\mathrm{d}\varrho}\right]^{-1} &= \frac{(2\lambda+m_1)\mathrm{e}^{\lambda\varrho}+q_1}{\lambda(o_1\mathrm{e}^{-\lambda\varrho}-(\lambda^2+m_1\lambda+m_2)\mathrm{e}^{\lambda\varrho})}-\frac{\varrho}{\lambda}\\ &= \frac{(2\lambda+m_1)\mathrm{e}^{\lambda\varrho}+q_1}{\lambda(2o_1\mathrm{e}^{-\lambda\varrho}+q_1\lambda+q_2)}-\frac{\varrho}{\lambda}\ \text{(using Eq. (36))}.\\ \left[\frac{\mathrm{d}\,(Re\lambda)}{\mathrm{d}\varrho}\right]\Big|_{\lambda=\mathrm{i}\kappa_0}^{-1} &= \frac{L_1L_3+L_2L_4}{L_3^2+L_4^2},\end{aligned}$$

where,

$$\begin{aligned}L_1&=m_1\cos\kappa_0\varrho_0-2\kappa_0\sin\kappa_0\varrho_0+q_1,\\ L_2&=2\kappa_0\cos\kappa_0\varrho_0+m_1\sin\kappa_0\varrho_0,\\ L_3&=2\kappa_0o_1\sin\kappa_0\varrho_0-q_1\kappa_0^2,\\ L_4&=2\kappa_0o_1\cos\kappa_0\varrho_0+\kappa_0q_2.\end{aligned}$$

If $\mathbf{H_{4_b}}: L_1L_3+L_2L_4\neq 0$, then $\left[\frac{\mathrm{d}\,(Re\lambda)}{\mathrm{d}\varrho}\right]\Big|_{\lambda=\mathrm{i}\kappa_0}^{-1}\neq 0$. Thus, the following theorem is stated using [41]:

Theorem 7 *Suppose that* $\mathbf{H_{4_a}}-\mathbf{H_{4_b}}$ *hold. Then, the endemic equilibrium* E_e *is locally asmptotically stable when* $\varrho\in[0,\varrho_0)$*; the system* (3) *undergoes a Hopf bifurcation at* E_e *when* $\varrho=\varrho_0$*, and a family of periodic solutions bifurcate from* E_e *near* $\varrho=\varrho_0$*.*

Case (v) $\rho_1>0,\ \rho_2>0$ and $\varrho_2\in(0,\varrho_2^*)$.

We consider Eq. (14) with ϱ_2 lies in stable interval and ϱ_1 is regarded as the parameter.

Let $\lambda=\mathrm{i}\kappa_*$, $(\kappa_*>0)$ be root of Eq. (14). Then,

$$G_1\sin\kappa_*\varrho_1+G_2\cos\kappa_*\varrho_1=G_3,\tag{45}$$
$$G_1\cos\kappa_*\varrho_1-G_2\sin\kappa_*\varrho_1=G_4,\tag{46}$$

where,

$$
\begin{aligned}
G_1 &= n_1\kappa_* - o_1 \sin \kappa_*\varrho_2,\\
G_2 &= n_2 + o_1 \cos \kappa_*\varrho_2,\\
G_3 &= \kappa_*^2 - m_2 - l_1\kappa_* \sin \kappa_*\varrho_2 - l_2 \cos \kappa_*\varrho_2,\\
G_4 &= -m_1\kappa_* - l_1\kappa_* \cos \kappa_*\varrho_2 + l_2 \sin \kappa_*\varrho_2.
\end{aligned}
$$

From Eqs. (45) and (46), we obtain

$$
\begin{aligned}
&\kappa_*^4 - 2\kappa_*^3 l_1 \sin(\kappa_*\varrho_2) + \kappa_*^2\left(-2(l_2 - l_1m_1)\cos(\kappa_*\varrho_2) + l_1^2 + m_1^2 - 2m_2 - n_1^2\right) + 2\kappa_* \sin(\kappa_*\varrho_2)\\
&\quad\times(-l_2m_1 + l_1m_2 + n_1o_1) + 2\cos(\kappa_*\varrho_2)(l_2m_2 - n_1o_1) + l_2^2 + m_2^2 - n_1^2 - o_1^2 = 0.
\end{aligned}
\tag{47}
$$

H5$_\mathbf{a}$: Let Eq. (47) has at least finite positive root. We denote the positive roots of Eq. (47) as $\kappa_{1*},\ \kappa_{2*},\ \kappa_{3*}, \ldots, \kappa_{k*}$. For every $\kappa_{i*},\ i = 1, 2, \ldots, k$, the corresponding critical value of time delay is

$$
\varrho_{1i*}^{(j)} = \frac{1}{\kappa_{i*}} \arccos\left\{\frac{G_1G_4 + G_2G_3}{G_1^2 + G_2^2}\right\}_{\kappa_*=\kappa_{i*}} + \frac{2\pi j}{\kappa_{i*}},\quad j = 0, 1, 2, \ldots. \tag{48}
$$

Let $\varrho_1^* = \min\left\{\varrho_{1i*}^{(0)}\right\}, \kappa^* = \kappa_{i*}\big|_{\varrho_1=\varrho_1^*},\ i = 1, 2, \ldots, k$. Differentiating Eq. (14) with respect to ϱ_1 yields:

$$
\left[\frac{\mathrm{d}\lambda}{\mathrm{d}\varrho_1}\right]^{-1} = \frac{2\lambda + m_1 + n_1\mathrm{e}^{-\lambda\varrho_1} + l_1\mathrm{e}^{-\lambda\varrho_2} - \varrho_2(l_1\lambda + l_2)\mathrm{e}^{\lambda\varrho_2} - o_1\varrho_2\mathrm{e}^{-\lambda(\varrho_1+\varrho_2)}}{\lambda\left((n_1\lambda + n_2)\mathrm{e}^{-\lambda\varrho_1} + o_1\mathrm{e}^{-\lambda(\varrho_1+\varrho_2)}\right)} - \frac{\varrho_1}{\lambda}.
$$

$$
\left[\frac{\mathrm{d}(Re\lambda)}{\mathrm{d}\varrho_1}\right]\Big|_{\lambda=\mathrm{i}\kappa^*}^{-1} = \frac{D_1D_3 + D_2D_4}{D_3^2 + D_4^2},
$$

where,

$$
\begin{aligned}
D_1 &= -\kappa^* l_1\varrho_2 \sin\left(\kappa^*\varrho_2\right) + l_1\cos\left(\kappa^*\varrho_2\right) - l_2\varrho_2\cos\left(\kappa^*\varrho_2\right) + m_1 + n_1\cos\left(\kappa^*\varrho_1\right) - o_1\varrho_2\cos\left(\kappa^*\left(\varrho_1^* + \varrho_2\right)\right),\\
D_2 &= 2\kappa^* - l_1\sin\left(\kappa^*\varrho_2\right) + l_2\varrho_2\sin\left(\kappa^*\varrho_2\right) + \kappa^* l_1\varrho_2\left(-\cos\left(\kappa^*\varrho_2\right)\right) - n_1\sin\left(\kappa^*\varrho_1^*\right) + o_1\varrho_2\sin\left(\kappa^*\left(\varrho_1^* + \varrho_2\right)\right),\\
D_3 &= -n_1\kappa^{*2}\cos\kappa^*\varrho_1^* + n_2\kappa^*\sin\kappa^*\varrho_1^* + o_1\kappa^*\sin\kappa^*(\varrho_1^* + \varrho_2),\\
D_4 &= n_2\kappa^*\cos\kappa^*\varrho_1^* + n_1\kappa^{*2}\sin\kappa^*\varrho_1^* + o_1\kappa^*\cos\kappa^*(\varrho_1^* + \varrho_2).
\end{aligned}
$$

H5$_\mathbf{b}$: $D_1D_3 + D_2D_4 \neq 0$.

If **H5$_\mathbf{b}$** holds, then $\left[\frac{\mathrm{d}(Re\lambda)}{\mathrm{d}\varrho_1}\right]\Big|_{\lambda=\mathrm{i}\kappa^*}^{-1} \neq 0$. Thus, using the Hopf bifurcation theorem in [41], we state the following theorem:

Theorem 8 *If the conditions* **H5$_\mathbf{a}$**$-$**H5$_\mathbf{b}$** *hold and* $\varrho_2 \in (0, \varrho_2^*)$*, then the endemic equilibrium* E_e *of system* (3) *is locally asymptotically stable for* $\varrho_1 \in [0, \varrho_1^*)$*, undergoes a Hopf bifurcation at* $\varrho_1 = \varrho_1^*$ *and a branch of periodic solutions bifurcate from* E_e *near* $\varrho_1 = \varrho_1^*$.

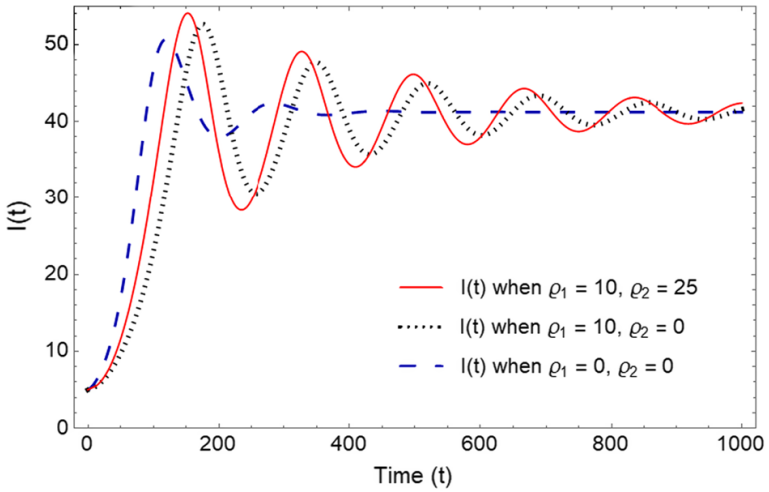


Fig. 2 Delay effects on $I(t)$

4 Numerical simulation

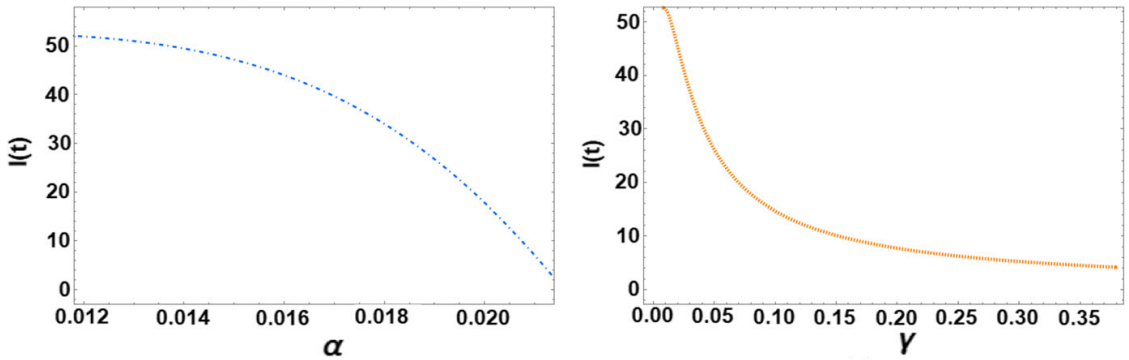
In this section, we simulate the system (3) numerically using the inbuilt command NDSolve of Mathematica 11 for solving the system of Delay differential equations (DDEs). There are many numerical schemes for solving a system of DDEs; for instance, see [42,43]. But, the objective of this section is the numerical validation of qualitative study only; therefore, we use the inbuilt command of Mathematica 11 to solve the system (3).

The following examples illustrates the analytical results:

Example 2 Let $r = 0.08$, $K = 100$, $\beta = 0.0012$, $\alpha = 0.002$, $\gamma = 0.001$, $\vartheta = 0.01$, $d = 0.008$, $\theta = 0.02$.

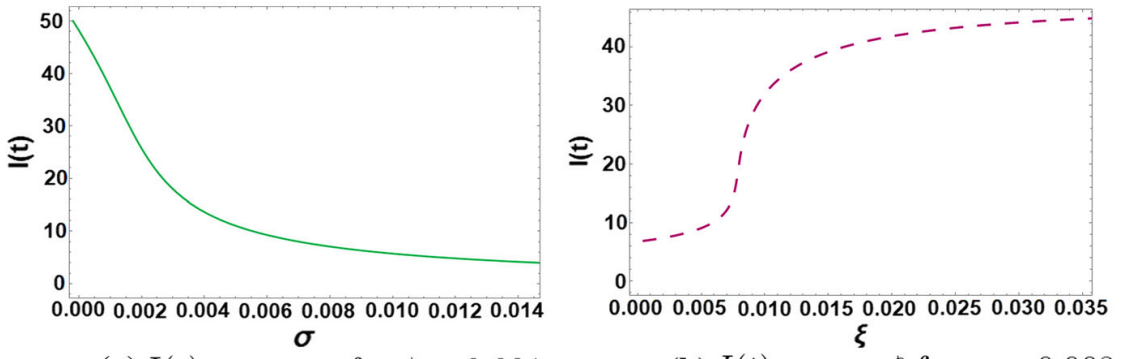
Figure 2 is plotted for $\sigma = 0.004$ and $\xi = 0.009$, which shows the effect of two time delays ϱ_1 and ϱ_2 on the infected population $I(t)$. It depicts that when there is no delay, i.e., $\varrho_1 = \varrho_2 = 0$, the infected individuals will increase and settle down to a steady state after a duration. If there is a time delay in incidence, i.e., $\varrho_1 \neq 0$, for instance, $\varrho_1 = 10$, then the infective will appear with oscillation and persist at a high level and stabilize after a long period. Lastly, if there are time delays in both incidence and treatment rates, i.e., $\varrho_1 > 0$, and $\varrho_2 > 0$, for instance, $\varrho_1 = 10$ and $\varrho_2 = 25$, then the disease will spread at a higher rate with higher oscillation. It shows the significance of considering both the time delays in studying the spread and control of an epidemic.

Figure 3 shows the effects of anti-epidemic preventive measures taken by susceptibles and infectives (recalling that α represents the rate of preventive measures adopted by susceptibles and γ is the rate of preventive measures adopted by infectives). The responsible actions taken by susceptibles and infectives will have different effects and impacts on how rapidly and extensively the infectious diseases will spread. It can be concluded that if both susceptibles and infectives follow effective protection measures, it can significantly increase the ability to control infectious diseases and prevent epidemics.



(a): $I(t)$ versus α. **(b)**: $I(t)$ versus γ.

Fig. 3 Effects of measure of inhibitions α and γ on infectives $I(t)$



(a) $I(t)$ versus σ for $\xi = 0.001$. **(b)** $I(t)$ versus ξ for $\sigma = 0.008$.

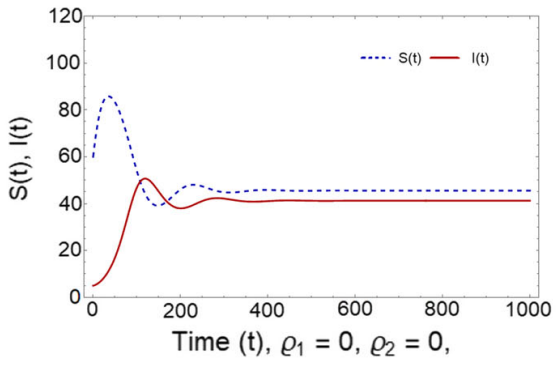
Fig. 4 Effects of cure rate σ and resouces limitation rate ξ on the infected individuals $I(t)$

Figure 4a is drawn to see the effect of the cure rate σ on the infectives, whereas Fig. 4b reflects the impact of limited cure resources on the infected community. It shows that the availability of the treatment can substantially control the disease spread and the inadequate access to treatment creates significant obstacles for eliminating the infection.

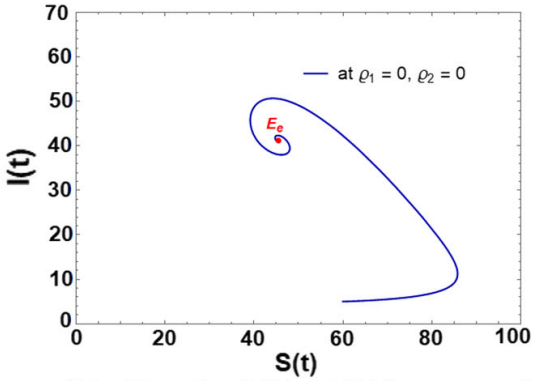
Example 3 Let $r = 0.08$, $K = 100$, $\beta = 0.0012$, $\alpha = 0.002$, $\gamma = 0.001$, $\vartheta = 0.01$, $d = 0.008$, $\theta = 0.02$, $\sigma = 0.004$, $\xi = 0.009$. Using these values of parameters, we obtain that $R_0 = 2.63158 > 1$ and the endemic equilibrium $E_e(S^*, I^*) = (45.554, 41.237)$.

We illustrate the Case (ii) of Sect. 3.3, i.e., $\varrho_1 > 0$ and $\varrho_2 = 0$. We evaluate that $A_{12} = l_{11}^2 - (n_2 + o_1)^2 = -3.51368 \times 10^{-6} < 0$ and $H'(\kappa_0^2) = 0.00405708 > 0$ which confirms the hypothesis $H2_b$ of Theorem 5. We calculate the critical value of ϱ_1 as $\varrho_1^* = 14.7333$. As exhibited in Fig. 5a–d, the endemic equilibrium $E_e(45.554, 41.237)$ is locally asymptotically stable when $\varrho_1 < \varrho_1^* = 14.7333$, and when the value of ϱ_1 crosses the critical value ϱ_1^*, then the periodic solutions bifurcates from E_e, which confirms the results of Theorem 5.

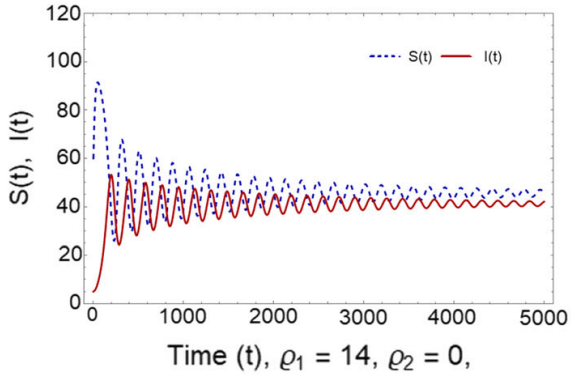
Example 4 To illustrate the Case (iii) of Sect. 3.3, i.e., $\varrho_1 = 0$ and $\varrho_2 > 0$, we consider the following parameters: $r = 0.08$, $K = 100$, $\beta = 0.004$, $\alpha = 0.002$, $\gamma = 0.001$, $\vartheta = 0.01$, $d = 0.008$, $\theta = 0.02$, $\sigma = 0.004$, $\xi = 0.009$. We evaluate that $R_0 = 8.77193 > 1$, $E_e(S^*, I^*) = (14.7911, 17.8592)$, $B_{12} = m_{02}^2 - m_{12}^2 =$



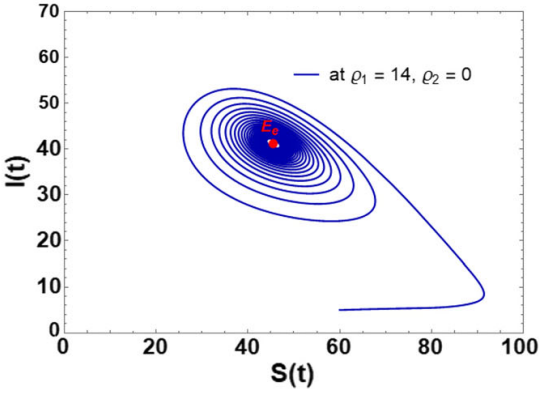
(a): Time series solution of $S(t)$ and $I(t)$ for $\varrho_1 = \varrho_2 = 0$.



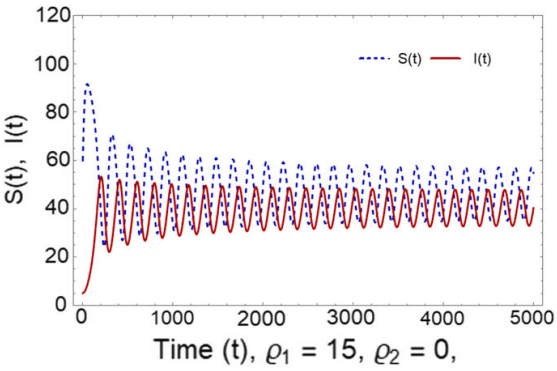
(b): Phase plot of $S(t)$ and $I(t)$ for $\varrho_1 = \varrho_2 = 0$.



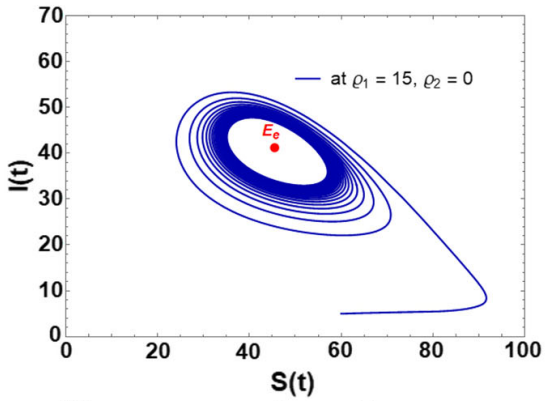
(c): Time series solution of $S(t)$ and $I(t)$ for $\varrho_1 = 14$ and $\varrho_2 = 0$.



(d): Phase plot of $S(t)$ and $I(t)$ for $\varrho_1 = 14$ and $\varrho_2 = 0$.



(e): Time series solution of $S(t)$ and $I(t)$ for $\varrho_1 = 15$ and $\varrho_2 = 0$.

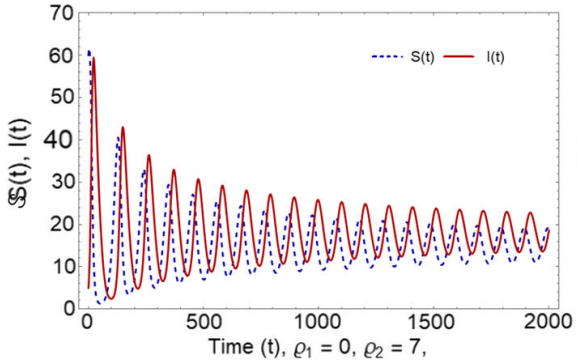


(f): Phase plot of $S(t)$ and $I(t)$ for $\varrho_1 = 15$ and $\varrho_2 = 0$.

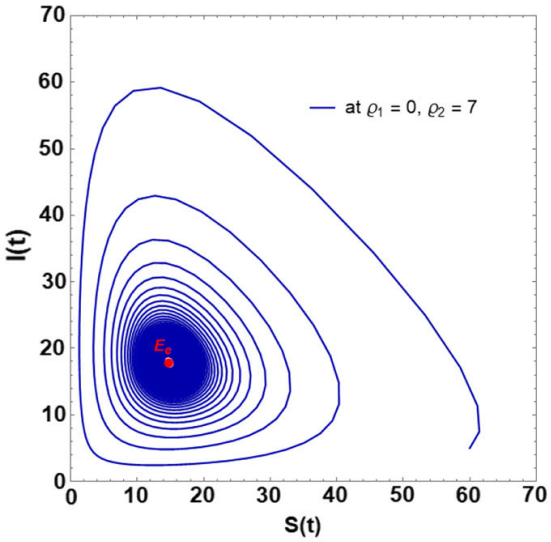
Fig. 5 Plots of Hopf bifurcation, illustrating case (ii) of Sect. 3.3.

$0.0000122401 > 0$, $B_{11}^2 - 4B_{12} = 5.03027 * 10^{-7} > 0$, and $G'(\kappa_0^2) = 0.117404 > 0$. Thus, the hypothesis $H3_b$ of the Theorem 6 holds. The critical value of ϱ_2 is calculated as $\varrho_2^* = 8.17751$.

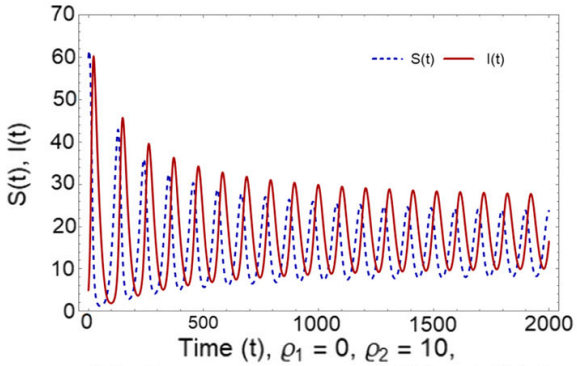
Figure 6a and b show the presence of periodic solutions, where the trajectories of $S(t)$ and $I(t)$ reach to the steady-state $E_e(14.7911, 17.8592)$, when the delay value ϱ_2 is less than its critical value ϱ_2^*. Figure 6c and d are plotted for $\varrho_2 > \varrho_2^*$, which shows that the periodic solutions appear and the trajectories of $S(t)$ and $I(t)$ go away from E_e, which shows the destabilization of the system (3) at E_e.



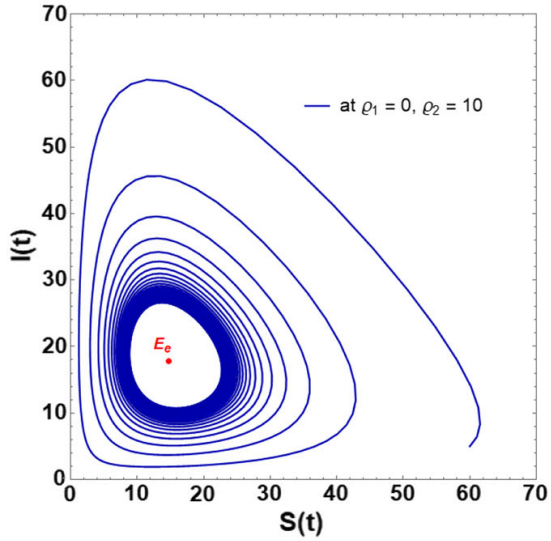
(a): Time series solution of $S(t)$ and $I(t)$ for $\varrho_1 = 0$ and $\varrho_2 = 7$.



(b): Phase plot of $S(t)$ and $I(t)$ for $\varrho_1 = 0$ and $\varrho_2 = 7$.



(c): Time series solution of $S(t)$ and $I(t)$ for $\varrho_1 = 0$ and $\varrho_2 = 10$.

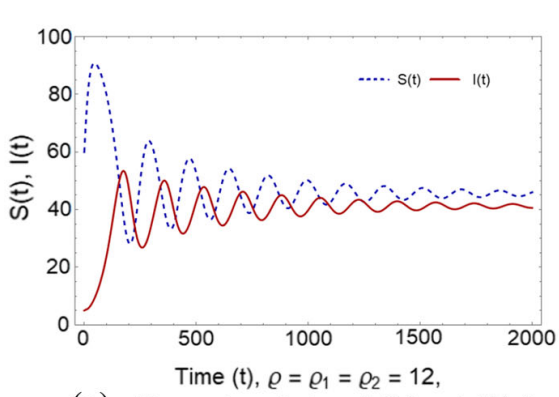


(d): Phase plot of $S(t)$ and $I(t)$ for $\varrho_1 = 0$ and $\varrho_2 = 10$.

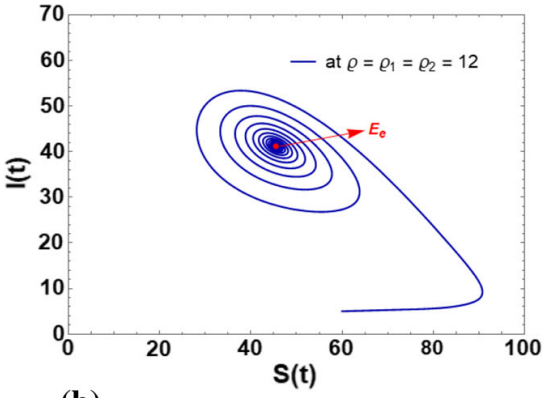
Fig. 6 Presence of Hopf bifurcation, illustrating case (iii) of Sect. 3.3.

Example 5 We consider the case (iv) of Sect. 3.3 (i.e., $\varrho_1 = \varrho_2 = \varrho$) with the same values of parameters as taken in Example (3). The critical value of delay ϱ is obtained as $\varrho_0 = 14.3942$. The time series solutions and the phase plane of $S(t)$ and $I(t)$ are plotted in the Fig. 7a and b for the delay $\varrho = 12$, where the trajectories reaches to the stable endemic equilibrium E_e. On the other hand, Fig. 7c and d show the instability of E_e when the time delay is higher than its critical values.

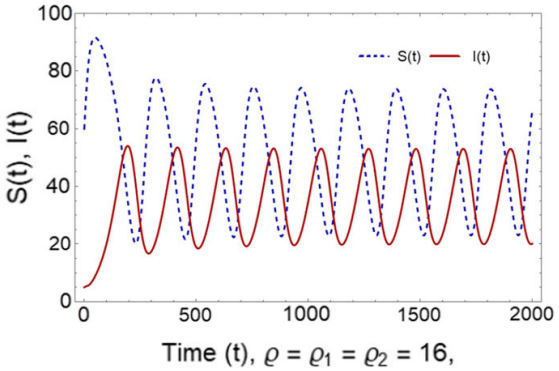
Example 6 In this example, Case (v) of Sect. 3.3 is illustrated with the same values of parameters as taken in Example (3). We fix $\varrho_2 = 7$. The critical value of ϱ_1 is obtained as $\varrho_1^* = 14.2$. Cleary, the impact of both the delay parameters ϱ_1 and ϱ_2 can be seen. Figure 8a and b show the oscillatory trajectories of $S(t)$ and $I(t)$, which stabilized to endemic equilibrium E_e. Figure 8c and d show the stable periodic solutions and



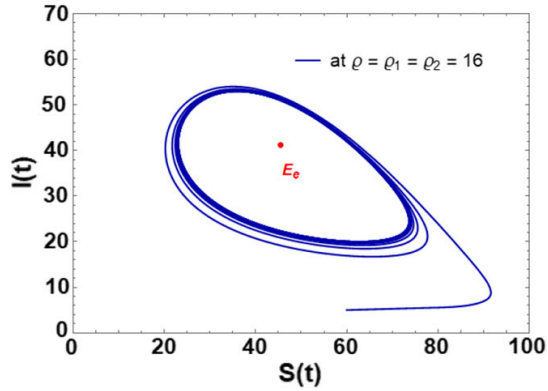
(a): Time series solution of $S(t)$ and $I(t)$ for $\varrho = \varrho_1 = \varrho_2 = 12$.



(b): Phase plot of $S(t)$ and $I(t)$ for $\varrho = \varrho_1 = \varrho_2 = 12$.



(c): Time series solution of $S(t)$ and $I(t)$ for $\varrho = \varrho_1 = \varrho_2 = 16$.



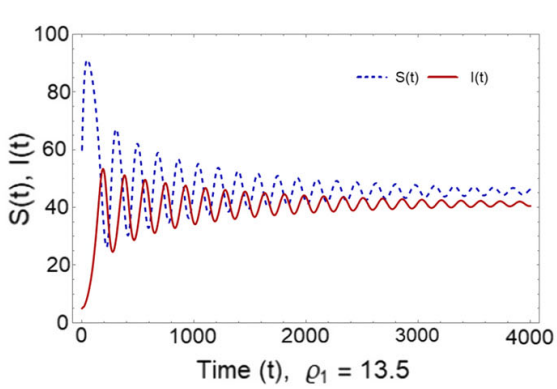
(d): Phase plot of $S(t)$ and $I(t)$ for $\varrho = \varrho_1 = \varrho_1 = 16$.

Fig. 7 Plots of Hopf bifurcation, illustrating case (iv) of Sect. 3.3.

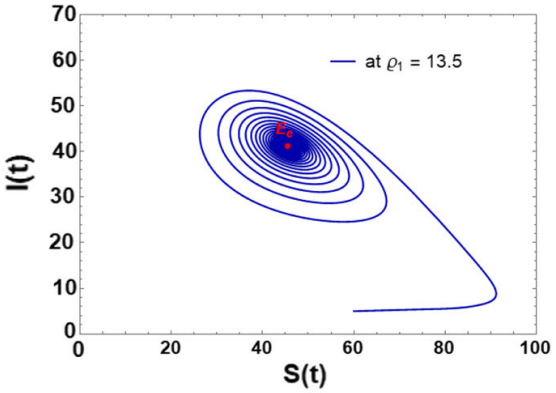
a stable limit cycle, which destabilized the endemic equilibrium E_e. It shows that the infectious diseases continue to proliferate periodically as ϱ_1 crosses ϱ_1^*. Thus, the consideration of both the delays majorly affects the number of infectives and the spread of infection.

5 Discussion and conclusion

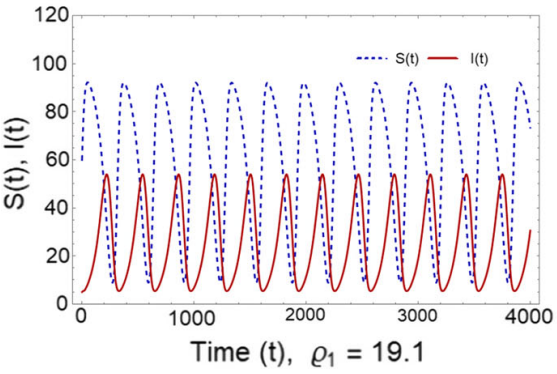
A time-delayed SIR epidemic model is proposed and analyzed mathematically with the logistic growth of susceptible individuals in the present article. The Crowley–Martin type nonlinear incidence rate is incorporated to examine the transmission of infectious diseases from the infected population to the susceptible population. The Holling type III treatment function is considered for the treatment of infectives, which incorporates the limited availability of medical resources such as ICU (intensive care unit) beds, medications, ventilators, hospital conditions, etc. We include two explicit time-delays in the present model: a time delay in the incidence rate, which represents the latent period; a time delay in the nonlinear treatment function, which measures the impact of delay in providing the appropriate therapy to infectives. We obtain the model's feasible equilibria, namely, disease-free equilibrium (DFE) and the endemic equilibrium (EE).



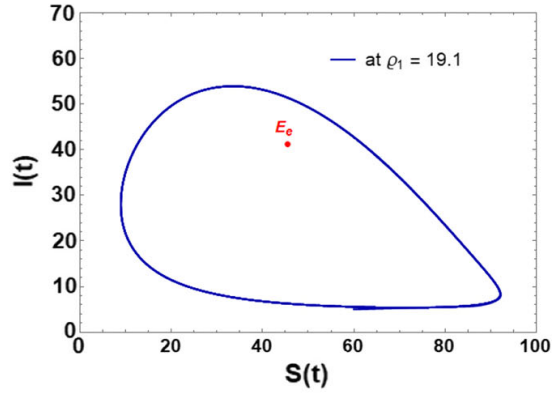
(a) : Time series solution of $S(t)$ and $I(t)$ for $\varrho_1 = 13.5$ and $\varrho_2 = 7$.



(b) : Phase plot of $S(t)$ and $I(t)$ for $\varrho_1 = 13.5$ and $\varrho_2 = 7$.



(c) : Time series solution of $S(t)$ and $I(t)$ for $\varrho_1 = 19.1$ and $\varrho_2 = 7$.



(d) : Phase plot of $S(t)$ and $I(t)$ for $\varrho_1 = 19.1$ and $\varrho_2 = 7$.

Fig. 8 Plots of Hopf bifurcation, illustrating case (v) of Sect. 3.3.

The basic reproduction number R_0 is derived. We show that the DFE of the delayed system is locally asymptotically stable when R_0 is less than unity, unstable when R_0 is greater than unity, and neutrally stable when R_0 is equal to unity. When there is no latent period, then the DFE for $R_0 = 1$ reveals the presence of forward bifurcation. The occurrence of forward bifurcation has an important implication in the epidemic models. It shows that the necessary and sufficient condition to eliminate the disease is to bring the basic reproduction number below unity. Further, the prerequisite for the existence of EE is obtained, and the stability of the EE is investigated. Hopf bifurcation near EE is shown for different possible cases of time delays, and the explicit formulas for the critical values of time-delays are derived.

Further, the numerical simulations verify the analytical results and show the importance of considering nonlinearity and time delays in the present paper. The latent period and the delay in treating patients have a significant impact on the number of infected individuals (Fig. 2). These delays result in spreading infections at a high rate, and so, the infection stays for a longer period. Considering the time delays as the bifurcation parameters can affect the obtained equilibrium's stability (Figs. 5, 6, 7 and 8). When the delay is suitably small, the system reaches its steady state, but stabilizing the system to its equilibrium will be out of control as it crosses the critical value. The course will produce a limit cycle and destabilize the endemic stability, making it dif-

ficult to control the spread of infection. The effect of preventive measures adopted by susceptibles and infectives has been shown numerically (Fig. 3). It shows that human behavior towards the disease exerts a powerful influence on pathogen invasion dynamics and infectious disease spread within human populations. A better understanding of risk factors for developing infectious disease in the general public is a requirement for disease prevention. The effects of cure rate and limitation in medical resources are also shown numerically (Fig. 4). It is observed that increasing medical resource acquisition has effects on eliminating infection from society. A reasonable and timely treatment can speed up the recovery process and prevent disease completely, whereas the limited medical resources availability increases the number of infected people. Thus, the treatment and the availability of its resources are essential to diminish the infection.

The findings of the model, consisting of explicit time delays, Crowley–Martin incidence, and Holling III treatment rates, are capable of demonstrating the significant role of the following nonlinearities: capturing legs between the exposure of disease, onset of its symptoms, and then providing treatment to infectives; susceptibles' and infectives' protection level against the infectious diseases; the limitation in the availability of the medical resources. The results further suggest control strategies to prevent the spread of infections. Our study reveals that the most robust defenses against infectious diseases are the preventive measures of susceptible and infected individuals and the timely treatment of infected individuals. Increasing knowledge, changing attitudes, and reducing risky behaviors towards the disease can reduce the death rate, disability due to illness, and the epidemic's growth. The role of government is crucial for addressing these challenges and achieving health equity. There is a need to strengthen community-wide education programs and other health promotion activities to make individuals aware of the disease and encourage them to make healthy choices. Timely treatment and resource accessibility need to be readily available. The earlier treatment's accessibility in the disease process can mitigate the disease spread and save humans' lives.

Acknowledgements The authors are grateful to the Delhi Technological University, Delhi, India, for providing financial support to carry out this research work. They are also indebted to the handling editor and the anonymous reviewers for their constructive comments and suggestions, which led to an improvement in the original manuscript.

Declarations

Conflict of interest The authors have no conflict of interest.

References

1. Owolabi, K.M.: Mathematical study of two-variable systems with adaptive numerical methods. Numer. Anal. Appl. **9**, 218–230 (2016)
2. Owolabi, K.M.: Numerical solution of diffusive HBV model in a fractional medium. SpringerPlus **5**, 1643 (2016)
3. Owolabi, K.M., Patidar, K.C.: Numerical simulations of multicomponent ecological models with adaptive methods. Theor. Biol. Med. Model. **13**, 1 (2016)

4. Owolabi, K.M., Patidar, K.C., Shikongo, A.: Mathematical analysis and numerical simulation of a tumor-host model with chemotherapy application. Commun. Math. Biol. Neurosci. **2018**, 21 (2018)
5. Owolabi, K.M., Patidar, K.C., Shikongo, A.: Efficient numerical method for a model arising in biological stoichiometry of tumour dynamics. Discrete Contin. Dyn. Syst.S **12**(3), 591–613 (2019)
6. Owolabi, K.M., Atangana, A.: Mathematical analysis and computational experiments for an epidemic system with nonlocal and nonsingular derivative. Chaos Solitons Fract. **126**, 41–49 (2019)
7. Owolabi, K.M., Shikongo, A.: Mathematical modelling of multi-mutation and drug resistance model with fractional derivative. Alex. Eng. J. **59**(4), 2291–2304 (2020)
8. Mishra, A.M., Purohit, S.D., Owolabi, K.M., Sharma, Y.D.: A nonlinear epidemiological model considering asymptotic and quarantine classes for SARS CoV-2 virus. Chaos Solitons Fract. **138**, 109953 (2020)
9. Kermack, W.O., Mckendrick, A.G.: A contribution to the mathematical theory of epidemic. Proc. R. Soc. Lond. **115**(772), 700–721 (1927)
10. Hethcote, H.W., van den Driessche, P.: An SIS epidemic model with variable population size and a delay. J. Math. Biol. **34**, 177–194 (1995)
11. d'Onofrio, A., Manfredi, P., Salinelli, E.: Vaccinating behaviour, information, and the dynamics of SIR vaccine preventable diseases. Theor. Popul. Biol. **71**(3), 301–317 (2007)
12. Zhang, J.Z., Jin, Z., Liu, Q.X., Zhang, Z.Y.: Analysis of a delayed SIR model with nonlinear incidence rate. Discrete Dyn. Nat. Soc. **2008**, 636153 (2009)
13. Wang, J.J., Zhang, J.Z., Jin, Z.: Analysis of an SIR model with bilinear incidence rate. Nonlinear Anal. Real World Appl. **11**(4), 2390–2402 (2010)
14. Xue, Y., Li, T.: Stability and Hopf bifurcation for a delayed SIR epidemic model with logistic growth. Abstr. Appl. Anal. **2013**, 916130 (2013)
15. Kumar, A., Nilam: Dynamic Behavior of an SIR epidemic model along with time delay; Crowley–Martin type incidence rate and Holling type II treatment rate. Int. J. Nonlinear Sci. Numer. Simul. **20**(7–8), 757–771 (2019)
16. Goel, K., Nilam: Stability behavior of a nonlinear mathematical epidemic transmission model with time delay. Nonlinear Dyn. **98**, 1501–1518 (2019)
17. Mena-Lorca, J., Hethcote, H.W.: Dynamic models of infectious disease as regulators of population size. J. Math. Biol. **30**(7), 693–716 (1992)
18. Gumel, A.B., McCluskey, C.C., Watmough, J.: An SVEIR model for assessing potential impact of an imperfect anti-SARS vaccine. Math. Biosci. Eng. **3**(3), 485–512 (2007)
19. Tipsri, S., Chinviriyasit, W.: Stability analysis of SEIR model with saturated incidence and time delay. Int. J. Appl. Phys. Math. **4**(1), 42–45 (2014)
20. Goel, K., Kumar, A., Nilam: A deterministic time-delayed SVIRS epidemic model with incidences and saturated treatment. J. Eng. Math. **121**, 19–38 (2020)
21. Goel, K., Kumar, A., Nilam: Nonlinear dynamics of a time-delayed epidemic model with two explicit aware classes, saturated incidences, and treatment. Nonlinear Dyn. **101**, 1693–1715 (2020)
22. Kumar, A.: Stability of a fractional-order epidemic model with nonlinear incidences and treatment rates. Iran. J. Sci. Technol. Trans. A Sci. **44**, 1505–1517 (2020)
23. Liu, W., Levin, S.A., Iwasa, Y.: Influence of nonlinear incidence rates upon the behavior of SIRS epidemiological models. J. Math. Biol. **23**, 187–204 (1986)
24. Dubey, B., Patra, A., Srivastava, P.K., Dubey, U.S.: Modeling and analysis of an SEIR model with different types of nonlinear treatment rates. J. Biol. Syst. **21**(03), 1350023 (2013)
25. Dubey, P., Dubey, B., Dubey, U.S.: An SIR model with nonlinear incidence rate and Holling type III treatment rate. In: Cushing, J., Saleem, M., Srivastava, H., Khan, M., Merajuddin, M. (eds.) Applied Analysis in Biological and Physical Sciences. Springer Proceedings in Mathematics and Statistics 186, Springer, New Delhi (2016)
26. Crowley, P.H., Martin, E.K.: Functional responses and interference within and between year classes of a dragonfly population. J. N. Am. Benthol. Soc. **8**(3), 211–221 (1989)
27. Song, X., Cheng, S.: A delay-differential equation model of HIV infection of CD4+ T-cells. J. Korean Math. Soc. **42**(5), 1071–1086 (2005)
28. Xu, R., Ma, Z.: Stability of a delayed SIRS epidemic model with a nonlinear incidence rate. Chaos Solitons Fract. **41**(5), 2319–2325 (2009)
29. Shan, C., Zhu, H.: Bifurcations and complex dynamics of an SIR model with the impact of the number of hospital beds. J. Differ. Equ. **257**(5), 1662–1688 (2014)

30. Wang, W., Ruan, S.: Bifurcation in an epidemic model with constant removal rates of the infectives. J. Math. Anal. Appl. **21**, 775–793 (2004)
31. Wang, W.: Backward bifurcation of an epidemic model with treatment. Math. Biosci. **201**, 58–71 (2006)
32. Zhou, L., Fan, M.: Dynamics of an SIR epidemic model with limited medical resources revisited. Nonlinear Anal. Real World Appl. **13**(1), 312–324 (2012)
33. Wu, K., Darcet, D., Wang, Q., Sornette, D.: Generalized logistic growth modeling of the COVID-19 outbreak: comparing the dynamics in the 29 provinces in China and in the rest of the world. Nonlinear Dyn. **101**, 1561–1581 (2020)
34. Li, J., Teng, Z., Wang, G., Zhang, L., Hu, C.: Stability and bifurcation analysis of an SIR epidemic model with logistic growth and saturated treatment. Chaos Solitons Fract. **99**, 63–71 (2017)
35. Avila-Vales, E., Pérez, Á.G.C.: Dynamics of a time-delayed SIR epidemic model with logistic growth and saturated treatment. Chaos Solitons Fract. **127**, 55–69 (2019)
36. Hale, J.K., Lunel, S.M.V.: Introduction to Functional Differential Equations. Springer, New York (1993)
37. van den Driessche, P., Watmough, J.: Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. **180**(1–2), 29–48 (2002)
38. Guckenheimer, J., Holmes, P.: Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields. Springer, Berlin (1983)
39. Castillo-Chavez, C., Song, B.: Dynamical models of tuberculosis and their applications. Math. Biosci. Eng. **1**(2), 361–404 (2004)
40. Wang, X.: A simple proof of Descartes's rule of signs. Am. Math. Mon. **111**(6), 525–526 (2004)
41. Hassard, B.D., Kazarinoff, N.D., Wan, Y.H.: Theory Applications of Hopf Bifurcation. Cambridge University Press, Cambridge (1981)
42. Shampine L.F., Thompson S.: Numerical solution of delay differential equations. In: Delay Differential Equations. Springer, Boston (2009). https://doi.org/10.1007/978-0-387-85595-0_9
43. Rebenda, J., Šmarda, Z.: Numerical algorithm for nonlinear delayed differential systems of nth order. Adv. Differ. Equ. **2019**, 26 (2019)

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

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-021-11045-1

State of the art content based image retrieval techniques using deep learning: a survey

Rajiv Kapoor[1] · Deepak Sharma[2] · Tarun Gulati[2]

Received: 26 January 2020 / Revised: 20 February 2021 / Accepted: 5 May 2021
Published online: 02 July 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

In the recent years the rapid growth of multimedia content makes the image retrieval a challenging research task. Content Based Image Retrieval (CBIR) is a technique which uses features of image to search user required image from large image dataset according to the user's request in the form of query image. Effective feature representation and similarity measures are very crucial to the retrieval performance of CBIR. The key challenge has been attributed to the well known semantic gap issue. The machine learning has been actively investigated as possible solution to bridge the semantic gap. The recent success of deep learning inspires as a hope for bridging the semantic gap in CBIR. In this paper, we investigate deep learning approach used for CBIR tasks under varied settings from our empirical studies; we find some encouraging conclusions and insights for future research.

Keywords Bits-scalable deep binary codes · Spatial maximal activator pooling · Semantic assisted visual hashing · Cosine distance · Deep hashing based on classification and quantization error (DHCQ) · Dot-diffused block truncation coding

1 Introduction

The multimedia researchers have broadly studied about feature representations and similarity measurement approaches which imperatively affect the retrieval pursuance of a Content Based Image Retrieval (CBIR) system. In spite of numerous researches carried out in the field of Content Based Image Retrieval (CBIR) system, the bridging of the semantic gap between low level features captured by machines and high level feature anticipated by human remains a vivid area of researcher's interest.

✉ Rajiv Kapoor
rajivkapoor.dtu@gmail.com

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

[2] Department of ECE, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, India

In last two decades, a number of Content Based Image Retrieval approaches were proposed to handle the challenge of semantic gap. The CBIR methods based on low level features like color, shape and texture are extensively explored and discussed in detail [6, 11, 37, 41, 63]. But CBIR based on color, shape and texture feature can limitedly address efficient retrieval as there is a considerable gap that exists between the analyzing ability using color, texture and shape features compared to the inability at the semantic level. Kokare et al. [28] explored the use of multidimensional indexing techniques for the Content Based Image Retrieval applications. As the database has grown with time, multidimensional indexing was used to speed up the queries. But the multidimensional indexing has proved to be of limited use for most image retrieval applications, due to its loss of efficiency as the number of dimensions increases. To further improve the retrieval efficiency, the focus of research shifted from low level feature extraction to the semantic gap reduction with high level semantics. Liu et al. [34] surveyed the distinct facets of research in this area like the use of object ontology to define high level concepts, the use of machine learning association methods, the introduction of relevance feedback in retrieval loop, generation of semantic templates and making use of text information. But an efficient complete CBIR system is not possible with only high level semantics. It requires the combination of high level semantics with low level features and efficient indexing tools. The research in the field of CBIR has focused on the traditional attributes. Another effective but not much explored method of CBIR, is browsing. Heesch [22] presented a survey to provide an organized overview of distinct browsing models that has been investigated over last one to two decades. Some of the browsing models solve retrieval problem better than others but there are some important issues with the browsing models like creating a browsing structure for the large collection can be computationally costly. The browsing models either use a pre computed structure that often are too rigid or initialize new query at every step that make the system slow for large collection. Another challenge with browsing models is the requirement of long-term learning. Rafiee et al. [50] presented a survey on CBIR using the techniques such as supervised, unsupervised learning and relevance feedback approaches to reduce the semantic gap. The supervised approach does not reasonably achieve desired level of generalization and accuracy, whereas in unsupervised approach, extra information has to be employed to seek desired retrievals. In case of relevance feedback, the efficient retrieval can be achieved by its integration with supervised / unsupervised learning techniques. Dharani et al. [10] also explored the application of various machine learning techniques in CBIR but in the survey they discussed Content Based Image Retrieval based on unlabeled images only. Another survey on CBIR based on different indexing techniques has been carried out by Mukherjee et al. [36]. They explored various bag of words based indexing approaches and concluded that the retrieval efficiency of L1 norm based similarity search is better for small size database. But the other methods like SR tree and Local Search Hashing approaches take less computation time than L1 or L2 norm based approach. One of the key challenges in CBIR is an accurate and fast indexing technique. Patel et al. [44] explored the various Content Based Image Retrieval system using hashing based indexing techniques. They illustrated that albeit hashing based indexing approaches performs better than tree based indexing methods still an efficient indexing method in CBIR requires of future research. There are different attributes that affects the retrieval performance of the CBIR system. Similarity techniques are one of the important attributes. Patel et al. [43] reviewed the various CBIR systems on the basis of distinct similarity measure techniques. The discussed methods have less computation cost but the performance of similarity approach depends on the number of relevant images related to the concept. Wan et al. [64] exploits a study on Content Based Image Retrieval system based

on deep learning. In this survey, CBIR systems using basic deep learning approaches are discussed. The basic deep learning performs better than shallow approaches but it would be very interesting to review CBIR system based on advanced deep learning approaches.

Even though the success of deep learning inspired researchers to explore deep learning based CBIR approaches, most applications that use deep learning are in the field of image recognition and classification and there is still limited focus on application of deep learning in the field of CBIR. This motivates to address questions like, whether deep learning is effective in CBIR work or not and how these methods perform in different domain image databases.

In recent years, the main focus of researchers was on efficient feature extraction methods and similarity metrics. It is very interesting to investigate its application in CBIR system; so this paper limits its focus to the description and analysis of the recent deep learning based CBIR methods. The deep binary codes performed better than other state of art methods in terms of accuracy by using late fusion approach. In case of large surveillance data the high retrieval performance achieved by using prioritized and segmented data with Spatial maximal activators (SMA) pooling. The semantic-assisted visual hashing approach shows superior performance by effectively integrating the visual and text semantics using offline learning. The bilinear CNN model is better in terms of learning more complex features by reducing feature maps to compact size. The VAE-GAN based deep hashing approach generates deep hash codes which contribute to high retrieval performance than other hashing approaches. The R-MAC based method performs better when used with the combination of query expansion (QE) and database side feature augmentation (DBA). Hopefully, this comprehensive study on CBIR systems using deep learning will provide lucrative knowledge and direction to the research in the field of CBIR. The remaining paper is organized as follows.

In Section 2, distinct deep learning based CBIR methods are reviewed and tabulated. The different databases used in CBIR systems based on deep learning are reviewed in Section 3. Section 4 & 5 summarizes the various similarity measures and evaluation parameters respectively used in CBIR. Section 6 contains the concluding remarks and future directions in the field of deep learning based CBIR systems.

2 Deep learning based CBIR approaches

In this Section, the content based image retrieval techniques using deep leaning methods are reviewed. The distinct content based image retrieval (CBIR) approaches based on deep learning are illustrated in Fig. 1.

2.1 Binary multi-scale multi-pooling approach

In this approach, deep binary codes [67] based on deep convolutional features are generated for image retrieval. The response of each feature map is compared to the average response of all features on the deep convolutional layer to create deep binary codes. The bit scalable binary codes are directly produced by the proposed spatial cross-summary strategy. When the image is passed through Convolutional Neural Network (CNN), the binary codes are produced on different deep layers. The deep binary codes generated at different layers contribute differently to search accuracy. The retrieval precision is further enhanced using late fusion approach. The deep learning codes generation is based on generic model and the late fusion approach is query adaptive.

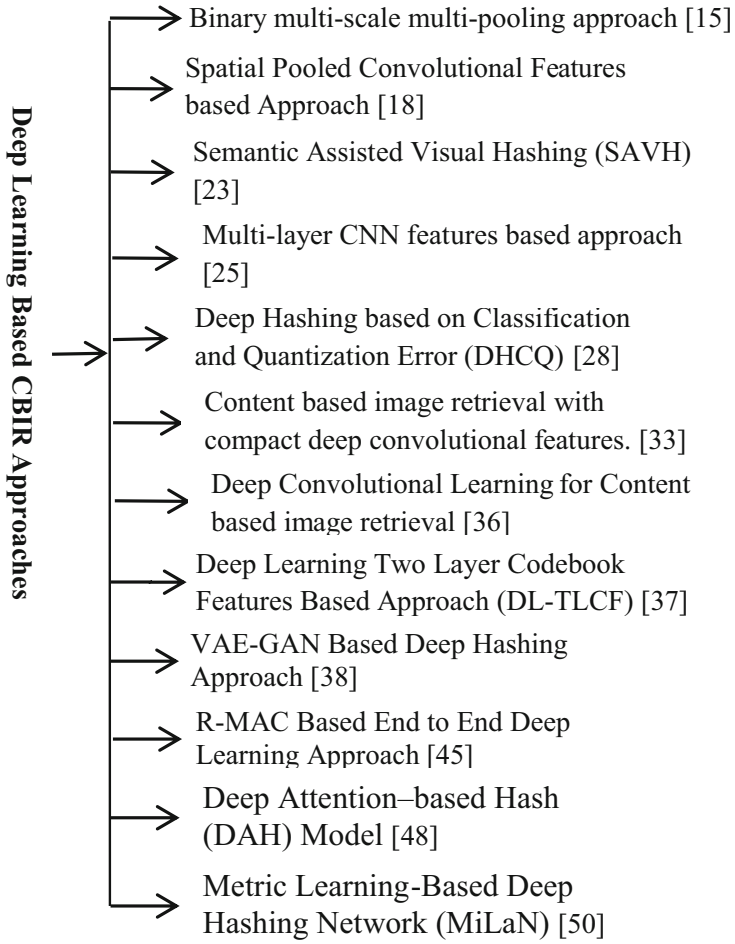


Fig. 1 Overview of Deep Learning based CBIR Approaches

2.1.1 Deep binary codes generation

To generate deep binary codes [67], a pre-trained CNN with 'L' deep convolutional layers is considered. When the input image 'I' is passed through the pre-trained network, the feature maps = $\{F_{i,j}\}$ are obtained for i = 1, 2, 3, −—— C_i. The feature map $F_{i,j}$ is the j^{th} feature map of i^{th} deep convolutional layer and C_i is number of convolutional kernels of deep convolutional layer. The size of feature map $F_{i,j}$ is $W_i \; x \; H_i \; x \; C_i$ where W_i and H_i are the width and height of each channel.

The feature maps generated by max-pooling or sum pooling are only translation invariant. The multi-scale pooling presents the scale-invariant feature maps. The features extracted from multiple scale regions are added and l2-normalized to represent the image. The feature maps are converted into binary codes. Since binary codes have less memory requirements and consume less retrieval time, theses codes are more suitable for large scale image search.

2.1.2 Spatial cross-summing

The different lengths of the feature maps result in a tradeoff between accuracy and efficiency. Low dimensional binary code have less requirement of computational and storage resources, while high dimensional binary codes attain more accuracy. The bit scalable seep binary codes are created by using a spatial cross summing strategy.

2.1.3 Dynamic late fusion

To further enhance the retrieval precision; the top-N search scores from deep binary codes [67] are fused. The hamming distance is used to measure the similarity between the images. The higher similarity means lower value of hamming distance. The modified hamming distance H_I=K-HI, where K in the size of deep learning codes.

2.2 Spatial pooled convolutional features based approach

Selective and low dimensional features are required for an efficient and accurate search of surveillance of data. Multi Trend Structure Descriptor (MTSD) [72] and Multi Scale Local Feature Pattern (MS-LSP) [2] apply pooling principle to local image features. Each sensor node of the modern surveillance network capture, prioritize and transmit visual data to central repositories. The transmission, storage and retrieval of a casual data become a difficult task due to the exponential growth of visual data from surveillance networks [1]. The most common but challenging task is the search of a particular visual object from the data. The proposed work is based on the extraction of optional kernels from the first layer of CNN that is pre trained on ImageNet dataset. The feature maps are generated by convolutional of input image with set of kernels. Then these feature maps are unified into single feature map using spatial maximal activator pooling technique. The particular surveillance objects are represented by computing a low dimensional feature vector. There is an improvement in performance over other state of art approaches for surveillance dataset. The potent features are required for the efficient retrieval of surveillance object. Before feature extraction the prioritization and segmentation of desired objects is necessary.

2.2.1 Image prioritization

All the data apprehended by the visual sensor is not useful. The important data is picked by using some efficient algorithms. Image prioritization scheme is used in this work. This approach is based on the concept of motion saliency.

2.2.2 Object segmentation

The prioritized data is further reduced by removing the unnecessary portion of the image using segmentation. Thus fewer amounts of data are used to extract efficient features transmitted to the server.

2.2.3 Feature extraction

The subnets of selected kernels are used to extract the features. The suitable kernels are selected on the basis of highest frequency usage frequency rate. Then, spatial maximal activator pooling strategy [1] is used to consolidate all the feature maps into a single feature map. The maximum activator value for each pixel position and each color channel produce by consolidated feature map have unique id that includes data, time sensor id and object identification.

2.2.4 Feature matching

Various similarity measures are used for image matching like Euclidean, Manhattan, Hellinger and Canberra distance. The performance of the histogram intersection approach and the Manhattan distance is very close. Therefore the Manhattan distance [1] is preferred due to less computational requirements.

2.3 Semantic assisted visual hashing (SAVH)

Hashing has become one of the active domains of research in the field of content based image retrieval. Single feature visual hashing can be categorized in two categories: data independent and data dependent hashing. Manifold based hashing [55] and deep learning based hashing [13] are in trend to extract binary hash codes. Multi View Latent Hashing (MVLH) [56] formulates binary representation learning of multimodal features. Multi View Alignment Hashing (MVAH) [33] incorporates regularized kernel non-negative matrix factorization in hash codes. The multi feature visual hashing provides comprehensive interpretation of visual contents and optional performance. In this Semantic Assisted Visual Hashing (SAVH) [73] method, the hash codes are integrated with the semantic assistance of auxiliary texts to characterize the correlation between images and shared topics. This unified structure enhances the effectiveness of the visual hashing approach. This method demonstrates superior performance over several state of the art approaches using standard dataset. The structure of the SAVH based Content Based Image Retrieval (CBIR) system has the phrases: offline learning and offline hashing.

2.3.1 Offline learning

The hash codes of database images and hash function of query images are generated in this phase. In the first step, the image pixels using visual and text features of image are identified. Then, with the help of topic hyper graph, a text-enhanced visual graph is constructed while text information guides the latent semantic topic detection. The framework having correlation of images and semantic topics is utilized for hash code of database images.

2.3.2 Online hashing

In online hashing [73], the extracted virtual features of the query image are mapped into binary codes with hash functions. The hashing space is used to calculate to similarity between query and database images. The most similar has minimum distance.

2.4 Multi-layer CNN features based approach

For image retrieval task, the CNN features are firstly used in Overfeat approach. All convolutional layers of GoogleNet [57] are evaluated to extend good results. Convolutional neural network features contain different information at different layers. Lower layer feature have more detailed information but contains semantic ambiguity also. Whereas higher layer features have less but more semantic information. The proposed work effectively attained the complementary strengths of the different layers of CNN [71]. The effectiveness of lower layer similarity is focused on a mapping function, where similarity between query image and its

nearest neighbor is measured in terms of similar semantics. This work outperforms renowned retrievals benchmark datasets. The fully connected layer features are the output of their corresponding layer whereas the aggregated features are the features of convolutional layers. The problem in direct concatenation is the inference of high layer features or low layer features.

The threshold t and exponent P is fixed to a certain value get the best performance from the system. The threshold t is threshold of similarity of the selected layers of the images for similarity matching and exponent P controls the degree of weakening in the mapping function of this approach. The value of P ranges between 0 and 1. The two layers (fc2 and conv 4) are used to evaluate the impact of t and P. The CNN is trained on ImageNet ILSVRC 2012. The framework of CNN has five convolutional layers and three fc [71]. The experiments are conducted using three convolutional layers (3, 4) and two fully connected layer (fc and fc2) to calculate similarity between images. The fully connected layer output of convolutional layer is the extracted set of local features. The Bag of Words (BoW) is used to encode the output of the convolutional layer in a universal representation of the image.

2.5 Deep hashing based on classification and quantization error (DHCQ)

The Supervised Discrete Hashing (SDH) [54] methods are handcrafted visual feature based methods. The deep hashing methods are also very important for image retrieval. Erin binary et al. [31] proposed a DH approach by minimizing the quantization error whereas Yang et al. [68] presented a DH method by minimizing classification error. Due to the fast retrieval speed and low storage cost, hashing approach has been very popular for Approx Nearest Neighbor search. A novel supervised approach is proposed for scalable face image retrieval by concurrently learning feature of images, hash codes and classifier in [59]: the Deep Hashing based on Classification and Quantization error (DHCQ). The deep convolutional network framework extracts discriminative features to generate hash codes and image labels. The learning process of the deep network is guided by the quantization error and the prediction error. Experiments are carried out to demonstrate the effectiveness of the presented work in comparison with the state of the art of existing hashing methods.

2.5.1 Deep binary function

A CNN with powerful feature extraction capabilities is used as a deep hashing learning framework [59]. The deep network has four convolutional layers with max pooling and two fully connected layers: deep feature layer and latent layer. The multi scale features are extracted by connecting third and fourth convolutional layers with deep featured fully connected layer. The features of the fourth layer are more global than third layer features. To learn hashing function, the output of the latent layer is quantized to discrete values. The number of nodes of the latent layer μ is set equal to the length of the desired hashing bits. The sigmoid function is used as activation function of the latent layer. The activation function maps the data into the range of (0, 1). The ReLU function is used as activation function to reset the layers.

2.6 Content based image retrieval with compact deep convolutional features

VLAD and Fisher Vector (FV) are popularly used due to their effectiveness in image retrieval. The VLAD and its variants are applied especially at high layer; in case of CNN image retrieval

applications. Ng et al. [39] encoded the multiple scale and layer features into single VLAD vector. Paulin et al. [45] utilize the patch level descriptor by using a convolutional kernel network. Lin et al. [30] also proposed a bilinear CNN-loss framework for image categorization using two features extractor. Gao et al. [12] recommended a compact pooling approach which is important in terms of computation and memory usage. The various CNN based approaches are utilized for image retrieval task. The image features are extracted at the last layer of the single CNN framework using order less quantization approaches. But this work presents bilinear CNN based structure using two parallel CNNs as feature extractors in context of CBIR. The image features are extracted at different locations and scales using activation of convolutional layers. The deep CNN framework [3] is initially pre-trained on a large generic image dataset and fine tune for CBIR application. The compact and highly discrimination features are produced by applying bilinear root pooling approach to low dimensional pooling layer. The fine tune of network is performed by end to end back propagation to tune parameters for image retrieval work. The work comprises of three steps: 1. Initialization of deep CNN network pre rained on million of images. 2. Fine tuning of bilinear CNN framework on image retrieval dataset. 3. Feature extractor of query and dataset image.

2.6.1 Framework of retrieval and deep learning

The bilinear CNN framework is based on two CNN variant: VGG-m and VGG-16. Both CNN has convolutional layers, pooling layers and fully connected layers [3]. Image size 224 × 224 is initialized as an input. The features are extracted from activation of convolutional feature maps. The fully connected layers are discarded from the fine tune bilinear framework for the sake of simplification. The CNNs are truncated at the last convolutional layer 14th layer of VGG-m and 30th layer from VGG-16. Then three additional layers are added at the end of resulting CNN framework. These three layers are: (1) root compact bilinear pooling to transfer the data in small size. (2) SQRT (3) L2 NORMALIZATION. The resulting network is fine tuned using specific data set. In the pre trained network image features are extracted in the last convolutional C14 and C30 in VGG-m and VGG-16 respectively. The default output is 512 for each of two CNNs. The bilinear model can be represented as $B = (f_A, f_B, P)$. Where f_A and f_B are features set and 'P' is the pooling function. The features f_A, f_B are combined at each location 'l' using bilinear function.

$$Bilinear(l, I, f_A, f_B) = f_A(l, I)^T . f_B(l, I) \tag{1}$$

A low dimensional projection I applied to the extracted features using the root compared bilinear pooling. The low dimensional descriptor then passed through next layer: SQRT and L2 NORMALIZATION respectively. The final fine tuned structure is used to extract image vector of quantization data sets to calculate rank of retrieved images.

This bilinear CNN-Based architecture uses two parallel CNN as feature extractor whereas most of the other related retrieval methods used single CNN architecture. This bilinear model is used to extract image features at various image locations and scales but the other methods extract image features at last layer of CNN. Therefore the bilinear architecture has more discriminative features due to local image pattern extraction from intermediate convolution layers. This bilinear model has also an efficient bilinear root pooling than other related methods, this pooling approach reduces the image features to compact size. The network is then fine-tuned on the domain specific dataset using end to end training. Hence this model works better in terms of accuracy.

2.7 Deep convolutional learning for content based image retrieval

Feature aggregation approach using spatial convolutional layer activation produced compact feature vectors. E. Mohedano et al. [35] proposed an approach having convolutional CNN features and bag of word aggregation scheme. W. Yu et al. [70] exploited the strength of CNN features at different layers which outperforms the concatenation of multiple layers. The features are extracted from the convolutional layer of deep CNN. The most efficient and compact image descriptors are produced in term of retrieval accuracy and memory requirements. This work [62] presents three different images of retraining the network: fully unsupervised retraining when only dataset is available; the retraining with relevance information if label in the dataset are also available and last is relevance feedback based retraining if user feedback is available. The proposed approach is effective as compare to the hand crafted feature based technique and other CNN based approaches. The training of deep networks is not suitable due to huge set of required parameters and the training procedure is also performed offline. This work [62] proposed smaller and faster framework which enhance the performance with less memory requirement.

2.8 Deep learning two layer codebook features based approach (DL-TLCF)

This paper [32] presents an effective image retrieval method by combining high-level features from convolutional neural network (CNN) model and low-level features from dot-diffused block truncation coding (DDBTC). The low-level features, e.g., texture and color, are constructed by vector quantization -indexed histogram from DDBTC bitmap, maximum, and minimum quantizers. Conversely, high-level features from CNN can effectively capture human perception. With the fusion of the DDBTC and CNN features, the extended deep learning two-layer codebook features is generated using the proposed two layer codebook, dimension reduction, and similarity reweighting to improve the overall retrieval rate. Two metrics, average precision rate and average recall rate (ARR), are employed to examine various data sets. As documented in the experimental results, the proposed schemes can achieve superior performance compared with the state-of-the-art methods with either lower high-level feature in terms of the retrieval rate. Thus, it can be a strong candidate for various image retrieval related applications. In the proposed flowchart, to begin with, OpenMP is adopted to extract both GL-FCF and two-layer codebook feature (TLCF) [32] with different threads. As the threads are completed, both features are combined into several fusion features of various dimensions, including low-dimensional deep learning two-layer codebook features (LD(DL-TLCF)), mid-dimensional DL-TLCF (MD(DL-TLCF)) and high dimensional DL-TLCF (HD(DL-TLCF)). In GL-FCF, the released Convolutional Neural Network framework [32] is adopted to train the GoogLeNet model, and extract the features from the last fully-connection layer. In TLCF, DDBTC features, bitmap, color maximum quantizer and minimum quantizer, are extracted from color images to form VQ-indexed histogram by the proposed two-layer codebook, namely color two-layer combination histogram feature (CTLCHF) and bit two-layer combination histogram feature (BTLCHF).

2.9 VAE-GAN based deep hashing approach

Albeit many deep hashing methods are suggested to learn similarity preserving binary codes but usually they suffer from the problems of meager labeled training data or unreliable

semantic constraints. These limitations can be tackled by VGH network [20] as auxiliary network to learn discriminative image representation and compact binary codes concurrently. This method does not need any manual annotation information like image labels or text DSH-GAN and ACMR approaches. The DSH-GAN [48] uses semi supervised generative adversarial network to produce synthetic hashing training data and ACMR [65] learns discriminative and modality in variant representations for retrieval. This architecture fuses variational auto encoder (VAE) [27] with generative adversarial network (GAN) [16] to produce content preserving image for pair-wise hash learning. A semantic preserving feature mapping model is learned by using a real and a synthesized image in pair-wise form under an adversarial generative process. The unsupervised learning is implemented by the network that takes as input pair-wise images $\{I, I', s\}$ and refines them by using a deep hashing pipeline. In the pairwise image description, I and I' are the input and synthesized images respectively and s is the class similarity of pair-wise images. If I and I' belong to same class then $s = 1$ otherwise $s = 0$. The binary hash codes are learned by the hashing network using k-dimensional representation of the hash layer. Once the deep architecture is trained, q-bit hash code can be generated for an input image. Then the final query binary code is obtained by simple quantization using the sign function.

2.10 R-MAC based end to end deep learning approach

Most of the time networks are used as local feature extractors in deep retrieval approaches. They concentrate on designing image representation for image retrieval. Some contributions allow a deep architecture to accurately represent input images of distinct sizes and aspect ratios [26, 60] and address the deficit of geometric invariance of CNN based features. These deep architectures are restricted for retrieval applications due to the lack of supervised learning for the particular instance level retrieval task. In [18], the regional maximum activation of computation (R-MAC) descriptor is used for representation. CNN based descriptors are computed for distinct image region at different scales. These descriptors are further aggregated into a compact feature vector of fixed length and are moderately robust to scale and translation. This approach can encode images at high resolution without distorting their aspect ratio. Originally the R-MAC descriptor uses a pre-trained CNN on ImageNet which may not be an optimal solution for retrieval tasks. The pre-trained CNN is further improved by fine tuning on external set of images. The research work in [17, 49] illustrates that fine-tuned models can bring significant improvement. Siamese network fuses three streams (i.e. query, relevant, irrelevant) with triplet loss to optimize weights for generation of well suited representation of retrieval tasks. Originally the R-MAC is used to find the location of region which is pooled to produce the final image level descriptor. For retrieval task, the network learns to choose regions using Region Proposal Network (RPN). Finally this architecture encodes an image into compact fixed length vector in a single forward pass. Then the dot product is used to compare the representation of different images. Hence this approach encodes information at different resolutions that significantly improves retrieval results.

2.11 Deep attention–based hash (DAH) model

In this work, a Deep Attention–based Hash (DAH) [29] retrieval model was proposed. It is a combination of an attention module and a convolution neural network that is used to obtain strongly representable hash codes. This method significantly improves the image retrieval

precision by effectively extracting the semantic features of the data. The paired batches of images are organized as input that contains classified one-hot tags and similarity information. These strong hash codes achieved the improved level of classification due to interconnection and optimization of semantic similarity representation and classification of recognition ability. The original sigmoid activation function is improved and logarithmic hash loss function is designed. These functions cause a smooth gradient change between the hash layer and the classification layer and also the trained model is allowed to produce similarity features. In the forward structure of DAH network, the core structure of 30 layers of the ResNet50 [21] convolution structure is replaced by the depth-wise separable convolution kernel, from Xception [8]. The back structure finds the direction of change of model. It calculates partial derivatives of the loss function in the opposite direction of the forward structure. The initial experiments in this study prove the positive effect of the design on the model parameter training.

2.12 Metric learning-based deep hashing network (MiLaN)

Recent deep neural network (DNNs) have shown success in optimize feature learning for the retrieval tasks in end to end fashion. But in case of training of DNNs with small labeled dataset can cause problem of over fitting. The metric learning-based deep hashing network (MiLaN) [52] used a pre-trained DNN as a intermediate feature representation step to reduce the problem of over fitting on small remote sensing labeled datasets without the retraining a fine tuning requirements. The big CNN (Inception Net [58]) pre-trained on Imagenet is used to extract in intermediate features of remote sensing images. Then, using these intermediate features a smaller network is trained in order to learn final hash function. Each image is fed to the Inception Net [58], where the last layer (pool 3) is applied just before the softmax layer. This last layer consist of 2049 neurons, whose activation shown the result of average pooling on the complete feature maps of the layer. It optimize the features by learning a semantic based metric space then compact binary hash codes are computed for fast retrieval task. The hamming distance is calculated between the query and each database image.

The overview of various state of the art content based image retrieval techniques using deep learning is summarized in Table 1. In this study, the mean average precision (mAP) is the main performance evaluation parameter. The bilinear root compact pooling [3] has higher mAP for Oxford 5 k database (0.886) and holiday dataset (0.951) than other approaches. The spatial pooling of deep convolutional feature approach [1] is used for the person retrieval applications. From large database, the supervised deep hashing method [59] performed better than other methods used for information retrieval for face images database in youtube (0.7786) and face-scrap (0278) database. The superior values of the mAP for large databases like MIRFlickr (0.6704) and NUS-Wide (0.5281) are obtained by using semantic assisted visual hashing [73] for retrieval applications. In terms of top N-score, the R-MAC based end to end deep learning based approach [18] has best score for UKBench (3.84) database.

3 Dataset

The performance of deep binary codes and the dynamic late fusion scheme are evaluated on four datasets: INRIA Holidays, Oxford, UKBench and MIRFLICKR 1 M [24]. The INRIA Holidays dataset [25] has been created for the ANR RAFFUT project i.e. the joint project of

Table 1 State of the Art Content Based Image Retrieval Techniques using Deep Learning

Author	Algorithm	Deep Network Used	Database	Similarity Measure	Evaluation Parameters	Description
Song Wu et al. [67]	Deep Binary Codes	VGGNet	INRIA Holidays, Oxford 5 k, UKBench, MIRFLICKR	Hamming distance	mean Average Precision (mAP), **0.7719** (INRIA Holidays+ FLICKr) **0.5204** (Oxford 5 k +FLICKr), **0.9156** (UKBench + MIRFLICKR)	Significantly reduces the memory requirements and computational cost compared to other state of the art methods and the dynamic late fusion scheme provides consistent improvement in accuracy.
Jamil Ahmad et al. [1]	Spatial Pooling of Deep Convolutional Features	AlexNet	Weizmann, ETHZ, CUHK, CAVIAR	Euclidean, Manhattan, Hellinger & Canberra distances, Histogram Intersection	mean Average Precision (mAP) **0.8150** (Weizmann) **0.6790** (ETHZ) **0.6810** (CUHK) **0.4760** (CAVIAR)	Effective mechanism to deal with surveillance data but a more improved feature extraction process can be used to extract more sophisticated features.
Lei Zhu et al. [73]	Semantic Assisted Visual Hashing	Proposed SAVH	WiKi, MIR Flickr, NUS-WIDE	Hamming space	mean Average Precision (mAP) **0.1991** (WiKi) **0.6704** (MIRFlickr) **0.5281** (NUS-WIDE)	Superior performance than other visual hashing methods. Its effectiveness can be further validated by involving more associated modalities.
Wei Yu et al. [71]	Multi-layer CNN	AlexNet, VGG-16	Oxford 5 K, Holiday, UKB	Cosine distance	mean Average Precision (mAP) **0.615** (Oxford 5 K) **0.912** (Holiday) top –N Score **3.69** (UKB)	Outperforms the methods using handcrafted features and single layer CNN but can be further improved using complementary information
		Proposed DHCQ		Hamming Ranking		

Table 1 (continued)

Author	Algorithm	Deep Network Used	Database	Similarity Measure	Evaluation Parameters	Description
Jinhui Tang et al. [59]	Supervised Deep Hashing		Youtube, FaceScrub & Cifar-10		mean Average Precision (mAP) **0.7786** (Youtube) **0.2728** (FaceScrub) **0.7094** (Cifar-10)	Retrieval Performance is enhanced using optimized objective function defined over classification and quantization error.
Ahmad Alzu'bi et al. [3]	Bilinear Root Compact Pooling of Deep Convolutional Features	VGG-16	Holidays, Oxford, UKBench	Euclidean distance, Manhattan distance	mean Average Precision (mAP) **0.886** (Oxford 5 K) **0.951** (Holiday) top –N Score **3.56** (UKB)	High retrieval performance in terms of extraction & search time and storage cost.
Maria Tzelepi et al. [62]	Deep convolutional learning using model retraining	CaffeNet	Paris 6 k, UKBench,	Euclidean distance, Cosine distance	mean Average Precision (mAP), **0.9859** (Paris 6 k) top-N score **3.9710** (UKB)	Outperforms others CNN-based retrieval techniques as well as conventional handcrafted feature based approaches.
Peizhong et al. [32]	Fusion of Deep features and compressed features	GoogLeNet	Corel 1000, Brodatz, Holidays, UKBench	Proposed similarity assessment algorithm	Average Precision, **0.7509** (Corel 1000) **0.8205** (Holidays) Average Recall **0.9225** (Brodatz) NS- Score **3.636** (UKBench)	Achieve the best performance compared to the other state of the art methods on both natural and textural databases and is a very robust and effective method for retrieval applications.
Jin Guoqiang et al. [20]	VAE-GAN based deep hashing approach	Variational Auto-encoder (VAE), Generative Adversarial Network (GAN)	CIFAR-10, Flickr, NUS-WIDE	Hamming Ranking	mean Average Precision (mAP), **0.129** (CIFAR-10), **0.529** (Flickr),	Attains state of the art retrieval performance by yielding accurate deep hash codes.

Table 1 (continued)

Author	Algorithm	Deep Network Used	Database	Similarity Measure	Evaluation Parameters	Description
Albert Gordo et al. [18]	R-MAC based end to end deep learning based approach	VGG-16, ResNet101	Oxford 5 k, Paris 6 k, Holidays, UKBench	Dot Product	**0.475** (NUS-WIDE) mean Average Precision (mAP), **0.861** (Oxford 5 k), **0.946** (Paris 6 k), **0.948** (Holidays) Recall@4 Score **3.84** (UKBench)	Faster and more memory efficient. It outperforms other state of the art approaches.
Xinlu Li et al. [29]	Deep Attention–based Hash (DAH) Model	ResNet50	CIFAR-10,	Hamming distance	mean Average Precision (mAP), **0.9267** (CIFAR-10),	The inclusion of the attention model in DAH improves the performance of this method in term of precision.
Subhankar Roy et al. [52]	Metric Learning-Based Deep Hashing Network (Mi-LaN)	Inception Net	UCMD, AID	Hamming distance	mean Average Precision (mAP), **0.991** (UCMD), **0.926** (AID),	The MiLaN method is more fast and accurate than other CBIR used for remote sensing data retrieval applications.

INRIA and Advestigo Company. This has 1491 personal holidays photos divided in 500 categories. This dataset is mainly used for evaluation of image search approaches. It is collection of personal holiday images of natural, man-made and water scenes etc. The Oxford building dataset [46] contains 5062 of 11 different Oxford landmarks that are collected from Flickr. This dataset has 55 queries in all having 5 queries for each landmark category. Henrik Stewenius and David Nister created the UK bench dataset [40] of 10,200 images of 2550 groups having 4 images each of size 640 × 480. The dataset blurred and rotated images for different evaluation purposes. This dataset is typically used of image retrieval applications using one query image from each group. MIRFLICKR 1 M dataset [24] has 1 million Flickr images introduced as an ACM Sponsored image retrieval evaluation. These datasets are popularly used for benchmarking. The different types of datasets are used to evaluate the performance of the spatial pooled convolutional feature based approach. The dataset 3DPeS (3D People Surveillance dataset) [5] mainly created for people re-identification multicar system having many videos capture from real surveillance set up of 8 cameras. This is capture by monitoring an area of the University of Modena and Reggio Emilia campus. The images are uncompressed with resolution of 704 × 576 pixels. VIPER (Visual Person Detection made Reliable) [19], data was acquired at Brugge railway station, Belgium with cooperation of the thermal imaging company FLIR. There are 25,000 images that are extracted as frames rom 28 separated videos. Multi-view car dataset [42] have images of cars and rotated by 360 degree. This dataset is captured by Nikon D70 camera a tripod at the Geneva international motor show'08. These comprehensive car dataset [69] are used for kernel selection. The vehicle image dataset [4] have 4000 vehicle image under real road environment. The Weizmam [14] dataset contains 1464 image of 22 different vehicles from different viewpoints. The ETHZ public surveillance image dataset [53] has 8500 object patches from 149 people included the background clutter and degree of occlusion. The most challenging dataset CAVIAR [7] and CUHK [1] person dataset are also utilized for the evaluation of proposed work. The CAVIAR dataset contains 10 images captured from different view of 72 different individuals in shopping mall using surveillance cameras. The CUHK dataset consists of 3884 images of 971 different people captured with surveillance camera from different views. The performance of the SAVH is validated by conducting experiments on the publically available databases: Wiki [51], MIR flicker [23] and NUS-WIDE [9]. All the datasets include image and text. The Wiki dataset [51] has 2866 Images of 10 different semantic categories whereas NUS-WIDE dataset [9] has 269,648 images categorized in 81 different concepts. It was developed by the lab for media search in National University of Singapore. The three well known dataset Oxford 5 K [46], Holiday [25] and UKB [41] are used to evaluate the performance of the multilayer CNN feature based approach. The DHCQ method [59] is evaluated by conducting experiments on dataset such as: Youtube face dataset [66], face scrub face datasets [38] and cifar-10 [61]. The Youtube dataset contains 1595 face images of people captured from 3425 videos. The face scrub dataset comprised of 107,818 face images of 530 different celebrities with about 200 images per person whereas the Cifar-10 contains 60,000 images of size32 ×32 each that are classified in 10 different categories. Cifar-10 was created by Alex Krizhevsky, Vinod Jain and Geoffrey Hinton. The standard datasets like the Holiday dataset [] and Oxford building dataset and UK-bench dataset are used to evaluate the performance of the proposed work in comparison to other state of the art approaches. The performance of deep learning approach [15] for CBIR has been tested using dataset Paris 6 K [47] and UK-bench [42]. The Paris 6 K dataset contains 6392 images collected from Flickr by search of Paris landmarks. University of California Merced [52] (UCMD) dataset contains 2100 aerial images from 21 different land-

cover categories, where each category includes 100 images. Another benchmark dataset is the aerial image data set [52] (AID) contains 10,000 images from 30 different categories, and the number of images in each category varies between 220 and 420. The accomplishments of DL-TLCF approach [32] has been accessed using various types of datasets like COREL 1000 with 1000 images divided in 10 categories of 100 images each, the Holiday, the UK-bench and the Brodatz-1856 dataset which has 1856 images divided in 116 categories of 16 images each etc. The comparison of various datasets used for the evaluation of CBIR techniques that use deep learning based on the dataset size and of the number of categories is shown in Figs. 2 and 3 respectively.

4 Similarity measures

In content based image retrieval approaches, the similarity measures play a very crucial role in retrieving relevant images from databases. In CBIR, the image contents (features) like color, shape or texture extracted from images are compared with the contents of the query image with the help of similarity measures. The accuracy of the retrieval results highly depends on the performance of the retrieval results similarity measures. The Manhattan distance [1, 3] is one of the simple and robust measures used to find similarities in image retrieval. The Manhattan distance is given by Eq. (2).

$$d_{MH}(a,b) = \sum_{j=1}^{m} a_j - b_j \tag{2}$$

Where a & b are the two set of image feature vectors and j denotes the j^{th} feature vector. The Euclidean distance [1, 3, 62] is another very popular similarity metric used in many applications particularly best suited for the measurement of image similarity in terms of best match. Equation (3) shows the formula for Euclidean distance as:

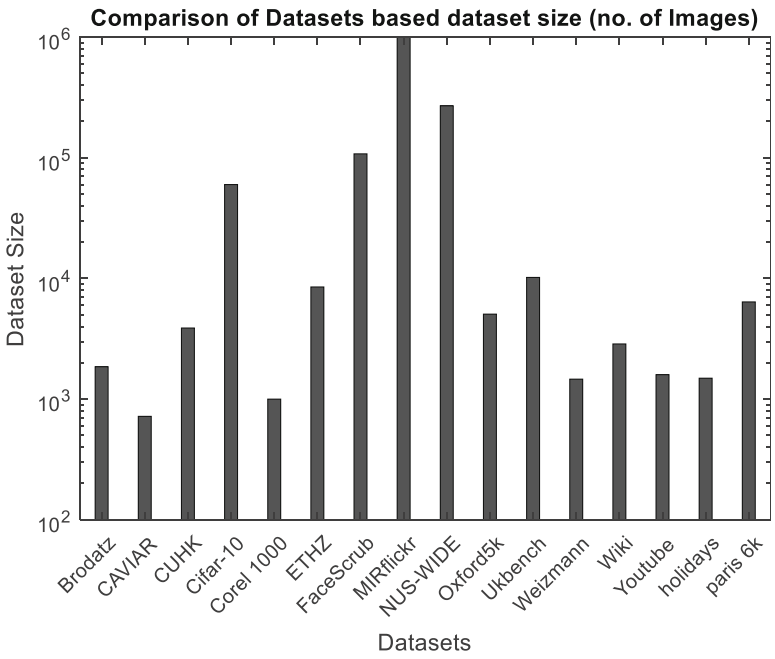


Fig. 2 Comparison of the dataset size of datasets that use Deep Learnig for CBIR

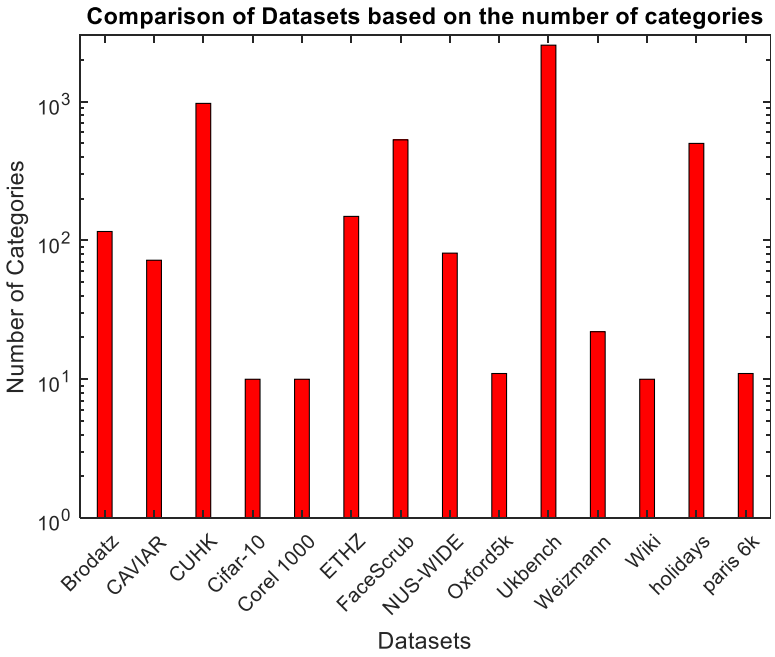


Fig. 3 Comparison based on the number of categories for the datasets that use Deep Learning for CBIR

$$d_E(a,b) = \sqrt{\sum_{j=1}^{m} \left(a_j - b_j\right)^2} \tag{3}$$

The histogram intersection distance [1] is a similarity metric proposed for the color image retrieval applications. This similarity metric is defined for color histograms as given in Eq. (4) where |a| and |b| are the respective histogram magnitudes.

$$d_{HI}(a,b) = \frac{\sum_x \sum_y \sum_z \min(a(x,y,z), b(x,y,z))}{\min(|a|,|b|)} \tag{4}$$

Where a & b are the RGB images, (x, y) are the pixel coordinates and z represents the colour channel of RGB image. Initially, however the Canberra distance [1] was not widely used but due to its significant performance in retrieval application it has become popular in CBIR applications in these days. The formula of the Canberra distance is shown in Eq. (5).

$$d_{CAN}(a,b) = \sum_{j=1}^{m} \frac{|a_j - b_j|}{|a_j| + |b_j|} \tag{5}$$

The cosine distance d_c [62, 71] is often used as similarity metric for image retrieval applications. It is calculated by using Eq. (6):

$$d_C(a,b) = 1 - CS(a,b) \tag{6}$$

CS is the cosine similarity and calculated by Eq. (7).

$$CS(a,b) = \frac{a.b}{\|a\|\|b\|} = \frac{\sum_{j=1}^{m} a_j b_j}{\sqrt{\sum_{j=1}^{m} a_j^2}\sqrt{\sum_{j=1}^{m} b_j^2}} \tag{7}$$

with a_j and b_j are the components of vector a and b respectively. Liu et al. [32] proposed a similarity assessment and a normalization algorithm for finding similarity between images with improved overall performance. The hamming distance is used to evaluate the similarity between two images if features of the images are represented as deep binary codes. It evaluates the number of bits difference between two binary code vectors: the lower the number of the dissimilar bits, the higher the similarity. This metric is used when features are in the form of binary code vectors.

The Hellinger distance [1] is also used as similarity measure in image retrieval applications where the image features are represented in the form of probability distribution function. Given two probability distribution functions A & B, the Hellinger distance between them is shown in Eq. (8).

$$d_H(A,B) = \frac{1}{\sqrt{2}} \left\| \sqrt{A} - \sqrt{B} \right\|_2 \tag{8}$$

5 Evaluation parameters

The performance of a content based image retrieval (CBIR) system can be evaluated by various evaluation parameters (performance metrics) such as precision, recall, average precision rate (APR), average recall rate (ARR), mean average precision (mAP) and top-N score. The precision [62] is defined as the ratio of relevant retrieved images to the total number of retrieved images.

$$P = \frac{A_R}{B_T} \tag{9}$$

where P is the precision of B_T retrieved images, A_R is the number of retrieved images and B_T is the total number of retrieved images for a query and the average precision [32] is the average of the precision scores of all the relevant images of a query image.

$$P_{avg} = \frac{\sum_{l=1}^{L} P(l) \times R(l)}{B_T} \tag{10}$$

where P_{avg} is the average precision of G retrieved images. P(g) is the precision of top g retrieved images and R(g) indicates the relevance of the image of rank g to the query image. The value of R(g) =1 for relevant image at rank g and 0 for irrelevant image. The mean average precision (mAP) [1, 3, 59, 67, 71, 73] is the mean of all the average precisions of different query images applied to the system.

$$mAP = \sum_{j=1}^{J} \frac{P_{avg}(j)}{J} \tag{11}$$

where mAP is the mean average precision score of J different queries and $P_{avg}(j)$ represents the average power of j^{th} query. The recall [32, 62] is the ratio of total number of relevant images retrieved and the total number of relevant images present in the database. The recall is given by the Eq. (12).

$$R = \frac{A_R}{C_T} \tag{12}$$

where A_R is the number of relevant retrieved images and C_T is the total number of relevant images in the database. The parameter, top-N score [62] refers to average number of relevant images, within the top-N ranked images.

6 Discussion

In recent past, in order to exploit potential advantage of different deep learning approaches, content based image retrieval system has been explored. The deep binary codes have many advantages over deep convolutional features as they do not need any additional training. The experiments carried out using standard dataset demonstrate that the deep binary codes are competitive to other state of art approaches. The dynamic late fusion approach presents considerable enhancement in accuracy. The prioritized and segmented surveillance image dataset is very helpful to extract compact and discriminative features. The features are utilized by selected kernel with SMA to achieve high retrieval performance. This technique is very effective when dealing with large collection of surveillance data. The SFVH and MFVH schemes utilize only visual features and do not take advantage of valuable text segments for performance enhancement. The UCMH technique can treat both visual and text equally but fails to achieve good performance. The SAVH approach effectively integrates the semantics of text by offline learning using image as input. The SAVH shows performance superiority over other techniques. The objective function based on classification and quantization error is optimized. The obtained learned hashing codes boost the retrieval performance. The bilinear CNN model shows higher efficiency in learning more complex image features. This model extracts more discriminative features as it uses local image pattern taken from intermediate convolutional layers. This model reduces the features map to compact size which enhances the performance in terms of retrieval time and storage. This shows high retrieval performance for large scale image retrieval. The VAE-GAN based deep hashing approach attains state-of-the-art retrieval performance by generating deep hash codes. The R-MAC based approach with its global descriptors performs well and the performance is improved by combining some complex methods like query expansion (QE) and database-side feature augmentation (DBA). These methods made the system faster and memory efficient but use costly matching and verification. The DAH learns supervised information end to end by developing a loss function for learning the correlation information between images. MiLaN avoids the over fitting problem in small – size labeled remote sensing dataset. This is more accurate and time efficient than the state of the art methods.

The different dataset for deep learning based CBIR system are summarized in Figs. 2 and 3 with different parameters. The datasets are generated under distinct conditions. In general, the research of deep learning based CBIR system is performed using standard dataset so that results can be reused for further study and meaningful comparison of performance of various CBIR methods can be performed. In this review, the various deep learning based methods are tested for different standard dataset or different domain like human face and surveillance datasets.

In this survey, the distinct similarity measures like Euclidean, Hamming, Manhattan, Canberra, Hellinger distance and histogram intersection for CBIR are discussed. Evaluation

parameters like mean average precision (mAP), top-N score, average precision and average recall are mostly considered for performance evaluation of deep learning based CBIR systems.

7 Conclusion and future direction

In present day content based image retrieval approaches, deep learning based CBIR methods have reached miraculous attainment due to its potent learning competence. The numerous latest advancements of deep neural network based CBIR techniques have been collected in this paper. The deep binary codes with late fusion have shown improved accuracy for large scale image retrieval. The convolutional neural network pre-trained on ImageNet database has shown better performance and accuracy in a surveillance dataset. Due to hashing methods, an unsupervised visual hashing approach, called semantic-assisted visual hashing, has been used in CBIR application. This method attains better performance than other state of the art approaches. The use of multilayer CNN features for the CBIR performs better than single layer features and also presents a competitive performance on various popular CBIR benchmarks. The hashing based CBIR approach, with a deep convolution network, is used to produce hash codes and predicts labels of images. The CBIR method based on the bilinear root pooling, utilizes the compact and highly discriminative image feature set. The deep CNN based CBIR system has efficient retrieval performance and memory efficiency due to compact image feature descriptor. The CBIR model based on fusion of deep learning and compressed feature also achieve superior performance in terms of retrieval rate for both low and high features. The inclusion of the attention model in DAH improves the performance of this method in term of precision. The different model structures can be considered and their effect on retrieval performance can be analyzed as a future work. The MiLaN method is more fast and accurate than other CBIR used for remote sensing data retrieval applications. The survey is concluded with the following remarks:

1. Deep learning approaches can handle CBIR problems more efficiently than other state of the art method.
2. The deep learning based CBIR methods attains advancement in real application.
3. The dataset with more associated modality, need further research for validation in CBIR.
4. Features extracted using deep learning methods are more compact and more discriminative for different retrieval scenarios.
5. The resolution between retrieval accuracy and storage space can be further exploited and improved.

Finally, in this survey, some of the key deep learning based CBIR approaches have been investigated. The most recent developments have been compared and displayed in tabular form. This review will clarify future directions of research in the ever growing domain of deep learning based CBIR systems.

References

1. Ahmad J, Mehmood I, Baik SW (2017) Efficient Object-Based Surveillance Image Search using Spatial Pooling of Convolutional Features. J Vis Commun Image R 45:62–76

2. Ahmad J, Sajjad M, Rho S, Baik SW (2016) Multi-scale local structure patterns histogram for describing visual contents in social image retrieval systems. Multimed Tools Appl 75(20):12669–12692
3. Alzu'bi A, Amira A, Ramzan N (2017) Content based image retrieval with compact deep convolutional features. Neurocomputing 249:95–105
4. Arrospide J, Salgado L, Nieto M (2012) Video Analysis Based Vehicle Detection and Tracking using an MCMC sampling Framework, EURASIP J Adv Signal Process, pp. 1–20
5. Baltieri D, Vezzani R, Cucchiara R (2011) 3dpes: 3d People Dataset for Surveillance and Forensics, In Proceedings of the 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding, ACM
6. Bhagyalaksluni A, Vijayachamundeeswari V (2014) A survey on content based image retrieval using various operators IEEE international conference on computer communication and systems (ICCCS '14)
7. Cheng DS et al. (2011) Custom pictorial structures for re-identification, In BMVC
8. Chollet F (2017) Xception: Deep Learning with Depth-wise Separable Convolutions, In IEEE Computer Vision and Pattern Recognition, Hawaii, USApp. 1251–1258
9. Chua TS, Tang J, Hong R, Li H, Luo Z, Zheng Y (2009) Nuswide: A Real-world Web Image Database from National University of Singapore In Proc ACM Int Conf Image Video Retrieval, pp. 48:1–48:9
10. Dharani T, Aroquiaraj IL (2013) A survey on content based image retrieval. In: Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, pp 485–490
11. Felci Rajam I, Valli S (2013) A survey on content based image retrieval. Life Sci J 10(2):2475–2487
12. Gao Y, Beijbom O, Zhang N, Darrell T (2016) Compact Bilinear Pooling In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, pp. 317–326
13. Gao L, Song J, Zou F, Zhang D, Shao J (2015) Scalable multimedia retrieval by deep learning hashing with relative similarity learning. In: Proc ACM Int Conf Multimedia, pp 903–906
14. Glasner D, Galun M, Alpert S, Basri R, Shakhnarovich G (2012) Viewpoint aware object detection and continuous pose estimation. Image Vis Comput 30(12):923–933
15. Goel R, Sharma A, Kapoor R (2019) Object recognition using deep learning. in J Comput Theor Nanosci 16(9):4044–4052
16. Goodfellow I, Pouget-Abadie J, Mirza M, et al. (2014) Generative adversarial nets, Proc of International Conference on Neural Information Processing Systems, pp.2672–2680
17. Gordo A, Almazán J, Revaud J, Larlus D (2016) Deep image retrieval: learning global representations for image search, In Proceedings European conference on computer vision (ECCV)
18. Gordo A, Almazan J, Revaud J, Larlus D (2017) End-to-end Learning of Deep Visual Representations for Image Retrieval. Int J Comput Vis 124(2):237–254
19. Gray D, Brennan S, Tao H (2007) Evaluating Appearance Models for Recognition, Reacquisition, and Tracking, In Proc IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS), Citeseer
20. Guoqing J, Yongdong Z, Ke L (2019) Deep hashing based on VAE-GAN for efficient similarity retrieval. Chi J Electron 28(6):1191–1197
21. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. IEEE Comput Vis Pattern Recognit, Las Vegas, NV, USA, pp 770–778
22. Heesch D (2008) A survey of browsing models for content based image retrieval. J Multimed Tools Appl (Springer) 40:261–284
23. Huiskes MJ, Lew MS (2008) The Mir Flickr retrieval evaluation. In: Proc ACM Int. Multimedia Inf Retrieval, pp 39–43
24. Huiskes MJ, Thomee B, Lew MS (2010) New trends and ideas in visual concept detection: the MssIRFLICKR retrieval evaluation initiative. In: Proceedings of the International Conference on Multimedia Information Retrieval, pp 527–536
25. Jegou H, Douze M, Schmid C (2008) Hamming embedding and weak geometric consistency for large scale image search, In Computer Vision–ECCV, pp. 304–317
26. Kalantidis Y, Mellina C, Osindero S (2016) Cross-dimensional weighting for aggregated deep convolutional features. In: Workshop on Web Scale Vision and Social Media (VSM), European Conference on Computer Vision (ECCV), pp 685–701
27. Kingma D, Welling M (2013) Auto-encoding Variational Bayes, arXiv: 1312.6114
28. Kokare M, Chatterji BN, Biswas PK (2002) A survey on current content based image retrieval methods. IETE J Res 48(3–4):261–271
29. Li X, Xu M, Xu J, Weise T, Zou L, Sun F, Wu Z (2020) Image retrieval using a deep attention-based hash. IEEE Access 8:142229–142242
30. Lin TY, Roy Chowdhury A, Maji S (2015) Bilinear CNN Models for Fine Grained Visual Recognition, In Proceedings of the IEEE International Conference on Computer Vision, pp. 1449–1457
31. Liong VE, Lu J, Wang G, Moulin P, Zhou J (2015) Deep Hashing for Compact Binary Codes Learning, In IEEE Comput Vis Pattern Recognit, pp. 2475–2483

32. Liu P, Guo JM, Wu CY, Cai D (2017) Fusion of Deep Learning and Compressed Domain Features for Content Based Image Retrieval, IEEE transactions on image processing, vol. 26, no. 12
33. Liu L, Yu M, Shao L (2015) Multi-view alignment hashing for efficient image search. IEEE Trans Image Process 24(3):956–966
34. Liu Y, Zhang D, Lu G, Ma W-Y (2007) A Survey of Content Based Image Retrieval with High-Level Semantics. J Pattern Recognit (Elsevier) 40:262–282
35. Mohedano E, McGuinness K, O'Connor NE, Salvador A, Marques F, Giroi Nieto X (2016) Bags of Local Convolutional Features for Scalable Instance Search, In Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval, ACM, pp. 327–331
36. Mukherjee J, Mukhopadhyay J, Mitra P (2014) A survey on image retrieval performance of different bag of visual words indexing techniques. In: Proceedings of the IEEE Students' Technology Symposium, pp 99–104
37. Muller H, Michoux N, Bandon D, Geissbuhler A (2004) A Review of Content Based Image Retrieval Systems in Medical Applications—Clinical Benefits and Future Directions. Int J Med Inform 73:1–23
38. Ng HW, Winkler S (2014) A Data-driven Approach to Cleaning Large Face Datasets, In IEEE International Conference on Image Processing pp. 343–347
39. Ng J, Yang F, Davis L (2015) Exploiting Local Features from Deep Networks for Image Retrieval, In Proceedings of the IEEE Conf Comput Vis Pattern Recognit Workshops, pp. 53–61
40. Nister D, Stewenius H (2006) Scalable Recognition with a Vocabulary Tree. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2:2161–2168
41. Oussalah M (2008) Content based image retrieval: review of state of art and future directions. In: First Workshops on Image Processing Theory, Tools and Applications (IEEE), pp 1–10
42. Ozuysal M, Lepetit V, Fua P (2009) Pose Estimation for Category Specific Multi-view Object Localization, In IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2009, IEEE
43. Patel T, Gandhi S (2017) A survey on content based similarity techniques for image retrieval. In: International Conference on Innovative Mechanisms for Industry Applications (IEEE), pp 219–223
44. Patel F, Kasat D (2017) Hashing based indexing techniques for content based image retrieval: a survey. In: International Conference on Innovative Mechanisms for Industry Applications (IEEE), pp 279–283
45. Paulin M, Douze M, Harchaoui Z, Mairal J, Perronin F, Schmid C (2015) Local Convolutional Features with Unsupervised Training for Image Retrieval, In Proceedings of the IEEE International Conference on Computer Vision, pp. 91–99
46. Philbin J, Chum O, Isard M, Sivic J, Zisserman A (2007) Object retrieval with large vocabularies and fast spatial matching. In: Computer Vision and Pattern Recognition (CVPR), pp. 1–8
47. Philbin J, Chum O, Isard M, Sivic J, Zisserman A (2008) Lost in quantization: improving particular object retrieval in large scale image databases. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, pp 1–8
48. Qiu Z, Pan Y, Yao T, et al. (2017) Deep Semantic Hashing with Generative Adversarial Networks, Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval, Tokyo, Japan, pp. 225–234
49. Radenovic F, Tolias G, Chum O (2016) CNN image retrieval learns from BoW: unsupervised fine-tuning with hard examples, In Proceedings European conference on computer vision (ECCV)
50. Rafiee G, Dlay SS, Woo WL (2010) A review of content-based image retrieval. In: 7th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP 2010), pp 775–779
51. Rasiwasia N, Costa Pereira J, Coviello E, Doyle G, Lanckriet GR, Levy R, Vasconcelos N (2010) A New Approach to Cross-modal Multimedia Retrieval, In Proc ACMInt Conf Multimedia, pp. 251–260
52. Roy S, Sangineto E, Demir B, Sebe N (2020) Metric learning based deep hashing network for content-based retrieval of remote sensing images. IEEE Geosci Remote Sens Lett 18(2):226–230
53. Schwartz WR, Davis LS (2009) Learning discriminative appearance based models using partial least squares. In: 2009 XXII Brazilian Symposium on Computer Graphics and Image Processing. IEEE, pp 322–329
54. Shen F, Shen C, Liu W, Shen HT (2015) Supervised discrete hashing, In IEEE Computer Vision and Pattern Recognition, pp. 37–45
55. Shen F, Shen C, Shi Q, van den Hengel A, Tang Z, Shen HT (2015) Hashing on nonlinear manifolds. IEEE Trans Image Process 24(6):1839–1851
56. Shen X, Shen F, Sun QS, Yuan YH (2015) Multi-view latent hashing for efficient multimedia search. In: Proc ACM Int Conf Multimedia, pp 831–834
57. Szegedy C, Liu W, Jia Y, Sermanet P (2015) Going deeper with convolutions. In: IEEE Computer Vision and Pattern Recognition (CVPR), pp 1–9

58. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the Inception Architecture for Computer Vision, In Proc IEEE Conference Computer Vision Pattern Recognition (CVPR), pp. 2818–2826
59. Tang J, Li Z, Zhu X (2018) Supervised deep hashing for scalable face image retrieval. Pattern Recogn 75: 25–32
60. Tolias G, Sicre R, Jegou H (2016) Particular object retrieval with integral max pooling of CNN activations. In: International Conference on Learning Representations (ICLR), pp 1–12
61. Torralba A, Fergus R, Freeman W (2008) 80 million tiny images: a large dataset for nonparametric object and scene recognition. IEEE Trans Pattern Anal Mach Intel 30(11):1958–1970
62. Tzelepi M, Tefas A (2018) Deep convolutional learning for content based image retrieval. Neurocomputing 275:2467–2478
63. Veltkamp RC, Tanase M (2002) Content-based image retrieval systems: A survey. Multimed Sys Appl Ser (Springer) 21:47–101
64. Wan J, Wang D, Hoi SCH, Wu P, Zhu J, Zhang Y, Li J (2014) Deep learning for content based image retrieval: a comprehensive study. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp 157–166
65. Wang B, Yang Y, Xing X, Alan H, Heng TS (2017) Adversarial cross-modal retrieval. In: Proc of ACM on Multimedia Conference, pp 154–162
66. Wolf L, Hassner T, Maoz I (2011) Face Recognition in Unconstrained Videos with Matched Background Similarity, In IEEE Computer Vision and Pattern Recognition, pp. 529–534
67. Wu S, Oerleman A, Bakker EM, Lew MS (2017) Deep binary codes for large scale image retrieval. Neurocomputing 257:5–15
68. Yang H, Lin K, Chen C (2017) Supervised learning of semantics-preserving hash via deep convolutional neural networks. IEEE Trans Pattern Anal Mach Intell 40(2):437–451
69. Yang L et al. (2015) A large-scale Car dataset for fine grained categorization and verification, In Proceedings of the IEEE conference on computer vision and pattern recognition
70. Yu W, Yang K, Yao H, Sun X, Xu P (2017) Exploiting the complementary strengths of multilayer CNN features for image retrieval. Neurocomputing 237:235–241
71. Yua W, Yangb K, Yaoa H, Suna X, Xub P (2017) Exploiting the complementary strengths of multilayer CNN features for image retrieval. Neurocomputing 237:235–241
72. Zhao M, Zhang H, Sun J (2016) A novel image retrieval method based on multi trend structure descriptor. J Vis Commun Image Represent 38:73–81
73. Zhu L, Shen J, Xie L, Cheng Z (2017) Unsupervised visual hashing with semantic assistant for content based image retrieval. IEEE Trans Knowl Data Eng 29(2):472–486

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

2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP - 2021)

2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP) | 978-1-6654-3277-1/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCCSP52374.2021.9465347

Survey and Comparison of Video Summarization Techniques

Anil Singh Parihar, Ritvik Mittal, Prashuk Jain and Himanshu
Dept. of Computer Science
Delhi Technological University
New Delhi, India
{parihar.anil, mittalritvik.rm, himanshubhushan775, prashuk156}@gmail.com

***Abstract*—Video summarization is a keenly intellective video compression technique to select a subset of keyframes or keyshot which are combined to represent shorter and compendious summary of the original input video without losing the contextual semantics of the same. In context of summary generated, it can be divided into static (static storyboard) and dynamic (video skimming) video summarization. Supervised ,unsupervised and reinforcement learning techniques have been introduced in the literature of video summarization. Earlier development in this field marked the introduction of various unsupervised techniques that used handcrafted heuristics to select mutually independent keyframes. In recent years, supervised techniques and deep reinforcement learning techniques have been introduced that models the structural semantics of the original video to generate summaries using frame - level ground truth generated by humans, so that we get the result as close as possible to the human understanding of the video. This paper aims to introduce different architectures proposed for video summarization and provides a qualitative and quantitative comparison of these methods.**

***Index Terms*—video, summarization, survey, quantitative, qualitative**

I. Introduction

With the advent of high speed Internet and low storage cost, the amplitude of data has incremented dramatically and the major part of it is represented in the form of visual data or videos. With such humongous magnitude of data, we require efficacious techniques and implements to take these videos and represent them in a more compact and concise way so that these can be utilized for sundry applications like video browsing, video surveillance analysis, egocentric video analysis to study human comportment, medical video analysis like endoscopy analysis, etc. These requisites magnetized researchers and technologists and paved the way for incipient research areas like video summarization.

Video summarization is usually considered to be a selection optimization problem and models are built to select such frames or shots which store the highest representative value amongst the frame. There are various representations that have been considered as an output to video summarization tasks like video synopsis [9], montages [1, 10, 11] and storyboards [12, 13]. This computer vision task of video summarization can be divided into precisely two subgroups - static video summarization and dynamic video summarization. Static video summarization works at the level of frames and considers it as a keyframe selection problem. In dynamic video summarization, keyshots are selected taking into account the domain representation of the frames. The major advantage of dynamic video summarization is that it introduces motion as well as audio element of the video into the summary. Audio and motion help to improve the expressiveness of the summary and keep the viewer more engaged and intrigued. But dynamic video summarization also needs synchronization of the keyshots and hence engenders bottlenecks on the flexibility of the output summary.

Initially research work was majorly carried out in static video summarization but now with the amplitude of resources available, researchers are moving towards producing models that work on dynamic video summarization. In recent years, deep learning has been extensively used for the task of video summarization. In this context, video summarization can be subdivided into supervised and unsupervised tasks. Earlier, usually unsupervised models were introduced for the task that leveraged low level feature extraction like color distribution, pixel values, etc and clustering techniques to select keyframes. But now supervised models are being designed for the task which use human annotated video dataset to model architecture that could generate summary as close as possible to human interpretation of the video. Due to the subjective nature of annotations, it is very difficult to design an established evaluation metric. Nevertheless, F-score is a generally accepted evaluation metric used for quantitative analysis of various models.

In this paper a detailed survey is provided for various models that have been introduced for video summarization and a comparative analysis is generated including both quantitative and qualitative aspects.

978-1-6654-3277-1/21/$31.00 ©2021 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:38:55 UTC from IEEE Xplore. Restrictions apply.

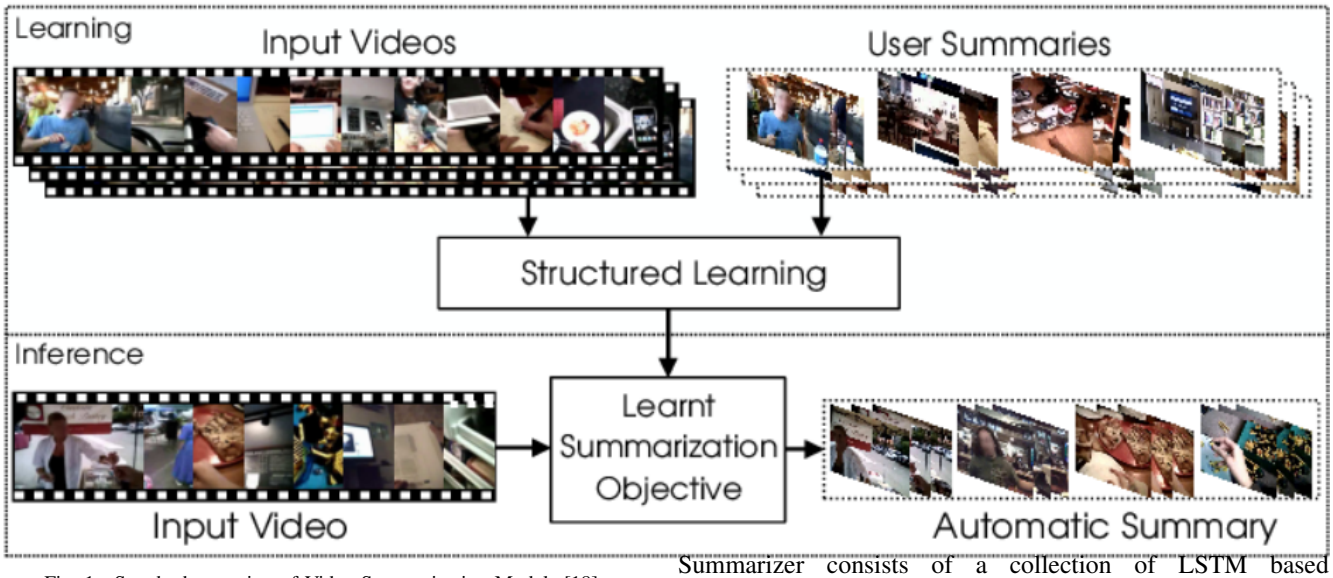


Fig. 1. Standard operation of Video Summarization Models [18]

II. Related Work

Sandra Eliza Fontes de Avila et al. [1] proposed an efficient unsupervised video summarization approach known as VSUMM. It extracts features from some sample video frames in the HSV color space and then applies K-means clustering on the extracted features to group them into different clusters. Then a representative frame is chosen for each cluster and the summary is generated. VSUMM also introduces a new evaluation technique known as CUS (Comparison of User Summaries). In CUS, users create a summary for the input video and that summary is compared with the summary created with the model. A frame in the user summary is considered similar to that created by the model if the Manhattan distance between them is less than some predefined value delta and consequently the accuracy and the error is calculated.

Ke Zhang et al. [2] leveraged the adeptness of LSTMs in extracting the structural and contextual knowledge from a sequence towards video summarization. A BiLSTM network is proposed to model the structure of the video considering both the directions. The network is trained in a supervised setting against the user generated summaries to distillate the semantics of how a user perceives summarizing a video. One other major contribution was standardization of techniques to deal with unavailability of user annotated video summaries.

Behrooz Mahasseni et al. [3] introduced a unique adversarial learning based model for unsupervised video summarization. The architecture focuses on learning a summarizer network to generate video summaries which are evaluated against the original video using a discriminator network. Summarizer consists of a collection of LSTM based network modules for frame selection, encoding and decoding. Discriminator is modeled as a LSTM network which takes as input original and generated summaries and tries to differentiate them. The summarizer is trained to maximally confuse the discriminator thus minimising the difference between original and generated summary. LSTMs are used as they can model long range semantic dependencies between the temporally ordered input frames.

Mrigank Rochan et al. [4] casts video summarization as a binary sequence labeling problem and establishes a relationship with a far more researched topic of semantic segmentation. [4] thus leverages the use of fully convolutional networks [9] for video summarization. The network architecture is in the form of an encoder decoder network. Stack of convolutions and pooling layers aggregate the contextual information between the frames and form the encoder. The reduced temporal dimensions are then recovered via upsampling of the encoder outputs. The whole architecture is then realised as an end-to-end convolutional network. Main intuition is repeated convolutions can effectively capture range and structural dependencies much like a LSTM but are easier to train.

Kaiyang Zhou et al. [5] proposed the first reinforcement learning based unsupervised video summarization technique. In this technique the reward function calculates how diverse and representative the selected frames are for the input video. The diversity function was proposed in such a way that two distant similar frames are considered different as they can be important for the storyline construction. Diversity value is calculated by taking the pairwise dissimilarity among selected video frames. No labelling of frames is required for this approach however the model is extended by considering log probability of selecting the annotated keyframes when labels are present. The paper proposed a DSN (Deep Summarization Network) which predicts a probability for each frame which represents the probability that the frame is selected. DSN is

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:38:55 UTC from IEEE Xplore. Restrictions apply.

an encoder decoder network where encoder is a CNN network which extracts features from the input video and the decoder is a Bi-LSTM network that produces the probabilities for frame selection. Pre-trained GoogleNet is used as the CNN encoder network and only decoder part is trained.

Zhong Ji et al. [6] proposed an attention based encoder decoder network where the encoder is a Bi-LSTM network and the decoder is an attention based LSTM network. The proposed model is named AVS (Attentive encoder decoder networks for video summarization). The input to AVS is the video frames and the output of the AVS is the frame scores. A key shot selection model is then applied on the frame level importance scores to generate a shot level video summary. The proposed model improves upon previous Bi-LSTM encoder decoder approaches by allowing the context vector, that flows from encoder to decoder, to be of variable size. In previous Bi-LSTM encoder decoder networks, no matter what kind of output shots/frames were to be predicted, all the shots/frames in the input video series had the same importance. Those methods risked missing much of the temporal context underlying the video because of this indiscriminate averaging of all the frames. The proposed model gives different importance to different frames using the attention based decoder network thus trying to generate the summary like a human does.

Jiri Fajtl et al. [7] proposed a supervised learning approach with a Soft Attention based Simple Neural Network as it hypothesizes that video summarization is rather a subjective task. The paper proposed a new soft attention based simple network for sequence to sequence transformation and demonstrates that the proposed network is better and is less complex than the already used Bi-LSTM based encoder decoder network with soft attention. The proposed network is a combination of a soft attention network and a regressor network. In some Bi-LSTM based encoder decoder networks for video summarization, the state that is passed from encoder to decoder network is of fixed length, so if the input to the network is a sequence of longer length then the state passed from encoder to decoder will have more information loss. The proposed network doesn't suffer from such losses.

Jungin Park et al. [8] emphasizes on modeling the semantic relationships between the frames with respect to the story of the whole video. It drifts from other methods in casting the problem as graph modeling problem in which individual frames represent nodes while edges represent the relationship between these nodes. An initial graph constructed using frame level features and represented by an adjacency matrix is iteratively refined using Graph Convolution outputs as adapted feature representations for each step. This architecture has an effect that the neighbourhood of each node is updated depending on the effective context at each step. The final frame features and relations are fed to a summary GCN to classify whether it is part of the generated summary. Network can be trained in supervised as well as in unsupervised setting.

III. Datasets

Supervised training requires user generated summaries of training videos. TVSum [16] and SumMe [15] are two annotated datasets which contain videos summarised by multiple users. These datasets are the benchmark for testing video summarization models. TVSum has 50 videos with categories ranging across news, documentaries, etc. and video length 1-5 minutes. It provides user generated summaries as frame level importance values. SumMe dataset has 25 videos of various events across holidays, sports etc. ranging from 1.5 to 6.5 minutes in length. User generated summaries are key shot based (set of frame intervals). OVP(open video project) [1], [17] dataset has 50 videos covering genres such as historical, educational, etc. These are 1-4 minutes in length.The YouTube [1] dataset contains 39 videos distributed among news, sports,commercials, etc. These are 1-10 minutes in length. Both OVP and YouTube dataset provide multiple user summaries as key-frames. Since different datasets provide user annotated summaries in different formats, we follow [2] for pre-processing data for training of supervised models and evaluation.

IV. Comparative Analysis

We follow the commonly used evaluation metric [2] and use F-Scores for quantitative comparison. It is most commonly used metric across the work we have considered and hence provide for fair analysis. F score is based on precision(P) and recall(R). Given G as ground truth summary and M as machine generated summary:

$$P = G \cap M / \ |M| \qquad R = G \cap M / \ |G| \tag{1}$$

where $|.|$ indicates summary length. F-Score is computed as:

$$F = (2P * R)/(P + R) * 100\% \tag{2}$$

F scores on augmented [2] setting of the datasets where training is performed on augmentation of OVP, YouTube, X1, 80% of X2 and testing on 20% of X2 are given. X1 and X2 being TVSum and SumMe interchangeably. Table I shows the quantitative analysis based on F-score of the various research models we have studied while Table II presents a brief qualitative analysis on these techniques.

TABLE I
F scores of Analysed models

Paper	SumMe	TvSum
[1]	33.5	48.6
[2]	42.9	59.6
[3]	43.6	61.2
[4]	51.1	59.2
[5]	43.9	59.8
[6]	46.1	61.8
[7]	51.1	62.4
[8]	52.9	65.8

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:38:55 UTC from IEEE Xplore. Restrictions apply.

TABLE II
QUALITATIVE ANALYSIS OF MODELS

Author	Approach	Advantages	Drawbacks
Sandra Eliza Fontes de Avila et al. [1]	In this approach, Hue color histograms as feature desciptors and K Mean clustering selects clusters which are visually diverse	1. Colour histograms are used as low level feature descriptors which are robust to small changes in camera position 2. This approach produces good results using very less resources as the time and computer resources required to form clusters is very less compared to neural net architectures.	1. Clustering Algorithm does not take into account the temporal order 2. Colour histograms do not take into account the orientation or the colour spread in the frame. Thus an image with the same colours distributed very differently gets considered similar to an image with same colours distributed differently
Ke Zhang et al. [2]	The model uses BiLSTM architecture. Output from these LSTM layers is coupled with visual frame feature into multi-layer perceptron to yield either binary frame label or frame level importance score at each temporal step.	LSTMs model the variable temporal dependencies effectively because of persistent past knowledge.	1. The computations of recurrent models cannot be parallelised hence leading to a waste of computing resources. 2. Poor performance on modelling dynamically changing scenarios as structural order is not maintained.
Behrooz Mahasseni et al. [3]	The architecture learns a keyframe selector LSTM by using variational autoencoder to generate corresponding summary which is then fed to confuse the discriminator in an adversarial setting.	1. Unsupervised Approach thus no need for annotated user data which is hard to obtain. 2. Learns Intermediate representations maybe useful for other applications. 3. Adversarial feedback maybe more useful than user ground truths in some cases. 4. Regularization can be introduced at various instances thus focusing on varying semantics.	1. GANs are notorious to train both in time and memory. 2. Subtle changes which may be important to users are hard to capture as focus is on achieving global representation. 3.User semantics in generating a video summary are not captured.
Mrigank Rochan et al. [4]	A fully convolutional model is adapted for video summarization.	1. Convolution models support parallel computation as they are not dependent on the previous results. This provides for efficient training compared to LSTM approaches. 2. Compared to LSTMs, CNNs can model whole range at much smaller depth allowing for high level context aggregation earlier in the network. In LSTMs, the last node is only considered once.	1. Repeated down-sampling of the inputs by the stack of convolution layers results in loss of resolution and low level semantics thus biasing the output towards contextual knowledge only disregarding local knowledge. 2. Upsampling by a large factor results in spreading of a small number of values to a large space thus affecting uniqueness and rendering many nodes similar.
Kaiyang Zhou et al. [5]	Reinforcement learning framework for training of Encoder Decoder Architecture. Rewards are based on diversity and representativeness of generated summary.	1. No requirement of annotated data for training. 2. Directly modelling diversity and representativenss provides better results than reconstruction loss [3].	1. Focus on diversity introduces problems in case of subtle or slow changes. 2. LSTMs cause inefficient training. 3. Does not include any reward towards long temporal dependencies.
Zhong Ji et al. [6]	Attention based encoder decoder network where the encoder is a Bi-LSTM network and the decoder is an attention based LSTM network.	Attention mechanism provides attention weights associated with individual encoder hidden states. Decoder output at a single instance can learn to combine different hidden units through these weights resulting in rich structural knowledge. This is particularly beneficial against encoded vector representation of inputs which provide fixed knowledge. The latter technique is common in other LSTM encoder decoder techniques [2], [3] and results in loss of temporal structural especially in case of large inputs.	The model is computationally expensive to train as both the encoder and decoder are LSTM based networks. More memory and time requirements to compute attentions at each time step based on all encoder outputs.
Jiri Fajtl et al. [7]	The model is based on evaluating a soft, global attention layer to evaluate a context vector using attention weights. Context vector along with residual sum is then used for frame level score regression.	1. Provides the advantage of attention mechanism as in [6] while being computationally efficient and easier in implementation. Does not involve any sequential processing and just 2 fully connected layers. 2. Proposed model prevents compression losses as in encoder decoder architectures by accessing the input features directly. 3. Impressive results without taking temporal structure into account.	1. No consideration of temporal semantics. 2. Slight increase in accuracy on the TvSum dataset as compared to the SumMe indicates global attention is not desirable for long videos. Hence local attention is required.
Jungin Park et al. [8]	Initialised graph from frame features affinity is refined recursively using Graph Convolutional Networks which follow aggregating features from previous steps and updation of adjacency matrix.	Frames are related to each other following exploration of similarities among themselves and also through global context as the graph is refined. The limited range of convolutional and recurrent operations cannot model these dependencies to this extent. Hence this model effectively captures frame level relationships.	1. Distance between frames is not accounted and temporal structure is not modeled. 2. Iterative refinement by computing pairwise scores between frames may lead to performance bottlenecks in case of long videos.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:38:55 UTC from IEEE Xplore. Restrictions apply.

V. Conclusion

We have presented different architectures published for video summarization. For each approach, we first introduce the theory and then analyse the architecture and results to furnish a lucid comparison. Through this work, we provide a base for exploring diverse video summarization techniques. We hope this study facilitates current apprehension and future advancement in video summarization.

Our major aim has been to exhibit diverse methodologies for video summarization. To this end, we could not consider some of the other excellent work that has been published for the scope of this study.

References

[1] S. E. F. De Avila, A. P. B. Lopes, A. da Luz, and A. de Albuquerque Araujo. Vsumm: A mechanism designed to produce ´ static video summaries and a novel evaluation method. Pattern Recognition Letters, 32(1):56–68, 2011.

[2] Zhang, K., Chao, W.L., Sha, F., Grauman, K.: Video summarization with long short-term memory. In: ECCV (2016)

[3] Mahasseni, B., Lam, M., Todorovic, S.: Unsupervised video summarization with adversarial lstm networks. In: CVPR (2017)

[4] Rochan, M., Ye, L., Wang, Y.: Video summarization using fully convolutional sequence networks. In: ECCV (2018)

[5] Zhou, K., Qiao, Y., Xiang, T.: Deep reinforcement learning for unsupervised video summarization with diversity-representativeness reward. In: AAAI (2018)

[6] Ji, Z., Xiong, K., Pang, Y., Li, X.: Video summarization with attention-based encoder-decoder networks. arXiv preprint arXiv:1708.09545 (2017)

[7] Fajtl, J., Sokeh, H.S., Argyriou, V., Monekosso, D., Remagnino, P.: Summarizing videos with attention. In: ACCVW (2018)

[8] Park, Jungin et al. "SumGraph: Video Summarization via Recursive Graph Modeling." Lecture Notes in Computer Science (2020): 647–663. Crossref. Web.

[9] Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition (2015)

[10] Y. Pritch, A. Rav-Acha, and S. Peleg. Nonchronological video synopsis and indexing. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(11):1971–1984, Nov 2008.

[11] H. wen Kang, Y. Matsushita, X. Tang, and X. quan Chen. Space-time video montage. In CVPR, pages 1331–1338, 2006.

[12] M. Sun, A. Farhadi, B. Taskar, and S. Seitz. Salient Mon- tages from Unconstrained Videos, pages 472–488. Springer International Publishing, Cham, 2014.

[13] D. B. Goldman, B. Curless, S. M. Seitz, and D. Salesin. Schematic storyboarding for video visualization and editing. ACM Transactions on Graphics (Proc. SIGGRAPH), 25(3), July 2006.

[14] Y. J. Lee, J. Ghosh, and K. Grauman. Discovering important people and objects for egocentric video summarization. In CVPR, pages 1346–1353, 2012.

[15] Gygli, M., Grabner, H., Riemenschneider, H., Van Gool, L.: Creating summaries from user videos. In: ECCV. (2014)

[16] Song, Y., Vallmitjana, J., Stent, A., Jaimes, A.: Tvsum: Summarizing web videos using titles. In: CVPR. (2015)

[17] : Open video project. http://www.open-video.org/

[18] Gygli, Michael et al. "Video summarization by learning submodular mixtures of objectives." 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015): 3090-3098.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:38:55 UTC from IEEE Xplore. Restrictions apply.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/352681664

Surviving International Competition: Analysis of Measures Employed By Local Companies

Research · June 2021

DOI: 10.13140/RG.2.2.31829.76000

CITATIONS
0

READS
19

2 authors:

Vanshika Notani
University of Delhi
1 PUBLICATION **0** CITATIONS
SEE PROFILE

Piyush Jain
Delhi Technological University
2 PUBLICATIONS **0** CITATIONS
SEE PROFILE

All content following this page was uploaded by Piyush Jain on 23 June 2021.

The user has requested enhancement of the downloaded file.

International Journal of Management (IJM)
Volume 12, Issue 6, June 2021, pp. 231-239, Article ID: IJM_12_06_020
Available online at https://iaeme.com/Home/issue/IJM?Volume=12&Issue=6
ISSN Print: 0976-6502 and ISSN Online: 0976-6510
DOI: 10.34218/IJM.12.6.2021.0020

© **IAEME** Publication **Scopus** Indexed

SURVIVING INTERNATIONAL COMPETITION: ANALYSIS OF MEASURES EMPLOYED BY LOCAL COMPANIES

Vanshika Notani
Department of Commerce, Sri Guru Tegh Bahadur Khalsa College,
University of Delhi, Delhi, India

Piyush Jain
Department of Production and Industrial Engineering,
Delhi Technological University, Delhi, India

ABSTRACT

The modern world is stepping towards a globalised economy, and not even a single company can shut its eyes to globalisation and escape the challenges caused by globalisation. This paper critically examines the challenges faced by the local companies due to the introduction of Globalisation. It also analyses the impact of global challenges on the local companies. The persistent outstanding barrier for the small and local companies is market competition. Globalisation stimulates competition from large ventures and international companies. MNCs have the capability to eliminate global challenges than small companies. In the matter of reduced barriers and receptiveness of the world, the inexorable procedure of globalisation portrays the projects which are touching the success potentials and risk potentials. Innumerable challenges have been created by the economic globalisation for small and local companies owing to expeditious climb in the competition. Hence, the collapse rate of small companies is climbing the ladder. Correspondingly, small and local companies need to ensue the survival strategies mentioned in the review in order to succeed in challenging the global challenges. Arising out of the industrial revolution until today, irrevocable economic integration persists. The rationale is straightforward, business and profits do not acknowledge the boundaries. Wherever nearly a give and take relationship is created, a business connection is promptly established.

Key words: Globalization, SMEs, Survival Strategies, International Competition, Internationalization, Economic Integration, Transnational Business

Cite this Article: Vanshika Notani and Piyush Jain, Surviving International Competition: Analysis of Measures Employed by Local Companies, *International Journal of Management (IJM)*, 12(6), 2021, pp. 231-239.
https://iaeme.com/Home/issue/IJM?Volume=12&Issue=6

1. INTRODUCTION

With the tumbling down/collapse of the isolationism of the emerging markets all over the globe, Multinational Companies started to look out for new opportunities. They rushed and rushed all around the world to grab those opportunities for their growth. The entrance of multinational companies was a boon to the consumers as they got the opportunity to choose from a wide variety of options. However, for the local/domestic companies, the inundation emerged to be a death sentence. Having regular dominant positions in the markets, the local companies suddenly faced the foreign rivals flashing a formidable array of benefits, including powerful brand name, expert management and marketing skills, advanced technology, sizable financial resources, and worthier products. With this, it was crystal clear that the continuance of the local companies was at stake.

A huge amount of funds with the Multinational Corporations facilitated them to hire such knowledgeable talent who proved to be an asset for them and advised them on how to enter the emerging markets by forming a detailed plan of the strategic measures while on the other hand, the managers of the local companies have had a very little knowledge and guidance on the same which proved to be a significant disadvantage for them. The critical question was, How can such local companies vanquish and even benefit from their differences with the competitors from ultra-modern industrial countries?

This was all possible because of Globalisation. After the turn of the millennium, Globalisation has become a ferocious and recurring issue for debate. Over a long period and discussion, it has been called a requisite and even an inescapable process that will act as a helping hand in curing all the world's ill, only if we accept and embrace it. Few have also negated the above opinion as they consider it to be enhancing the few but causing sufferings for many.

Even after acting as a turmoil for the local companies, it is undeniable that globalisation has and continues to have an intense effect/consequence on the world. It has completely changed how companies and the government do their business. As per the World Bank data, trade accounted for more than 70% of Global Gross Domestic Product in 2017, from less than 25% in 1960.

1.1. What Is Globalisation ?

Globalisation is a concept that illustrates the means by which regional and national economies, cultures and societies have become unified with the help of a global network of transportation, communication, information, trade and immigration. There is still a debate concerning the beginning of globalisation, where few sources place it as long as the bronze age, approx 4000 years ago. Generally, this is contemplated as the first wave of globalisation wherein the trade crossed seas and stayed in for more than a century.

By the 19th century, Globalisation landed into a new chapter. The emergence of technology-led international trade to move a step further, accelerate the pace and the range with which the business could be managed profitably. A range of institutional changes came with the technological developments that led to the integration of trade even further. The establishment of the United Nations in 1945 focussed on bridging the gaps between countries having economic development as their primary belief.

Simultaneously, the formation of the General Agreement On Tariffs and Trade, which was the first worldwide multi-trade free trade agreement, led to the facilitation of trade through the elimination of tariffs. Eventually, it was triumphed by the formation of the World Trade Organisation.

All these are the foundation of the current wave of globalisation whereby the figures of computers are also very heavy. Though the computers were existent in some form since at least the 1940s, the formation of Intel's first microprocessor in 1971 pronounced the great leap forward towards 'Modern Computing.' This led to the beginning of the trend in which companies were able to utilise computing to raise access to information and eventually unseal the economic potential.

The phase after World War II saw prompt integration and development of global trade. The surge of business and technological innovation fueled the sudden uptick in globalisation, which resulted in cross border trading flourishing at a pace unseen formerly. Eventually, this led to an expansion in the value of trade of over 4000% between 1913 and 2014, whereby the more significant share of the expansion occurred after the 1950s.

There exist two key domains of globalisation, that is, Globalisation Of Production which is labelled as a source of goods from various countries in the world to procure the gains from differentiation in cost of production, and the other one is Globalisation Of Market Places which lessens the cross border trade constraints.

Globalisation was first felt in the 1990s by India when Dr Manmohan Singh, then Finance Minister, commenced the economic liberalisation plan. Prior to 1990, India's economy concentrated on equity overgrowth. In 1991, India started to liberalise its economy by reducing the control of the government over many of its local industries and expanding its openness to the rest of the world. When combined, these measures led to the expansion of competitiveness.

Prior to this scenario, primarily, local companies were able to sell anything and everything without facing much competition, but the crammed international market means that a specialised offering in a product that offers a comparative advantage is easier to find a home for – however niche that might be. Competitive Advantage means the factors or the country's ability to produce goods or services better or more efficiently and cheaper than the others or its rivals.

This switch has led to great importance for international trade of intermediary products and additionally for the expert professional services, which is possible due to distribution of competitive advantage, for example, developing nation such as India in general relish competitive advantage in labour, which enables them to produce simpler intermediary goods effortlessly. This also affected the competition. After Globalisation, anyone who has enough capital and other resources can easily enter the market. Due to this, the monopoly of domestic companies has come under fire from beyond a country's borders.

1.2. After Effects of Globalisation

Such small businesses and domestic companies have to face many challenges, the most prominent ones are the competition and the price wars. An analogy may define this situation: suddenly, small businesses are like small fish in a big sea with other small and large fish. The other smaller species compete for the same food, while the larger fish simply eat the smaller ones to keep themselves going.

Due to Globalisation, large corporations have expanded into the countries that formerly might have been off the radar. For instance, 'Coca Cola Sold Here' signboards can be seen even on the mountain roads of India; streets of even small towns are littered with KFC and McDonald's. Taking advantage of economies of scale, such big corporations are able to provide their products at a fraction of cost and to compete with them, local companies are also forced to reduce their prices.

1.3. Downturns of Globalization

Market Floods

While competing with large scale international companies, one of the major difficulties that small local organisations experience is the global reach of those big businesses. For instance, almost every country struggles to surmount China's capability to mass export the goods at such prices which are undefeatable and almost impossible to compete with, even including the packaging and shipping costs, which makes it a difficult task for local companies to stay alive. Despite the attempt to provide unique customer solutions and superior customer service, the local companies eventually have to cut down their prices in competition with the MNCs.

Few facets of the service industry grieve as well, as is the concern with many phone-based consumer support businesses who fail to compete with foreign markets.

Internationalisation of Local Competition

It was uncomplicated for the local businesses to sustain a robust existence in a community due to the lack of acceleration of businesses across boundaries and the nature of the spread of information. In recent times, the Internet has taken over all the world, the whole world can read each and every piece of information in just a couple of minutes. In such times, simply being the one and the only shop in the town is no longer a guarantee of success.

As global entities have huge resources which enable them to access the consumer's data, taste and preferences. Utilising that pool of data, the global entities can tailor the products and services according to customers' choices, which is difficult for a local/small business owner to counter. It is not just the issue of price wars, but rather how a large company with adequate data, research, and facilities can eradicate a local market.

Global Market Chain Reactions

Every business feels the sting of large-scale changes through shifts in local customer sentiments, changes in the cost of raw materials or simply a shift from local suppliers to distant competitors. Recent examples are reasonably varied, but even situations that may not have an apparent economic impact can cause ripples that reach unlikely businesses; The recent EU focus on Brexit, for example, has led to uncertainty and damage in the realm of small UK businesses.

With the increase in speed and universality of Globalisation, the local companies had no other option left with them but to internationalise themselves. Those companies a pool of opportunities were awaiting them to perform in the international markets. Developing and adapting to the strategies for internationalisation becomes a necessity to reshape their organisation into a globally competitive company completely. They need to harmonise the production, marketing, supply and other activities on the basis of international activities. Globally assembling and organising the company is a huge global challenge and necessitates organisational skills, strategic positioning, a high degree of coordination and integration, attentiveness to the wants and needs of individual markets.

2. THEORETICAL APPROACHES

The below-mentioned theories explain the Impact of Global Challenges and Survival Strategies for the local companies:-

2.1. Keynesian Business Cycle Theory

There is no distinctive theory in Economic Literature to explain the effect of global challenges on firms and the strategic measures employed by the firms, majorly due to the dynamic nature of global challenges. Keynes brought forward a new theory named Keynesian Business Cycle

Theory, on the basis of the Great Economic Depression experienced in the 1930s. The theory elucidated the economic crisis (Keynes 1933). One of the most important global challenges that most business entities face is the economic and financial crisis, particularly SMEs in a globalised economy.

The Keynesian Theory uses two concepts to explain the situation: Marginal Efficiency Of Capital and Effective Demand. Particularly, the theory holds low demand in the economy as a root of economic depression, which further leads to low production and high unemployment due to stock hoarding in the economy. Additionally, it also associates with the inadequacy of financial liquidity or liquidity trap among the investors in the economy. Due to the extra vulnerability of small enterprises to external shocks, it becomes challenging for them to sustain their production level or standstill the production for a certain period of time at the time of the global economic crisis.

Therefore, Keynesian Theory recommends the government's intercession in the economy to magnify aggregate demand, particularly using fiscal and monetary instruments. Hence, the Keynesian Theory of Business Cycles puts forward an elucidation on the economic crisis that forms a global challenge for business enterprises and policy intervention. Most of the countries employed Keynesian view in economic crisis management during the latest economic downturn due to Covid-19.

2.2. Theory Of Comparative Advantage

The Theory Of Comparative Advantage can be utilised as an instrument to alter and adapt the production level in accordance with that of the competition as this helps in allocating the resources to the efficient sectors of the economy. Whenever a country produces goods/services at a lower opportunity cost than that of other countries, then the comparative advantage is achieved. Hence, this lets the company sell their goods/services at a lower price than that of its competitors, which further helps them increase their sales margin. In small enterprises, the trade of low-cost items, considering all the other specific attributes such as quality and price, is salient to achieve a competitive advantage from business internationalisation.

2.3. SWOT Analysis

SWOT Analysis stands for Strengths, Weaknesses, Opportunities and Threats, which lets an organisation recognise its external environmental factors (Opportunities and Threats) and internal environment factors (Strengths and Weaknesses). Many organisations perform SWOT Analysis as a part of their strategic planning process prior to implementing those strategies.

The situation inside the company or what occurs inside the company forms a great source of data for the strengths and weakness category of SWOT. The Internal Environment Analysis is associated with the factors relating to the contentment of consumer needs, which can be ascertained by the company's resources, culture, structure, and functional areas.

What occurs outside the company is equally important as the internal factors to the success of the company. The External Environment Analysis serves as a helping hand to the organisations in identifying the opportunities and threats in the outside environment for the business. However, the data available indicates that the small companies have minimal knowledge of both the external and internal environment operating in the negligible understanding of the SWOT Analysis. Nonetheless, big corporations and small companies can perform SWOT Analysis for their business prior to implementing their corporate strategies as this strategic planning will ultimately lead to a sustainable competitive advantage for the firms.

2.4. Porter Generic Strategies

The generic strategies are given by Porter assist in determining the strategic position of a business in order to acquire a competitive advantage. On the basis of the market where organisations compete (narrow or broad) and the source of competitive advantage (differentiation and cost), Michael Porter stated four generic strategies, named, Cost Focus, Cost Leadership, Differentiation and Differentiation Focus. The choice of the competitive scope is entirely upon the organisation, as the scope differentiates between the organisations aiming at a narrow industry segment or a broad industry segment. It is crucial not only for the large firms but also for the small enterprises to analyse it's the strategic position of the firm.

2.5. PESTEL Analysis

PESTEL symbolises Political, Economic, Social, Technological, Environmental and Legal Factors. PESTEL can be seen as a strategic planning and market research instrument that assists organisations in recognising the macro-environmental components that have a significant impact on the business and helps in making better and sound decisions.

The two fundamental functions of PESTEL include identifying the environment in which business is operating and furnishing the insights and details to the businesses that enable them to predict future circumstances and situations. Small companies need to scan the external environment in order to scrutinise the political, economic, social, technological, environmental and legal impact on the business prior to the formation and implementation of business plans. Simultaneously, the small companies need to perform PESTEL Analysis to be mindful of their competitors as such large organisations have an extremely high potential for promoting entrepreneurship, providing employment and earning high foreign exchange to the country.

2.6. Resource-Based View

A resource-based view can be considered a technique to draw out a strategy that aims to achieve a competitive advantage in an organisation's resources. Such an approach lets the organisation find a competitive advantage within themselves rather than outside the organisation. Simultaneously, a resource-based view also facilitates firms to forecast that the firm's specific resources can give rise to superior performance and competitive advantage.

Generally, small companies face additional hardships while entering the international market as they have additional hindrances to access export resources and capabilities than large firms. The resource-based view proposes that inter-organisational relationships may pave the way for small companies to access the resources, whether tangible or intangible, which will ultimately positively influence the organisation's performance. Hence, it is crystal clear that it would be advantageous for the small companies to acquire the righteous network to acquire needful resources for their better performance.

2.7. Stakeholder Theory

The Stakeholder Theory centralises the firm's relationship with its critical stakeholders for it is smooth working and successful. Further data and research support the vision that the satisfaction of employees, customers, suppliers and partners is the foundation for an organisation's success.

In a broad view, the above-explained theories denote the significance of establishing and improving a firm's stability conditions by way of a firm's amplified competitiveness.

3. SURVIVAL STRATEGIES

When the local companies search for better opportunities, internationalise themselves, and compete outside their own country, they have to face many hardships and challenges. In order to maintain competitiveness in the international market, these challenges and pressures require the local companies to cut their costs so that the customers do not set their minds against the product for being too costly. This situation requires the companies to locate their production houses and facilities in such areas where the production costs are lower.

Along with such ongoing pressures, the managers of local companies must make an effort to respond quickly to the local pressures in order to adapt their product to the local needs and specifications, where the company is active. This necessitates the local companies to differentiate their offers and strategies, distribution channels, human resources, practices and regulation for different countries to match the needs of differentiated customers.

The survival strategies acknowledged in this review suggests the significance of setting up the indispensable conditions in order to maintain the small companies led business environment in the economy and polishing up the competitiveness of the local companies in the WTO led trade regime era. The conditions would strengthen the potential to safeguard their markets from various global challenges. The survival strategies of the local companies were majorly centred around the fundamental theories.

Foremost, through Porter's Generic Strategies, the local companies can identify their strategic position to face the challenge of economic and global competition either differentiation, differentiation focus, cost leadership or cost focus. As stated in the past literature, cost leadership is the most efficacious strategy for small companies since the anticipated objectives of the small companies is to increase the profits by reducing the costs and setting the prices in accordance with the industry average. The local companies can elevate their market share by putting in low prices and earning sufficient profit from each sale due to previously pressed costs. It is challenging for the local companies to achieve economies of scale through a volume of orders initially.

Simultaneously, local companies can carry out SWOT Analysis to find out the organisational external and internal environmental factors to face the challenges of MNCs, economic and global competition, and transnational corporations. Furthermore, small companies can also execute PESTEL Analysis to be mindful of their competitors. In order to tackle the challenge of information and technology, small companies can target their strengths, weaknesses, opportunities and threats subsisting in their external and internal environment as it allows identification of the ability of small companies to take on those facilities in their companies and regulate the availability of the information in the marketplace as information and communication technology has become a critical deciding factor for both the small and large firms. Moreover, small companies can utilise their internal resources and inter-organisational relationships to the maximum as these relationships will help them get access to resources, whether tangible or intangible, which will positively impact the company's performance.

4. CONCLUSION

Under Shri Manmohan Singh, the Union Finance Minister, India first began the procedure of Globalisation and Liberalisation in 1991. Initial 5 years of globalisation didn't turn out to be much appreciable. The arrival of cold beverages manufacturers such as Pepsi, Coke and other companies such as KFC, McDonald's, Boomer etc. only controlled the show.

The arrival of giant car manufacturers such as Ford, Hyundai, Daewoo Motors, Honda which were the result of further liberalisation of trade and privatisation, led to the availability

of a diverse range of cars and ultimately a reduction of domestic car prices. World leaders in telecommunication sectors such as Nokia, Aiwa, Ericsson etc. and electronic giants such as IBM furnished a diversified range of quality products at affordable amounts.

The rate level declined owing to cut-throat competition and this served as an advantage for the Indian consumers. Globalisation turned out to be a bonanza for the customers but a grave for the Indian companies. On account of Liberalisation and Globalisation, the Indian market was inundated with affordable foreign products which eventually affected the Indian local companies adversely. It turned out to be as if the small fishes were left in a big sea with other small and big fishes whereby the small fishes compete for the same food and the large ones simply eat the small ones in order to keep themselves going.

REFERENCES

[1] Dawar N. & Frost T. (1999, April 1), *Competing with Giants: Survival Strategies for Local Companies in Emerging Markets*, Harvard Business Review, https://hbr.org/1999/03/competing-with-giants-survival-strategies-for-local-companies-in-emerging-markets, Last Accessed on 04/24/2021

[2] Lum B. W. (2017), *Business Strategies for Small Business Survival*, Walden University College of Management and Technology https://scholarworks.waldenu.edu/cgi/viewcontent.cgi?article=5635&context=dissertations

[3] *The Impact of Globalisation on Economies and Business*, (2019, March 14), GetSmarter Blog, https://www.getsmarter.com/blog/market-trends/the-historical-impact-of-globalisation-on-economies-and-business/, Last Accessed on 04/15/2021

[4] M. (2017, June 12), *Globalization and its impacts on small businesses – it's not all bad*, Unleashed Software, https://www.unleashedsoftware.com/blog/globalization-impacts-small-businesses, Last Accessed 03/29/2021

[5] Piletic P. (2018, June 28), *How Globalization Impacts Small Businesses More Than You Might Think*. CustomerThink, https://customerthink.com/how-globalization-impacts-small-businesses-more-than-you-might-think/, Last Accessed on 03/29/2021

[6] Zedillo, E. (2019). *The Past Decade and the Future of Globalization*. OpenMind. https://www.bbvaopenmind.com/en/articles/the-past-decade-and-the-future-of-globalization/, Last Accessed on 04/05/2021

[7] Guttal S. (2007), Globalisation, *Development in Practice*, *17*(4/5), 523–531, https://www.jstor.org/stable/25548249

[8] Kanter, R. M. (2003, August 1). *Thriving Locally in the Global Economy*. Harvard Business Review. https://hbr.org/2003/08/thriving-locally-in-the-global-economy, Last Accessed on 04/05/2021

[9] Hirasawa, K. (2019). Globalization and small businesses. *Ordnungspolitische Diskurse, OPO*, 1–21. http://hdl.handle.net/10419/200142

[10] V.A. (2017, August 23) *Impact of Globalisation on India*, Civilsdaily, https://www.civilsdaily.com/impact-of-globalisation-on-india/, Last Accessed on 04/24/2021

[11] Dr. Anbu S. (Ed.), (2010), In Proceedings: *Impact of Globalization in Indian Business and Culture*, Business Horizon

[12] Shanmukhappa S. (2016), Globalisation and Its Impact On Indian Economy: Developments And Challenges, *International Journal of Creative Research Thoughts*, *4*(4), 926–934, https://www.ijcrt.org/papers/IJCRT1133184.pdf

[13] Chand S. (2014, January 3), *Essay on Effects of Globalization on Indian Economy*, Your Article Library, https://www.yourarticlelibrary.com/globalization/essay-on-effects-of-globalization-on-indian-economy/22644, Last Accessed on 04/24/2021

View publication stats

The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1934-8835.htm

Sustainable human resource planning for hospitals in tier 2/3 cities: evidence from India

Sanjay Bhattacharya
USAP, Guru Gobind Singh Indraprastha University, New Delhi, India, and
Chandrani Bhattacharya
Department of Administration and Teaching, Sapphire International School, Noida, India

Received 5 July 2020
Revised 13 November 2020
9 April 2021
6 June 2021
Accepted 7 June 2021

Abstract

Purpose – To evolve sustainable practices for human resource planning (HRP) for hospitals being set up in Tier 2/3 cities in India. The purpose of this study is to determine the strategic challenges are faced by hospitals in Tier 2/3 cities. To align HRP/workforce planning to departmental strategic/business planning of the hospital in the defined context. To address the dilemmas of the prospective health-care workforce in Tier 2/3 cities.

Design/methodology/approach – The integrated methodology adopted uses the grounded approach for building theory from practice. Due to the lack of research in the identified context, the study relies on a review of literature related to health-care practices, HRP and general management inputs to evolve the exact interview questionnaire. The subsequent empirical analysis is based on interviews of niche human resource experts in private hospitals.

Findings – The traditional and successful model of private hospitals in the metro cities cannot be transplanted to the Tier 2/3 cities. The thumb rule on a ratio between the doctors, paramedic staff, nurses and other support staff requires to be modified as necessary. Uncertainty in the new context requires the outlook to be basic, uncomplicated, flexible and agile. Attracting and recruitment strategies require adopting an innovative and multipronged approach. Sourcing high-end specialists need not follow the retainership model of employment.

Practical implications – The need to penetrate health-care services to the Tier 2/3 cities within India cannot be disputed. With the increase in disposable income and standard of living in these cities, access to quality and affordable health care is also imperative. The study is a pioneering effort to suggest a practicable and uncomplicated model for hospital setup in the identified context, with a focus on HRP.

Originality/value – This paper offers novel perspectives to the HRP of health-care personnel for a hospital setup in Tier 2/3 cities in India. For the success of private hospitals and their financial viability, this planning is of utmost importance. There is an attempt to fill the knowledge gap in the context which has been rarely explored or ventured. In addition, the policy initiatives suggested are expected to encourage the growth of private hospitals within the Tier 2/3 cities in India.

Keywords Hospitals, Strategy, Economic sustainability, Policy making

Paper type Research paper

1. Introduction

Health care is perceived as a sector of immense possibilities and potential. Rising income levels, greater health awareness, increased precedence of lifestyle diseases and improved access to insurance are expected to be the key contributors to growth. The Indian health-care sector is predicted to reach US$280bn by the year 2020 (IBEF, 2019).

International Journal of Organizational Analysis
© Emerald Publishing Limited
1934-8835
DOI 10.1108/IJOA-07-2020-2289

The private sector has over the years emerged as a major player in India's health-care business, whilst earning both national and international visibility. It takes care of almost 74% of the country's total health-care outlay (IBEF, 2019). Telemedicine too is a fast-emerging trend in India. Major hospital groups (such as Apollo, AIIMS and Narayana Hrudayalaya) have adopted telemedicine services and entered into many public-private partnerships. Further, the presence of world-class hospitals and skilled medical professionals has strengthened India's position as a preferred destination for medical tourism (IBEF, 2019). Consequently, it is mandated that the Indian health-care sector grows at a brisk pace due to its strengthening coverage, services and increasing expenditure by public as well private players.

But, the capability of the Indian health-care system to handle this is questionable, especially from the point of view of the human resources (HRs) skill-set at its disposal. Examination of a few key figures makes this very evident. A World Health Organization (WHO) database has put India into the "critical shortage of health-care providers" category. There is one for every 10,189 people, the WHO recommends a ratio of 1:1,000) or a deficit of 600,000 doctors and the nurse: patient ratio is 1:483, implying a shortage of two million nurses (Table 1).

Further, there is an acute shortage of paramedical and administrative professionals. This may be a smaller problem, as this can still be dealt with within the short term through training and skill development programmes. Generating qualified doctors, however, is not an easy proposition. Drawing from another source, as of 30 September 2017, there were 1,041,395 doctors registered under different Indian state medical councils. That means India needs 497,189 more doctors to meet the WHO standards. The gap is huge and there is a need to increase the number by more than 48% more from the current workforce (CBHI, 2018). This problem is compounded by the fact that most of the skilled Indian medical workforce is actively being sought out by countries in Europe and the Middle East and retained at attractive compensation packages. The recently announced Ayushman Bharat Scheme by the Government of India is expected to encounter severe challenges on account of this. This gap between the existing and required strengths of HRs can potentially spell doom for the nascent growth of this sector in India. The recent challenges due to Covid-19 have put forward fresh questions and challenges to the health-care sector. It underscores the need for penetration of health-care facilities beyond the metro cities and also the need for exponential growth in the per capita bed capacity within the country.

Table 1.
Status of human resources in the health-care sector

Category	Estimated no.	Health workforce registered up to 2017	% of estimated to registered
Allopathic doctors	770,277	1,023,413	75.26
Dental practice	95,959	197,734	48.53
AYUSH system	530,919	771,468	68.82
Physiotherapy, diagnosis and others	86,508	60,000	144.18
Nursing and midwife	1,317,669	1980,536	66.63
Pharmacists	214,744	907,132	23.67
Health associate professionals	811,744	NA	NA
Auxiliary nurse-midwife	NA	811,279	NA
All health workforce	3,827,820	5,751,562	66.55

Source: Based on NSSO Data 2016

One may argue that India's expenditure on health-care information technology (HIT) is considerably low, currently (Ganapathy, 2021). A due increase in IT spend and intensive use of technology, namely, artificial intelligence, IoT, block chain, etc., can considerably improve the provision of improved and patient-centric services. Persons in the medical fraternity, however, counter that administering medical protocols only allows active human interventions by doctors and this solely, is legally tenable. Hence, there is a need for proactive strategies and developing process administration to mitigate the existing gap of skilled medical professionals. Whilst Christensen *et al.* (2000) have called for disruptive innovations in health care, Porter and Lee (2013) have since long called for restructuring of health-care-delivery services, organizations and in the right locations, to deliver what they call "the value agenda".

Another perspective to this scenario is the penetration of health care into rural areas, which is highly skewed currently. The situation is further compounded by the concentration of medical professionals in urban areas, catering to only 30% of India's population. Many patients, especially those living in rural and semi-urban areas, are still receiving services from unqualified practitioners. However, the Indian economic landscape in the small towns is also undergoing a tectonic shift. Higher disposable income, demographic dividend of a younger generation, heightened aspirations and increasing urbanization resulting in a growing number of nuclear families. Interestingly, the focus on hygiene, health, personal grooming and convenience seems to be driving the rapid growth in these towns (Neilson Report, 2012; Figure 1). Over the next few decades, opportunities in the metro areas will become relatively less enticing, whilst the demand revolution will continue in the smaller towns or the Tiers 2 and 3 cities, as the Indian middle class rises. The "Industry" status accorded to the health-care sector by the government recently (Business Today, 2019), is also expected to boost growth in the tier-2 and tier-3 cities, besides helping private hospitals gain access to land for projects and funding.

It is only predictable that the demand for corporate hospitals in the Tier 2/3 cities will increase in due course. However, it cannot be confidently stated that the successful model of corporate hospitals of metro cities should be replicated or transplanted in Tier 2/3 cities.

There are few companies so far that have tested the ground in such contexts. The authors came across a hospital setup, one of the first in Tier 2/3 cities in the country. M/s

Source: Nielsen <http://www.nielsen.com/content/dam/corporate/india/reports/2012/Featured%20Insights_Demand%20Spotting-The%20Rise%20of%20the%20Indian%20Small%20Town.pdf>

Figure 1.
Importance of Tier 2/3 cities

ABC Health-care (name changed) was one of the first movers in setting up a multispecialty hospital at Mathura. M/S ABC Health care and Research Private Limited were incorporated under the provisions of the Companies Act, 1956 with the Registrar of Companies, Delhi, incidentally decided to enter this unexplored market to set up their first hospital. The venture was ultimately characterized by skill acquisition difficulties, huge recurring operation expenses and accumulating financial losses. It was, therefore, felt worthwhile by the authors, to examine the case in detail to investigate the challenges faced by the venture.

All the facts stated above, are pointers towards the need for meticulous human resource planning (HRP) of the much in demand and valuable resources of skilled and specialized medical personnel. Whilst entering new markets, the model of business needs to be innovative, whilst drawing upon the traditional metro city strategies. It is, however, expected that the metro-based strategies cannot be transplanted or replicated in totality to the Tier 2/3 cities. The case of M/s ABC Health-care and Research Pvt. Ltd. further strengthened this line of hypothesis.

Based on the above-cited motivations, the primary objective of this research was to develop recommendations for sustainable practices for HRP in hospitals being set up in Tier 2/3 cities. Considering the gaps and contexts mentioned, the emerging research questions may be as follows:

- What strategic challenges are faced by hospitals in Tier 2/3 cities?
- How to align HRP/workforce planning to departmental strategic/business planning of the hospital in the context?
- What are the ways of managing the dilemmas of the prospective health-care workforce in Tier 2/3 cities in India?

The scope of this research was, therefore, limited to HRP in private hospitals in Tier 2/3 cities.

2. Literature review

2.1 Human resource planning research

The congruence between HR practice and business strategy has been emphasized in studies related to strategic human resource management (HRM) (Harrison and Bazzy, 2017; Miles and Snow, 1984; Baird and Meshoulam, 1988; Wright and McMahan, 1992). Successful implementation of any strategy has been given importance by various researchers time and again (Lengnick-Hall and Lengnick-Hall, 1988; Farnham, 2006; Legge, 2001; Wheelen *et al.*, 2006; Wright *et al.*, 2004). There are also increasing numbers of empirical studies, conducted in various contexts, on the relationship between strategic HRM and organizational performance or competitive advantage (Aswathappa, 2017; Chowhan, 2016; Chang and Huang, 2005; Huselid, 1995; Bjorkman and Fan, 2002; Chan *et al.*, 2004). Essentially, strategic HRM emphasizes developing the firm's capacity in response to the external environment through the appropriate deployment of HRs. The importance of a human capital pool with a specific range of skills that need to be compatible with corporate strategy, to fulfil the strategic goals has been reinforced through several researchers (Lucas, 1978; Hess, 1987). Dainty (2000) postulated that a well-designed workforce can save the organization a lot of money and contribute significantly towards surpassing business targets. On the converse, the penalties for not staffing an organization correctly can be costly and damaging. Understaffing loses the business economies of scale and specialization, orders, customers and profits, whereas overstaffing, on the other hand, is wasteful and expensive (Aina, 2005).

Baird and Meshoulam (1988) opined that the recruitment strategy for the start-ups should be in sync with the organizational stage of development and immediate objectives. However, a single recruitment strategy across industries cannot be generalized for all kinds of start-ups as different start-ups will have different requirements of HRs (Fombrun *et al.*, 1984). The success and failure of a new setup/start-up are strongly dependent on the quality of HRs and financial factors (Nucci, 1999). Bamford *et al.* (1996) suggest that the success of a start-up is dependent on the quality of product/services it offers to the market, which is usually a surrogate of competent HRs available. Organizational capabilities and their allied functional areas such as HRs, find relevance across industries and sectors for the evolution of powerful tools such as best practices and benchmarking (Seeck and Diehl, 2017; Zairi and Leonard, 1996; Zairi, 2019). This is irrespective of the size of the organization (Hornsby and Kuratko, 1990).

The job design, role, assessment, reward and engagement of an employee should be carried out in the context of the key business objectives of the start-ups (Schuler *et al.*, 2001). Backed up by the presence of an internal career ladder, a formal training system, result-oriented performance appraisal, employment security and broadly defined jobs will enhance the effectiveness and productivity of the employees and help the organization to achieve its objectives (Paul *et al.*, 2019; Mullins, 2002).

2.2 Human resource planning in health care

In the context of health care, there have been a few limited studies for HRP and capability (Zairi, 1998; Dreesch, 2005; Ono *et al.*, 2013; Lopes *et al.*, 2015). HRP in health care needs to be challenged with new ideas on thoughts with change in contexts (Murphy *et al.*, 2009). On the other hand, it is planning itself is a very complex and challenging proposition (Maynard, 2006; Shantz *et al.*, 2016). Evidence about the contribution of lean bundles to hospital performance and productivities have been researched from various perspectives (Parkhi, 2019; Gandhi and Sharma, 2018; Portela *et al.*, 2016; Abdallah and Alkhaldi, 2019). HRM is listed as one of them in the study and the means to establish the lean bundles along with its allied features, need further investigation. HR-related aspects in the Indian health-care industry (Swain, 2019; Ramadevi *et al.*, 2016; Kapoor and Goyal, 2013) have been recently researched, highlighting the focus on the current research problem.

2.3 Health-care business models for tiers 2 and 3 cities of India

The idea of corporate model hospitals in Tier 2/3 cities seems to hold merit as has already been discussed earlier. There is a case for taking medical services and treatment to masses, a majority of India, especially Tier 2/3 cities, do not have access to quality health care. The first merit is the rapid response at which treatment can be delivered, when available in the local vicinity (Porter and Lee, 2013). The problem becomes more acute when patients need to access medical attention in emergent situations, whereby carrying the patient over long hours and a few hundred kilometres is extremely detrimental to his or her well-being. The second merit is the chronic nature of treatments. Many patients need to access treatment regularly, as in the case of cancer. It is not possible for them to regularly travel long distances to access treatment, on a weekly or even fortnightly basis. Travelling to another city would involve expenses towards transport and accommodation, apart from the man-hour/earning loss for members of the family, who are required to accompany the patient. This is apart from the expenses towards boarding and lodging requirements. Therefore, it is imperative that they need to have access to world-class health care, that too, in close vicinity.

Kaplan and Porter (2011) have also emphasized the possibility of delivering the services cost-effectively.

The gap in the available services and their quality is pretty evident through the data cited earlier. A quick survey of the local and government hospitals and clinics and nursing homes in the areas around the case subject had been carried out. The standard of services provided and the facilities available to treat various health-related problems revealed a severe lack of even basic infrastructure. This has already been evidenced in non-cosmopolitan areas of many developing countries (Zhang *et al.*, 2017). Besides, even the availability of pharmaceutical support in terms of medical equipment, medicines and implants are in doubt. Recent research has also revealed extreme challenges in the preparedness of hospitals in the Tier 2/3 cities during the Covid pandemic (Singh *et al.*, 2020).

3. Research methodology

The research is based on an integrated approach of studying existing literature, a pilot case study and empirical research, as a part of a modified Delphi process. The literature review included studies on HRP, its importance and the necessity to be aligned with the overall organization strategy and objectives. Besides, a review of literature on start-ups and ventures into virgin or so far unexplored markets and for companies entering a new business was carried out to draw relevant insights.

For theory development, the field of management has always relied on a broad array of research methods. The strongest relevance of theory building is in the investigation of relatively unknown territories and areas not researched earlier (Yoder, 2019). The objective of such research is to gain a fresh perspective without getting biased by existing theories, schools and research (Eisenhardt, 1989). The severe lack of research related to health-care forays in Tier 2/3 cities revealed in the literature review, compelled the authors to rely on the grounded approach for building theory from practice based on an expert panel opinion. There was a clear need to prioritize specialized knowledge derived from practice in such research problems (Bogner and Menz, 2009). The Research Design has been exhibited in Figure 2.

The above methodology necessarily required improved credibility by basing the research on inputs from practitioners with niche HRs expertise in the corporate/private hospital domain. This has been particularly kept in mind during the course of this research. The selected persons for the interview were with a prerequisite of at least 15 years of experience and the Heads of HRs in Corporate/Private Hospital groups in Northern India. The number of experts was significantly constrained to 4 Nos, as they needed to be knowledgeable in the domain of non-metro health-care markets, with some background outreach experience in such areas. The use of this niche knowledge expertise as per convenience has been justified for data collection in research of such nature (Yin, 2018), which was expected to be a very pragmatic approach to solving the research problem. The interactions with these experts included in-depth structured interview questions, combined with a few open-ended questions (Appendix 1). The interviews allowed questions for critical and confirmatory views of knowledge gathered from the literature review also. Such tools of research have already been used in the context of the health-care industry earlier also (Dickinson *et al.*, 2017; Millar and Dickinson, 2016).

An iterative process was adopted for multiple interactions, such that the range of responses decreased and converged towards a consensus on feasible solutions. The responses recorded have been tabulated and presented in Appendix 2, to this article. The entire exercise involved extensive knowledge distillation in the specific context selected.

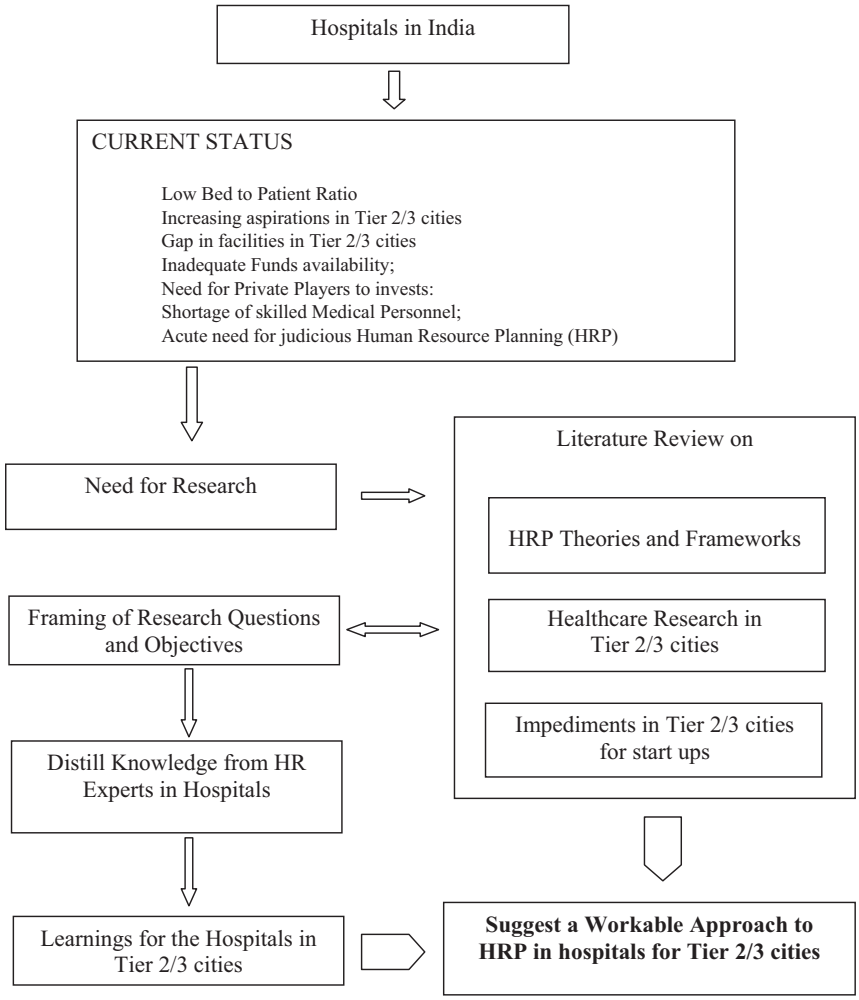


Figure 2.
Research design

4. Collated knowledge from niche experts: practical implications

On basis of the Literature Review, a local survey of the health-care setups and a detailed review of the performance of M/s ABC Health-care at Mathura, an interview questionnaire was developed. HR experts in the hospital domain were interviewed on basis of this development, specifically to seek inputs regarding HRP and the major challenges, whilst setting up a new hospital, in the context of Tier 2/3 cities.

The responses received from four niche experts, who are a part of the top management of leading corporate hospitals have been tabulated and presented in Appendix 2. The responses received and their interpretation towards HRP *vis-a-vis* the research questions have been collated in the following sections.

4.1 Aligning the market scenario to plans for business and operation goals

The HR department has a definitive role in an organization and the HR strategy affects the bottom line. The biggest cost in any organization is human intellect as the prime asset, more so in the case of a service industry. The HR strategies necessarily need to be aligned to the profitability and sustainability of each department and also the organization as a whole. This agrees with the conclusions derived by Harris *et al.* (2007) in research conducted in the UK. Therefore, HR strategies are an integral part of organizational strategy. HR strategies

develop practices, programmes and policies, competencies and employee performance that contribute to organizational objectives. Specifically, HRP for a new company is based on the organization's strategic plan, as deemed suitable to business objectives. This learning has already been highlighted by research carried out by Gowen *et al.* (2006), thereby reinforcing its applicability across contexts.

It was expected that in Tier 2/3 cities the paying power of the population could be limited. The typical clientele of corporate hospitals in Tier 2/3 cities was expected to be cash payment patients. In the current market scenario, clients with medical insurance (personal or group) were also expected to be in minority. It was but obvious, that it would be futile to replicate the corporate hospital model of the metros, for such cities. Manpower planning, which follows typical thumb rules in the metro cities had to be modified according to needs, as per the patient's profile.

Further, HRP in Tier 2/3 cities would also depend on the local availability of resources. HR priorities and department goals, both need to go hand in hand. The hiring is, therefore, to be done on basis of the medical programmes and specialty services. These are arrived at, on basis of the disease profile in the drainage area and the gap in the desired resources. The cost of hiring doctors, technicians can be expected to be higher, as the availability of trained resources is scarce in Tier2/3 cities. Hence, it becomes very critical to carefully prioritize the recruitment on basis of regular and periodic needs. For the HR department, it is a very innovative, but rewarding experience to discover and strike a balance among costs and numbers and to manage a workforce that possesses diverse skill levels.

The specific approach of HRP in new hospitals in Tier 2/3 cities could be influenced by a higher proportion of critical beds, as compared to larger cities. As per standards of manpower planning, the patient versus nurse ratio is done according to bed strength and the medical services provided. Therefore, new hospitals in Tier 2/3 cities could have 5–10% extra manpower at the paramedic and supporting level at all times, to take into account the attrition rate and the lead time in recruitment.

To achieve departmental goals, a clinician would always expect the availability of more than adequate skilled manpower to establish a competitive advantage in the market. On the contrary, HRs have also to align to the organizational goals and the overall financial viability. Recruiting anomalous personnel especially high-value doctors, without adequate numbers of respective department patients, can severely dent the viability and sustainability. HR experts feel that managing the manpower cost holds the key to the financial success of the hospital. Like any management skill, the contribution of HR has to be a focus on the fine balance of an efficient workforce, backed up with high quality, agility and faster delivery and better than the competitors. All this has to be achieved in complete alignment with top-line growth, to bring about competitive advantage. Starting lean has been overemphasized by researchers irrespective of the industry, stands justified here too (Blank, 2013; Gans *et al.*, 2018).

In the case of Tier 2/3 cities, a contingency of a minimum three-year cash burn should be readily available and provided for, as a cash reserve. After 3 years, the promoters should expect the operations to stabilize and probably from the fourth year onwards, the unit should be able to earn the revenues to meet the salary bills. In total, 5 to 6 years is the projected optimum time within which the venture can be expected to turn profitable. This duration is higher as compared to the metro cities. In the long run, it is expected that the profit margins should be assured, once service levels are maintained similar to the metro cities. The HR expenses and salary outgo, therefore, needs to be balanced with the prospective revenues and cash flow. This is a tough balancing act.

4.2 Deployment of local skills and manpower and gaps thereof

As has already been mentioned earlier, the availability of skilled and specialized doctors in Tier 2/3 cities is a challenge. Experienced and established medical practitioners, based in the metro cities are usually hesitant to shift to Tier 2/3 cities.

The strategy for recruiting personnel for a hospital in Tier 2/3 cities, therefore, needs to be multipronged. The mode of recruitment may also be varied. Apart from the traditionally established methods, channels such as the existing employee network, referrals and word of mouth can be used. Social media, digital media, web advertisements, social campaigns, conferences and medical camps can also be tapped to publicize and catch attention in general.

The HR experts suggested additional initiatives as mentioned below:

- All the prospective candidates need to be sensitized to the advantages of living in tier 2/3 cities. Whereby all the facilities, amenities and social infrastructure such as schools, colleges, malls, quality goods outlets, entertainment and leisure, already existing and established in the vicinity need to be highlighted. Better quality of local infrastructure exists consequent to the increased aspirations of the local population, their education levels and disposable income. Conversely, the cost of living is also much lower as compared to the metros. This is because rents, property prices and availability of economical labor/domestic help are much lower. Similarly, the quality of life is expected to be much better with all social amenities, larger houses, self-sufficient campuses and most significantly, negligible commuting. Health and physical well-being advantages are huge with fresh air and much lesser pollution, as compared to the larger cities. Water and electricity issues are also expected to be non-existent.
- Key doctors, branding and reputation are a few differentiating factors for hospitals in the metro cities, which may not be true for tier 2/3 cities. Even if senior clinicians become available on a full-time basis, they would prove to be extremely expensive. Therefore, appointing very senior and expensive doctors needs to be avoided unless mandated by strong logic or due to other compelling circumstances. Recent researchers, Stokes *et al.* (2019) and Malik *et al.* (2017) also hint at adopting ambidexterity as an approach. Hence, it would be worthwhile to attract young talented doctors, who can be hired at lower remuneration and given positions of department heads in the hospital. They can be lured with freedom from rigid hierarchies, better opportunities, the latest medical equipment and state-of-the-art facilities. For example, in Tier 1 city a postgraduate doctor with 3 years of experience would be a clinical associate with no patient admitting rights. Whilst for the same experience level, he/she can be given admitting rights in a Tier 2/3 city. According to the experts, the freedom to function and lead departments is an attractive proposition for a younger generation. These, however, need to be backed by good skills (Trong Luu and Rowley, 2016; Paul *et al.*, 2019). Further, doctors with aspirations other than only money and willing to create a mark for themselves, form an attractive proposition. Small agile groups with an entrepreneurial mindset, who quickly adapt to changing conditions will result in high productivity and morality (Parkhi, 2019; Rigby *et al.*, 2018). Research has emphasized that doctors necessarily have to be a part of any transformation like such (Ali *et al.*, 2018; Dickinson *et al.*, 2016; Lee and Cosgrove, 2014). It will result in better service delivery, better quality and lower risks.

- As a simultaneous measure, it may be worthwhile to recruit a few good local doctors who may have been running established practices locally. Whilst drawing the local patients to the hospitals, they should be counselled on the good practices and better range of facilities and diagnostics available at the hospital to support them. All this has to, however, take place in conjunction with the terms and conditions of fair and ethical practices and towards the best business interests of the hospital is set up. The locally recruited doctors in return, need to necessarily adopt the corporate setup and style of functioning, without any conflict of interest.
- The lower-level staff and paramedics can also be sourced and relocated from the metro easily, as they are usually not high-value personnel. However, their loyalty for the long term and attrition can be highly cost-sensitive. Hence, giving them better facilities and a higher net income saving is expected to lure them in favour of Tier 2/3 cities. An attempt can be made to locate local support staff, nurses and technicians, with reasonable skill levels. However, it would be necessary to train and mould them as per the standard procedures and sustainable operating practices to be followed at the hospital (Cooper *et al.*, 2017; Ghosh *et al.*, 2013). Based on the growth plan, support services like engineering needs to be hired in full strength. House-keeping, security front office, HR marketing can be strengthened phase-wise based on expansion plans.

Differentiating factors in HRP in Tier 2/3 cities are the challenges of sourcing talent like super-specialty doctors. High-end surgeries will always remain a challenge. Beginning with basic medical services, the offerings need to be ramped up to offer super-specialty services at later stages.

4.3 Sustainable long term human resource planning processes

To retain manpower and limit attrition in Tier 2/3 hospitals, there is a need for an alternative and effective approach in continuous HRP and development. These may involve some unconventional methods. Social networks, Tier 2/3 cities targeted recruitment, attitude screening, psychometric texting, alumni groups, conferences, skill academics and online simulation may be a few such methods. The focus, therefore, should be on building rapport with prospective employees and genuinely look after their needs. Education, training, experience and skill-sets are the top-ranking desirable characteristics to be enhanced continuously. For the nurses, supporting staff, administration and management staff, their attitude towards work, experience and team spirit appears to be most important for continuous development.

Better training and skill acquisition and exposure to top medical forums can serve as a good attraction. Other approaches towards sustainable relationships can be instilling a positive culture, sponsoring advanced training, mentoring by renowned doctors, knowledge networking and referrals and innovative employee engagement activities. Emphasis on any kind of career development programme is likely to be popular. There would be a need to create a dynamic work culture that promotes excellence and a people-centric organization.

The additional perks offered to employees in Tier 2/3 cities, could also be an attraction such as tie-ups with good schools, clubs, free housing or reimbursement to employees. For nurses, boarding, lodging, recreation facilities, canteens, transport and food can be a part of the compensation. In some cases, a one-time moving bonus may also serve the purpose of closing employer agreements.

4.4 Values and culture

The management's actions need to be driven by the values such as dedication to every patient's successful treatment, trust and personal responsibilities in all relations, integrity, teamwork and finally, accountability of results. These remain the unique core HR values and culture of the organization that revolve around providing quality care to patients and fulfilling their expectations. Looking inward, it has to focus on a long-term approach by recruiting outstanding talents, grooming unskilled ones and focusing on safety, professional development, teamwork, accountability, sustainability, health and wellness. From the point of view of the employees, it needs to be a rewarding and learning experience.

4.5 Training and development

To keep abreast of the latest developments, trends, enhancement of knowledge, skills up-gradation and exposure to the latest equipment and technology, there is a continuous need to train personnel. These may be related to new standards, benchmarks and advancements in the field of medicine. These may be achieved by encouraging research and participation in conferences regularly. These may need to be conducted locally or budgets allocated to send the personnel to participate in such forums outstation/abroad. Technological advancements require working in close collaboration with major equipment and diagnostic companies. Continuing Medical Education and Continuing Nursing Education programmes and other training related to communication skills, service excellence, stress management, use of world-class high-end medical technology, will ensure that the recruited staff is adequately exposed to contemporary knowledge and skills. Technology can enable conferences and collaborations between specialized doctors sitting at opposite corners of the world. They can discuss critical operations and action plans during the surgeries.

4.6 Performance metrics

To monitor the progress there is a need to establish key metrics. Metrics of progress, performance monitoring and reporting are expected to be similar to that of corporate hospitals of big cities. The acceptable levels of performance metrics may differ. These may be evolved over a whilst, with increasing thresholds. Experts believe that these have much to do with the prevailing organizational culture and professionalism. As typically a third of the staff stationed in these cities can be expected to be regional, the performance evaluation procedures and systems would need to be strongly and religiously inculcated in their mindset.

Many of the HR functions related to its strategy contribute to productivity. Pay, appraisal systems, training, selection and job design are HR activities that directly contribute to productivity. Competitive advantage in a hospital is achieved from diverse sources of customer services, which reflect the core capabilities of an organization. The core capabilities are underpinned by the intellectual capital of the organization in the form of employee competencies. The HR strategy focuses on how these competencies can be further enhanced to achieve the organization's strategic objectives. One expert believes that excellence in services is directly proportional to the talent pool one has. Doubled up with a happy and engaged workforce, the organization becomes a force to reckon with.

5. Policy implications

Whilst the focus of the current study is limited to private hospitals, the legislative authorities need to support their initiatives and business. This is especially given the health and well-being of the public is a social subject. The lack of direct intervention and necessary resources at its disposal makes them morally responsible to put in place policy

recommendations to enhance health care in the remotest locations within the country. Some of the suggested policy initiatives that can be thought of emerging from the current study are as follows:

- A framework can be proposed to subsidize and provide medical treatment for the citizens of the country through private hospitals. Ayushman Bharat proposed recently is a step towards such initiatives.
- There is a need to benchmark compensations of medical personnel at the entry-level, especially in the non-metros. This has to be commensurate to the educational commitment and hard work put in by individuals choosing the profession. This would go a long way to motivate them to serve in smaller cities.
- There is a need to promote and strengthen the network of associated medical/paramedical education and training institutes locally to maintain a sustainable pipeline of skilled manpower. This will also enhance skilled manpower within the country and generate employment.
- Suitable enactments need to be brought forth to legitimize protocols for use of technology in medical treatment. These days technology enables the best medical advice from the remotest locations. Setting up and accessing knowledge management networks, mentoring, a repository of cases, etc., can unfold new paradigms for medical treatment.
- The physical safety of doctors is again a burning issue and this can result in alarming consequences in the non-metro areas. Strong legislation towards the protection of medical personnel can go a long way to reassure them and encourage them to take up assignments of such nature.

6. Summary and recommendations

Like all other management functions and their components, the HRP of an organization needs sustainable practices in a particular context, to be aligned with the organizational goals. The overall objective like any other business remains the overall viability of the business venture and its long-term sustainability.

The idea of moving treatment/health care to the people undoubtedly holds merit according to the experts in the hospital business. However, challenges are many. These may revolve around a longer gestation period in tier 2/3 cities, a larger breakeven period, unavailability of skilled medical personnel locally, lack of support services to the hospital, etc. All the mentioned issues require a huge amount of financial discipline and planning in all aspects, including the vital cog in the entire business wheel of corporate hospitals. Innovative, flexible and agile thinking is required in HRP unlike the already established models of corporate hospitals in metros. The recent challenges due to Covid-19 have underscored the need for penetration of health-care facilities beyond the metro cities and also the need for exponential growth in the per capita bed capacity within the country.

Some of the major recommendations for *sustainable practices* made by the HR experts for private hospitals in Tier 2/3 cities are as follows:

(1) Manpower planning for a corporate hospital in a metro city follows a basic thumb-rule on a ratio between the doctors, paramedic staff, nurses and other support staff. A setup in tier 2/3 cities may require a varied ratio as per local needs, e.g. typically, it is noticed that the proportion of critical beds and patients are expected to be higher here. This, in turn, affects the staff numbers as per medicare norms.

(2) The uncertainty in the new context should require a cautious approach at the outset whereby, only the basic medical services should be commenced, followed by super-specialty services after a reasonable amount of time, post-market assessment.

(3) The availability of skilled and specialized doctors in Tier2/3 cities is a challenge. This may require a multipronged approach as follows:

- Personnel to be counselled on the financial advantages and enhanced quality of life in the tier 2/3 cities come to smaller towns or by offering better perks and remuneration. The prospective employees have to be sensitized to the attractiveness of moving into the Tier 2/3 cities, as compared to the hectic metro life.
- Attract younger talented doctors from the metros, who can be hired at lower remuneration and given better/responsible/higher position in the hospital. They should be provided with better opportunities, equipment and facilities of all kinds. Junior consultants with aspirations and higher salary brackets come to Tier 2/3 cities. The freedom to function and lead departments for a younger generation is an attractive proposition if backed by good skills.
- Possibilities of bringing onboard local doctors having a successful practice and with a good patient base should be explored. They can be lured with better facilities, advanced medical equipment and backup services, for better patients care.

(4) Arranging for high-end specialists and specialized surgeries can always be a challenge as the number of critical cases for full-time doctors may not be achieved or at least till some amount of time from commissioning. It is, therefore, suggested that such cases may be dealt with on a case-to-case basis by visiting doctors, who can be on panel from metros. Such sharing of resources has also been recommended by recent research (Liu *et al.*, 2018). Needless to say, these aspects need to be planned carefully for financial viability.

All the above suggestions stem out of the need for long-term sustainability, which can arise only out of carefully monitored financial discipline. This can be achieved only if every department set up is duly analysed and evaluated as a profit centre.

Finally, a set of policy legislation recommendations have also been suggested towards the well-being and encouragement of medical personnel, to serve in the smaller cities. The suggested plan of actions also includes initiatives for skill development and maintaining a sustainable pipeline of manpower, whilst generating employment within the country.

7. Limitations of research and concluding remarks

In the current study, the opinion of the private hospital HR heads is assumed to be niche expert opinion in the context selected. It relies heavily on the opinions, perspectives and recollections of respondents, which emerges as a limitation of the research. However, this grounded approach can be carried forward to study several units as cases, to further strengthen theory for the evolution of sustainable practices, so that these can be adopted universally as an HRP knowledge base. Other studies in the context of hospitals can also study the interrelationships of the allied management disciplines for a fully integrated approach to the viability and sustainability of corporate hospitals setup up in the context of tier 2/3 cities.

The study will also contribute towards redesign the value chain in HRP. The findings will be examined in light of policy guidelines and recommendations of the Government of India for supporting policy initiatives to enhance health-care facilities in the country. This will enable the health-care companies in India to take advantage of the potential growth crests to establish a competitive edge. The above combined with the already existing knowledge of the local market and customers can act as a potential springboard to the availability of quality health care in the country.

References

Abdallah, A.B. and Alkhaldi, R.Z. (2019), "Lean bundles in health-care: a scoping review", *Journal of Health Organization and Management*, Vol. 33 No. 4, pp. 488-510, available at: https://doi.org/10.1108/JHOM-09-2018-0263.

Aina, S. (2005), *Managing Human Capital in Nigeria*, Fountain Training, Lagos.

Ali, S.S., Basu, A. and Ware, N. (2018), "Quality measurement of Indian commercial hospitals – using a SERVQUAL framework", *Benchmarking: An International Journal*, Vol. 25 No. 3, pp. 815-837, available at: https://doi.org/10.1108/BIJ-05-2016-0060

Aswathappa, K. (2017), *Human Resource Management-Text and Cases*, Tata Mcgraw Hill Education Pvt Ltd, New Delhi.

Baird, L. and Meshoulam, I. (1988), "Managing two fits of strategic human resource management", *Academy of Management Review*, Vol. 13 No. 1, pp. 116-128. Issue

Bamford, C.E., Dean, T.J. and McDougall, P.P. (1996), "Initial founding conditions and new firm performance: a longitudinal study integrating predictions from multiple perspectives", *Frontiers of Entrepreneurship Research*, pp. 465-479.

Bjorkman, I. and Fan, X. (2002), "Human resource management and the performance of Western firms in China", *International Journal of Human Resource Management*, Vol. 13, pp. 853-864.

Blank, S. (2013), "Why the lean start-up changes everything", *Harvard Business Review*, Vol. 91 No. 5, pp. 65-72.

Bogner, A. and Menz, W. (2009), "The theory-generating expert interview: epistemological interest, forms of knowledge, interaction", in Bogner, A., Littig, B. and Menz, W. (Eds), *Interviewing Experts, Research Methods Series*, Palgrave Macmillan, London, pp. 43-80.

Business Today (2019), "Industry status for health-care sector", available at: www.businesstoday.in/sectors/pharma/industry-status-for-health-care-sector/story/310373.html (accessed 26 February 2019).

CBHI (2018), "National health profile", available at: https://cdn.downtoearth.org.in/pdf/NHP-2018.pdf (accessed 26 July 2019).

Chan, L.L.M., Shaffer, M.A. and Snape, E. (2004), "In search of sustained competitive advantage: the impact of organisational culture, competitive strategy and human resource management practices on firm performance", *The International Journal of Human Resource Management*, Vol. 15 No. 1, pp. 17-35.

Chang, W.J. and Huang, T. (2005), "Relationship between strategic human resource management and firm performance: a contingency perspective", *International Journal of Manpower*, Vol. 25 No. 5, pp. 434-449.

Chowhan, J. (2016), "Unpacking the black box: understanding the relationship between strategy, HRM practices, innovation and organizational performance", *Human Resource Management Journal*, Vol. 26 No. 2, pp. 112-133.

Cooper, C.L., Stokes, P., Liu, Y. and Tarba, S.Y. (2017), "Sustainability and organizational behavior: a micro-foundational perspective", *Journal of Organizational Behavior*, Vol. 38 No. 9, pp. 1297-1301, available at: https://doi.org/10.1002/job.2242

Christensen, C., Bohmer, R. and Kenagy, J. (2000), "Will disruptive innovations cure health-care?", *Harvard Business Review*, Vol. 78 No. 5, pp. 102-112.

Dainty, A. (2000), "Improving employee sourcing within construction organization", *Proceedings of the ARCOM 2000 Conference*, Glasgow.

Dickinson, H., Bismark, M., Phelps, G. and Loh, E. (2016), "Future of medical engagement", *Australian Health Review*, Vol. 40 No. 4, pp. 443-446, available at: https://doi.org/10.1071/AH14204

Dickinson, H., Snelling, I., Ham, C. and Spurgeon, P.C. (2017), "Are we nearly there yet? A study of the english national health service as professional bureaucracies", *Journal of Health Organization and Management*, Vol. 31 No. 4, pp. 430-444, available at: https://doi.org/10.1108/JHOM-01-2017-0023

Dreesch, N., Dolea, C., Poz, M.R.D., Goubarev, A., Adams, O., Aregawi, M., Bergstrom, K., Fogstad, H., Sheratt, D., Linkins, J., Scherpbier, R. and Youssef-Fox, M. (2005), "An approach to estimating human resource requirements to achieve the millennium development goals", *Health Policy and Planning*, Vol. 20 No. 5, pp. 267-276, available at: https://doi.org/10.1093/heapol/czi036

Eisenhardt, K.M. (1989), "Building theories from case study research", *Academy of Management Review*, Vol. 14 No. 4, pp. 532-550.

Farnham, D. (2006), *Examiner's Report*, CIPD Co. U.K.

Fombrum, C., Tichy, N.M. and Devanna, M.A. (1984), *Strategic Human Resource Management*, *John Wiley and Sons*, New York, NY.

Ganapathy, K. (2021), "Health-care IT in India: an optimistic outlook", available at: www.asianhhm.com/information-technology/health-careit-india (accessed 8 April 2021).

Gandhi, A.V. and Sharma, D. (2018), "Technical efficiency of private sector hospitals in India using data envelopment analysis", *Benchmarking: An International Journal*, Vol. 25 No. 9, available at: https://doi.org/10.1108/BIJ-06-2017-0135, pp. 3570-3591.

Gans, J., Scott, E.L. and Stern, S. (2018), "Strategy for start-ups", *Harvard Business Review*, Vol. 96 No. 3, pp. 44-51.

Ghosh, P., Satyawadi, R., Prasad Joshi, J. and Shadman, M. (2013), "Who stays with you? Factors predicting employees' intention to stay", *International Journal of Organizational Analysis*, Vol. 21 No. 3, pp. 288-312, available at: https://doi.org/10.1108/IJOA-Sep-2011-0511

Gowen, C.R., McFadden, K.L. and Tallon, W.J. (2006), "On the centrality of strategic human resource management for health-care quality results and competitive advantage", *Journal of Management Development*, Vol. 25 No. 8, pp. 806-826, available at: https://doi.org/10.1108/02621710610684277

Harris, C., Cortvriend, P. and Hyde, P. (2007), "Human resource management and performance in health-care organisations", *Journal of Health Organization and Management*, Vol. 21 Nos 4/5, pp. 448-459, available at: https://doi.org/10.1108/14777260710778961

Harrison, T. and Bazzy, J.D. (2017), "Aligning organizational culture and strategic human resource management", *Journal of Management Development*, Vol. 36 No. 10, pp. 1260-1269, available at: https://doi.org/10.1108/JMD-12-2016-0335.

Hess, D.W. (1987), "Relevance of small business courses to management needs", *Journal of Small Business Management*, Vol. 25 No. 1, pp. 26-34.

Hornsby, J.S. and Kuratko, D.F. (1990), "Human resource management in small business: critical issues for the 1990's", *Journal of Small Business Management*, Vol. 28 No. 3, pp. 9-18.

Huselid, M.A. (1995), "The impact of human resource management practices on turnover, productivity, and corporate financial performance", *Academy of Management Journal*, Vol. 38 No. 3, pp. 635-672.

IBEF (2019), "Health-care presentation", available at: www.ibef.org/download/Health-care-August-2019.pdf (accessed 8 April 2021).

Kaplan, R.S. and Porter, M.E. (2011), "How to solve the cost crisis in health-care", *Harvard Business Review*, Vol. 89 No. 9, pp. 47-64.

Kapoor, A. and Goyal, S. (2013), "Inclusive health-care at base of the pyramid (BoP) in India", *International Journal of Trade and Global Markets*, Vol. 6 No. 1, pp. 22-39.

Lee, T.H. and Cosgrove, T. (2014), "Engaging doctors in the health-care revolution", *Harvard Business Review*, Vol. 92 No. 6, pp. 104-111.

Legge, K. (2001), "Silver bullet or spent round? Assessing the meaning of the high commitment management/performance relationship", in Storey, J. (Ed.), *Human Resource Management: A Critical Text*, 2nd ed.

Lengnick-Hall, C.A. and Lengnick-Hall, M.L. (1988), "Strategic human resource management: a review of the literature and a proposed typology", *Academy of Management Review*, Vol. 13 No. 3, pp. 454-470.

Liu, K., Zhang, R., Yin, L., Sun, L., Pan, X., Hou, L., Shang, X. and Yang, X. (2018), "Medical service unity: an effective approach for medical care in rural areas in China", *Rural and Remote Health*, Vol. 18 No. 3, p. 4483.

Lopes, M.A., Almeida, Á.S. and Almada-Lobo, B. (2015), "Handling health-care workforce planning with care: where do we stand?", *Human Resource for Health*, Vol. 13 No. 38, available at: https://doi.org/10.1186/s12960-015-0028-0

Lucas, R. (1978), "On the size distribution of business firms", *The Bell Journal of Economics*, Vol. 9 No. 2, pp. 508-523.

Malik, A., Boyle, B. and Mitchell, R. (2017), "Contextual ambidexterity and innovation in health-care in India: the role of HRM", *Personnel Review*, Vol. 46 No. 7, pp. 1358-1380, available at: https://doi.org/10.1108/PR-06-2017-0194

Maynard, A. (2006), "Medical workforce planning: some forecasting challenges", *The Australian Economic Review*, Vol. 39 No. 3, pp. 323-329.

Miles, R. and Snow, C. (1984), "Designing strategic human resource systems", *Organizational Dynamics*, Vol. 13 No. 1, pp. 36-52.

Millar, R. and Dickinson, H. (2016), "Planes, straws and oysters: the use of metaphors in health-care reform", *Journal of Health Organization and Management*, Vol. 30 No. 1, pp. 117-132, available at: https://doi.org/10.1108/JHOM-11-2013-0242

Mullins, L. (2002), *Management and Organisation Behaviour*, 6th ed., Financial Times/Prentice Hall, London.

Murphy, G.T., Kephart, G., Lethbridge, L., O'Brien-Pallas, L. and Birch, S. (2009), "Planning for what? Challenging the assumptions of health human resources planning", *Health Policy*, Vol. 92 No. 2-3, pp. 225-233, available at: https://doi.org/10.1016/j.healthpol.2009.04.001

Neilson Report (2012), "The rise of the Indian small town", available at: www.nielsen.com/content/dam/corporate/india/reports/2012/Featured%20Insights_Demand%20Spotting-The%20Rise%20of%20the%20Indian%20Small%20Town.pdf (accessed 26 July 2019).

Nucci, A.R. (1999), "The demography of business closings", *Small Business Economics*, Vol. 12 No. 1, pp. 25-39.

Ono, T., Lafortune, G. and Schoenstein, M. (2013), "Health workforce planning in OECD countries: a review of 26 projection models from 18 countries", OECD Health Working Papers, No. 62, OECD Publishing, Paris, available at: https://doi.org/10.1787/5k44t787zcwb-en

Parkhi, S.S. (2019), "Lean management practices in health-care sector: a literature review", *Benchmarking: An International Journal*, Vol. 26 No. 4, pp. 1275-1289, available at: https://doi.org/10.1108/BIJ-06-2018-0166

Paul, H., Bamel, U., Ashta, A. and Stokes, P. (2019), "Examining an integrative model of resilience, subjective well-being and commitment as predictors of organizational citizenship behaviours", *International Journal of Organizational Analysis*, Vol. 27 No. 5, pp. 1274-1297.

Portela, M.C.A.S., Camanho, A.S., Almeida, D.Q., Lopes, L., Silva, S.N. and Castro, R. (2016), "Benchmarking hospitals through a web based platform", *Benchmarking: An International Journal*, Vol. 23 No. 3, pp. 722-739, available at: https://doi.org/10.1108/BIJ-07-2014-0067

Porter, M.E. and Lee, T.H. (2013), "The strategy that will fix health care", *Harvard Business Review*, Vol. 91 No. 12, pp. 50-67.

Ramadevi, D., Gunasekaran, A., Roy, M., Rai, B.K. and Senthilkumar, S.A. (2016), "Human resource management in a health-care environment: framework and case study", *Industrial and Commercial Training*, Vol. 48 No. 8, pp. 387-393, available at: https://doi.org/10.1108/ICT-03-2016-0014

Rigby, D.K., Sutherland, J. and Noble, A. (2018), "Agile at scale", *Harvard Business Review*, Vol. 96 No. 3, pp. 88-96.

Schuler, R., Jackson, S.E., Jackofsky, E. and Slocum, J.W. (2001),. "Managing human resource in Mexico: a cultural understanding", in Albrecht, M. (Ed.), *International HRM: Managing Diversity in the Workplace*, Blackwell, London, pp. 245-270.

Seeck, H. and Diehl, M.R. (2017), "A literature review on HRM and innovation – taking stock and future directions", *The International Journal of Human Resource Management*, Vol. 28 No. 6, pp. 913-944.

Shantz, A., Alfes, K. and Arevshatian, L. (2016), "HRM in health-care: the role of work engagement", *Personnel Review*, Vol. 45 No. 2, pp. 274-295, available at: https://doi.org/10.1108/PR-09-2014-0203.

Singh, H.K., Patil, V., Chaitanya, G. and Nair, D. (2020), "Preparedness of the cancer hospitals and changes in oncosurgical practices during COVID-19 pandemic in India: a cross-sectional study", *Journal of Surgical Oncology*, Vol. 122 No. 7, available at: https://doi.org/10.1002/jso.26174.

Stokes, P., Smith, S., Wall, T., Moore, N., Rowland, C., Ward, T. and Cronshaw, S. (2019), "Resilience and the (micro-)dynamics of organizational ambidexterity: implications for strategic HRM", *The International Journal of Human Resource Management*, Vol. 30 No. 8, pp. 1287-1322,. Available at: https://doi.org/10.1080/09585192.2018.1474939

Swain, S. (2019), "Do patients really perceive better quality of service in private hospitals than public hospitals in India?", *Benchmarking: An International Journal*, Vol. 26 No. 2, pp. 590-613, available at: https://doi.org/10.1108/BIJ-03-2018-0055.

Trong Luu, T. and Rowley, C. (2016), "The relationship between cultural intelligence and i-deals: trust as a mediator and HR localization as a moderator", *International Journal of Organizational Analysis*, Vol. 24 No. 5, pp. 908-931.

Wheelen, T.L., Hunger, J.D. and Rangarajan, K. (2006), *Concepts in Strategic Management and Business Policy*, Pearson Education.

Wright, P.M. and McMahan, G.C. (1992), "Theoretical perspectives for strategic human resource management", *Journal of Management*, Vol. 18 No. 2, pp. 295-320.

Wright, P.M., Snell, S.A. and Jacobson, (2004), "Current approaches to human resource strategies inside-out versus Outside", *Human Resource Planning*, Vol. 27 No. 4, pp. 2-22.

Yin, R.K. (2018), *Case Study Research: design and Methods*, 6th ed., Sage, Los Angeles, CA.

Yoder, M.E. (2019), "Better together: complementarity between theory and practice in strategic management education", *Journal of Education for Business*, Vol. 94 No. 5, pp. 324-332.

Zairi, M. (1998), "Building human resources capability in health-care: a global analysis of best practice - part I", *Health Manpower Management*, Vol. 24 No. 3, pp. 88-99, available at: https://doi.org/10.1108/09552069810207033.

Zairi, M. (Ed.) (2019), "The question of organisational capability", *Building Organisational Capability*, European Centre of Best Practice Management, Keighley, pp. 1-30.

Zairi, M. and Leonard, P. (1996), *Practical Benchmarking: The Complete Guide*, Chapman and Hall, Springer.

Zhang, T., Xu, Y., Ren, J., Sun, L. and Liu, C. (2017), "Inequality in the distribution of health resources and health services in China: hospitals versus primary care institutions", *International Journal of Equity Health*, Vol. 16 No. 1, p. 42, available at: https://doi.org/10.1186/s12939-017-0543-9

Further reading

Argyris, C. (2003), "Actionable knowledge", in Tsoukas, H. and Knudsen, C. (Eds), *The Oxford Handbook of Organization Theory*, Oxford University Press, Oxford, 423-452.

Bartunek, J.M., Rynes, S.L. and Ireland, R.D. (2006), "What makes management research interesting and why does it matter?", *Academy of Management Journal*, Vol. 49 No. 1, pp. 9-15. Issue

Chandler, A.D. (1962), *Strategy and Structure*, MIT Press, Cambridge, MA.

Chandler, G.N. and McEnvoy, G.M. (2000), "Human resource management, TQM and firm performance in small and medium-size enterprises", *Entrepreneurship Theory and Practice*, Vol. 25 No. 1, pp. 43-58.

Deloitte (2015), "2015 Health-care outlook India", available at: www2.deloitte.com/content/dam/Deloitte/global/Documents/Life-Sciences-Health-Care/gx-lshc-2015-health-care-outlook-india.pdf (accessed 26 July 2019).

Dutton, J.E. and Dukerich, J.M. (1991), "Keeping an eye on the mirror: the role of image and identity in organizational adaptation", *Academy of Management Journal*, Vol. 34 No. 3, pp. 517-554.

Dyer, L. (1985), pp. 488-511. "Strategic Human Resource Management and Planning", in Rowland, K. and Ferris, G. (Eds), *Research in Personnel and Human Resource Management, Vol. 31*.

Edmondson, A.C., Bohmer, R.M. and Pisano, G.P. (2001), "Disrupted routines: team learning and new technology implementation in hospitals", *Administrative Science Quarterly*, Vol. 46 No. 4, pp. 685-716.

Fortis (2014), "Annual report 2013-14", available at: www.fortishealthcare.com/pdf/Fortis_AR_2013-14.pdf (accessed 26 July 2019).

Hamel, G. and Prahalad, C.K. (1994), *Competing for the Future*, Harvard Business Press, Boston, MA.

Galunic, D.C. and Eisenhardt, K.M. (2001), "Architectural innovation and modular corporate forms", *Academy of Management Journal*, Vol. 44 No. 6, pp. 1229-1249.

Gilbert, C.G. (2005), "Unbundling the structure of inertia: resource versus routine rigidity", *Academy of Management Journal*, Vol. 48 No. 5, pp. 741-763.

Learned, E.P., Christensen, C.R., Andrews, K.R. and Guth, E. (1965), *Business Policy*, Homewood.

Lewin, K. (1946), "Action research and minority problems", *Journal of Social Issues*, Vol. 2 No. 4, pp. 34-46.

Mintzberg, H. and Waters, J.A. (1982), "Tracking strategy in an entrepreneurial firm", *Academy of Management Journal*, Vol. 25 No. 3, pp. 465-499.

Ray, B. (2001), "Human resource planning in health-care", *Health Administrator*, Vol. 9/10 Nos Nos 1/2, pp. 14-16.

Shrivastava, P. and Lim, G.E. (1989), "A profile of doctoral dissertations in strategic management", *Journal of Management Studies*, Vol. 26 No. 5, pp. 531-540.

Snow, C.C. and Thomas, J.B. (1994), "Field research methods in strategic management: contributions to theory building and testing", *Journal of Management Studies*, Vol. 31 No. 4, pp. 458-480.

Sutton, R.I. and Raphaeli, A. (1988), "Untangling the relationship between displayed emotions and organizational sales: the case of convenience stores", *Academy of Management Journal*, Vol. 31 No. 3, pp. 461-487.

Whyte, W.F. (1941), "Corner boys: a study in clique behavior", *American Journal of Sociology*, Vol. 46 No. 5, pp. 647-664.

Appendix 1

EXPERT INTERVIEW FORMAT

PERSONAL EXPERIENCE RELATED

1. Can you give a brief background of your experience in Human Resource Planning of Organisations specifically with respect to the number of years in various industries?
2. Are there any differentiating factors in Human Resource Planning for a new company?
3. Similarly are there any differentiating factors in Human Resource Planning for a company in Tier 2 / 3 cities?
4. Can you elaborate on your experience in the Healthcare industry especially if any in corporate hospitals?
5. Are there any key takeaways you have experienced in your Human Resource Planning career?

HR PLANNING IN HOSPITALS RELATED

6. How is the HR Strategy linked to the overall Strategy of the Organisation?
7. What are the unique core HR values and culture of an organization running hospitals?
8. Does the Hospital have a full fledged HR department or depends on outside organization for some functions?
9. Does Digital Media or Social Media form an important part of the process of reaching out to the prospective candidates? If yes, which are the categories of these personnel; Senior Doctors, Junior Doctors, Nurses, Administrative, Top Management, etc.
10. Do you believe that HR contributes to the competitive advantage in hospitals? If yes how?
11. As a corporate hospital, who are the typical clientele?
12. What is the typical attrition rate for a hospital at various levels of personnel?

NEW HOSPITALS IN TIER 2/ 3 CITIES RELATED

13. How does Human Resource Planning Process differ in case of a New Hospital in Tier 2 / 3 cities in the Indian Context with respect to the following stages
    - Identifying Goals
    - Analysis of Environment
    - Gap Analysis
    - Departmental Goals vs. HR Priorities
    - Metrics of Progress, Performance Monitoring and Reporting
14. What do you feel are the challenges specific to the above context in terms of Availability of manpower, remuneration, relocation, etc or any other significant driver.?
15. For a new Hospital coming up in a Tier 2 or 3 city what do you feel is the specific approach towards
    - Human Resource Planning
    - Searching of prospective candidates

(*continuned*)

- Promotion, attraction and retaining of candidates
- Additional sops offered to employees.

16. Do you feel the need of an alternative and effective approach? What can be that?

17. While recruiting what do you look for in the specific category of staff? Please rank in order ; 1 highest and 6 lowest

	Education	Experience	Skill Set	Aptitude	Attitude	Team Spirit
Sr Doctors						
Jr Doctors						
Management Staff						
Admn Staff						
Nurses and Supporting Medical Staff						

18. What is the level of difficulty experienced in HR Planning for the following categories of staff for a hospital in Tier 2 /3 cities (1: Maximum Difficulty, 3: Least Difficulty)

	Availability	Retaining / Counter Attrition	Replaceability
Sr Doctors			
Jr Doctors			
Management Staff			
Admn Staff			
Nurses and Supporting Medical Staff			

19. Do you feel that the recruited staff in specific context of Tier 2 / 3 cities get limited in terms of exposure, enhancement of knowledge, skills and updation on the latest development and trends?
20. What is your view on the availability of local talent specific to hospitals in Tier 2/3 cities?
21. How high is the reluctance to relocate to Tier 2/ 3 cities in the corporate hospital sector?
22. Do you believe that skills are scalable in the Tier 2 / 3 cities? Which ones and How?
23. Do you feel that there is a need to reinforce the benefits of lower cost of living and subsistence in these cities?

(*continued*)

24. Conversely do you feel the need to improve the basic infrastructure and amenities locally? What can the management do to offset and compensate the same?

25. Does the very basic idea of taking the medical service / treatment to the masses hold any merit?

GENERIC INPUTS ON MANAGEMENT THINKING (opinion only)

26. An idea of order of investment required (number of months of cash burn) by a promoter in terms of the monthly salary bill as the benchmark for sustaining operations? How many months of cash flow should be readily available as standby?

27. What is an ideal time frame within which revenues should be able to finance operations? (i.e. the promotor is not required to pump in more money)

28. What should be optimum time period within which the venture should turn profitable i.e. reach breakeven? Can you provide any inputs quantitatively in terms of EBITA etc.?

29. Finally, do you believe that the HR expenses and salary outgo need to be balanced with respect to prospective revenues? How difficult is the task for a new hospital in Tier 2 / 3 cities?

Thanks!!!

Table A1.
Responses of HR experts in the hospital domain

S. no	Description	HR Expert 1	HR Expert 2	HR Expert 3	HR Expert 4
1.	Personal experience	25 years	20 years	22 years	21 years
2.	Field of experience	Health-care, hotels and education	Health sector, health awareness projects and construction projects	Health-care and Hospital conception	Health sector
3.	Differentiating factors in HRP for a new company	Based on the organization strategy plan as the current HR capacity is assessed as deemed fit by promoters and business objectives	Based on the growth plan, support services like engineering needs to be hired in full strength. Housekeeping, security front office, HR, marketing can be strengthened based on expansion	Key doctors, branding and attraction of talent	HRP for a start-up is depended on the vision, mission, objective, strategy of the new company. Depending on the business objective, a strategy has to be planned to arrange for skilled doctors to giving training to unskilled local talents for making a dynamic environment
4.	Differentiating factors in HRP in Tier 2/3 cities	Getting talent is a challenge, need to create a dynamic work culture that promotes excellence and people-centric organization Give fringe benefits as perks and remuneration	Getting super speciality doctors is a challenge Basic medical support such as general physicians, orthopaedics, pediatrics, OBGY, ENT, opthal, chest physician and physio therapy can start off High-end surgeries will always be a challenge	Local availability, disease profiles in the proposed area, proposed specialities	HRP in Tier 2/3 cities is not an easy task. Perks and remuneration along with fringe benefits are given to attracting experienced doctors from big cities. Looking for local talents and giving them training, followed by creating a dynamic work culture is a challenge in itself
5.	Experience in corporate hospitals	Rewarding experience that taught to strike a balance among cost, numbers and manage a work force that is diversed from unskilled to super-specialist clinicians	Corporate hospitals can transfer their second level doctors, skilled technicians and nurses to the Tier 2/3 cities on higher remuneration/ designation and	Almost 10 plus years of varied experience	Had experience of 11 years. Had the opportunity of managing an unskilled workforce, providing training programmes and also

(*continued*)

Table A1.

S. no	Description	HR Expert 1	HR Expert 2	HR Expert 3	HR Expert 4
			independent rights to doctor		dealt with top class doctors and specialists
6.	Key takeaways	Striking a perfect balance between the symbiotic forces of rationalization and revitalization and to manage and strategize them is a skill	Manpower planning has a thumb rule but according to situations the ratio needs to be changed as per the criticality of the patients' requirement	Operations and functions based on key financials, notable part is the overall viability of each department and the transfer pricing within the department for logging of revenues	Success of any organization depends on its mission, objectives, strategies, values and culture. HRP has to be made and changed depending on the criticality of the situation
7.	Link between HR strategies and organizational strategy	Prime asset of any organization is human intellect. HR department has a more defined role in organization and HR strategy affects the bottom line	The biggest cost in any hospital is manpower cost	As mentioned earlier the HR strategies to be aligned to the profitability and sustainability of each department	HR strategies are an integral part of organizational strategy. HR strategies develop practices, programmes and policies, competencies and employee performance that help achieve organizational objectives
8.	Unique core HR values and culture of hospitals	Long term career oriented approach-recruiting outstanding talents, focus on safety, professional development, diversity and inclusion	Patient-centric/ accountability/integrity/ excellence/respect/ teamwork/sustainability/ health and wellness	Has to revolve around providing quality care to patients and patient satisfaction. Conversely for the employees, it has to be a rewarding and learning experience	Our actions are driven by the values: dedication to every patient's success. Trust and personal responsibilities in all relations. Integrity, teamwork and accountability of results
9.	Is there a full-fledged HR department in your hospital?	Yes	Yes	Yes	Yes
10.	Role of digital media	Hospital websites and social networking sites are its first identity.	Hiring of doctors through word of mouth, reputation and coordination with	Usual use of social media for marketing and publicity. Digitization	Internet has enabled us to search for talents cutting across time and distance

(*continued*)

Table A1.

S. no	Description	HR Expert 1	HR Expert 2	HR Expert 3	HR Expert 4
		Informative websites help in recruiting talents. Conferences between specialized doctors sitting at two corners of the world are made possible due to digitization, they discuss critical operations and possibilities of action to be taken during the surgery	house senior doctors, nurses through interviews and personal contacts of other nurses. A centralized call centre of HR specialists answers questions of employees, reducing the need for HR departments at each location	quickly becoming a part of daily operations and patient and doctor scheduling. It can also help establish accountability	constraints. Recruiting through online job postings, applications and candidate screening from the hospital's website
11.	Contribution of HR to the competitive advantage of the organization	Services are directly proportional to the talent pool one has Happy and engaged workforce becomes a force to reckon with	Managing the manpower cost have a positive impact on the profit and loss statement of the hospital	Like any management skill, it has to be quality delivery, agile and faster and better than others. It has to ultimately align with the organization top-line growth	Competitive advantage stems from diverse sources as customer services, skills of the workforce in communication, trust and personal responsibility along with teamwork and dedication to every patient's success
12.	Typical clientele of corporate hospitals	Clients with medical insurance, patients with chronic and acute problems with specialized needs, foreign patients	Patients with medical insurance (personal or group), people who can afford cash treatment at a premium pricing	Patients across categories, more so with the advent of medical insurance	Most of the clients of corporate hospitals are the ones with medical insurances or with acute problems or chronic diseases
13.	Typical attrition rate at various levels of personnel	Highest attrition is among nurses and resident doctors 20–30% Senior doctors/HOD 1–2%	Nurses 35–40% Junior consultants/senior residents/resident doctors 30–35% Technicians/staff 15–18%	Higher in the lower staff which may be 35 to 40%. Doctors movement is more individual and strategic in nature	Attrition rate is highest among the nurses and lower staffs, 30 to 35%. Attrition is lower among the senior doctors, 5 to 10%

(*continued*)

S. no	Description	HR Expert 1	HR Expert 2	HR Expert 3	HR Expert 4
14.	HRP in tier 2/3 cities-identifying goals	Beginning with basic medical services followed by super speciality services after sometime	One needs to plan in advance the manpower and organizational goal so that the strategy, costing and plan can be formulated	Problem is the availability of skilled and specialized manpower in these areas. However, inhibitions in moving to these towns decreasing with improvement in quality of life. They have to be counseled	We can get younger talented Doctors and provide them the opportunities and advanced equipment and gain a name for the institution. It allows a firm to lower its cost of expenditure
15.	Analysis of environment	The services of government hospitals, private hospitals and nursing homes are surveyed Town and villages adjacent are surveyed Kind of clientele in and around the city Availability of pharmaceutical support in terms of medicines and implants	Gathering information about the economy, government, laws and demographic factors such as population size Analysing the competition in terms of existing employment opportunities	The state of local hospitals in the smaller towns is extremely poor. The level of skill available and the equipment is very poor. Adoption of standards is practically non-existent Market analysis helps in estimating the patient demand that the proposed hospital will service	Through surveys of the local and government hospitals and clinics to gather information about the standard of services provided and facilities available to deal with health-related problems and analysing the competition in the market
16.	Gap analysis	After analysis of the environment, points are discussed with the team and then decided on starting different medical services and doctor/ technician/nurses hiring	Gap analysis helps leaders to read on future employment by identifying the competencies their vision requires	The gap undoubtedly exists but so is the penetration of medical insurance, and hence the paying capability of patients	After analysing the gap present in the existing environment, the higher authorities analyse organizations' strengths, weaknesses, opportunities and threats of setting up a hospital in the town
17.	Department goals vs HR priorities	Both goals go hand in hand. The hiring is done based on the medical programme and speciality	HR priorities are to fulfill department goals	HR without doubt has to align to the organizational goals and the overall financial viability. May	To achieve departmental goals one has to give priorities to the availability of skilled

(*continued*)

Table A1.

Table A1.

S. no	Description	HR Expert 1	HR Expert 2	HR Expert 3	HR Expert 4
		services. Hiring of doctors, technicians are at a higher cost as the availability of trained resources is minimal		require a completely new model of treatment where extremely super specialized doctors may not be desirable	manpower for competitive advantage in the market. Training and development programmes become part of the organization working with unskilled workers
18.	Metrics of progress, performance monitoring and reporting	Metrics are the same as what is required of the corporate hospital in Tier 1 cities	These are an integral part of the organization and are much to do with organizational culture and professionalism. As 1/3 manpower are regional or outsourced staff the approach is to slowly drill their mindset	The metrics have to be similar to other management or organization metrics such as revenues, productivity and qualitatively the service levels and patient satisfaction	Metrics of progress, performance monitoring and reporting are similar to that of big cities. However, as the local employees are not used to it, they are trained and their mindset is changed through training and development programmes
19.	Challenges specific to availability of manpower, remuneration and relocation	Challenges are of doctors, nurses and technicians at a higher cost as the availability of trained resources are minimal in these cities	Challenges of making doctor, nurses and technicians leave their private practices and nursing homes to join such institutes	Local challenges include good doctors locally give up their practice to enter corporate setup and style of functioning. Convincing doctors from the metros to move to such towns is another	Availability of talented doctors locally is a big challenge and relocating them from big cities is another challenge. Preparing high standard lower staff from locals is also a challenge
20.	Specific approach of HRP in new hospitals in Tier 2/3 city	New hospitals in Tier 2/3 cities have 5–10% extra manpower at all times, taking into account the attrition rate and the lead time	The hospitals may have less no. of critical beds and the manpower planning(patient vs nurse ratio) is done according to bed strength and the medical services provided	Approach has to be mixed based on the above response. Need to bring on board local doctors with a good patient base and successful practice. Also get young skilled doctors from metros, who may	Along with attracting talented and experienced doctors from metro cities, the experienced and famous local doctors can be attracted by giving better perks and

(*continued*)

S. no	Description	HR Expert 1	HR Expert 2	HR Expert 3	HR Expert 4
				move to these towns for better responsibilities and freedom in decision-making	remuneration along with the better positions
21.	Search of prospective candidates for new hospitals in Tier 2/3 city	Candidates originally from Tier 2/3 cities willing to create a mark for themselves Referrals from existing employees in addition to the website and social networking Tie up with nursing colleges	Searching is a herculean task as seasoned medical practitioners do not want to shift to Tier 2/3 cities Only junior consultants with aspirations and higher salary brackets come to Tier 2/3 cities	Search is usually very difficult and conversion of established doctors from metros. Efforts are required to convince them across all levels for better life quality, better equipment and freedom from the rigid hierarchy	Social networking and digital media play an important role in it. Local educated unskilled manpower is trained according to requirements. Contact is maintained with nursing colleges
22.	Promotion, attraction and retaining of candidates in Tier 2/3 city	Candidates are given higher salary brackets compared to their experience and qualification In Tier 1 city PG doctor with 3 years experience would be clinical associate with no patient admitting rights whereas given admitting rights of patients in Tier 2/3 city	The approach is to use training, mentoring, instilling a positive culture, encourage referrals and recruit from within and use employee engagement activities	The freedom to function and lead departments for a younger generation is attractive if backed by good skills. Better training and skill acquisition and exposure to top medical forums also can serve as a good mix	For retaining talented candidates in Tier 2/3 cities, they are given better perks and remuneration. Emphasis is given on their career development programmes. They are equipped will better equipment and given freedom of decision-making
23.	Additional sops offered to employees	The amount of additional sops offered to employees is higher in Tier 2/3 cities Through additional sops, the candidate is made to realize the benefits,	Tie up with good schools, clubs, free housing or its reimbursement to employees. For nurses, boarding lodging,	Sops will be the quality of life accommodation, club facilities, schools for children, etc. One-time moving bonus may also serve a purpose	Additional sops enable an organization to attract, retain and motivate competent employees. Perquisites are allowed to doctors such as car, club

(*continued*)

Table A1.

Table A1.

S. no	Description	HR Expert 1	HR Expert 2	HR Expert 3	HR Expert 4
		savings and career growth potential.	transport and food are part of CTC		membership and furnished house. For nurses, boarding, transportation facility, canteen, recreation
24.	Need of alternative and effective approach in HRP? What?	Yes, the focus should be on building a rapport with the prospective employee and genuinely look after their needs	Yes, unconventional methods of sourcing manpower can be social networks, Tier 2/3 cities targeted recruitment, psychometric texting, alumni groups, conferences, skill academics and online simulation	HRP methods can be more or less the same. However, the prospective employees have to be sensitized to the attractiveness and advantages of moving into the Tier 2/3 cities as compared to the hectic metro life	Yes, alternative and effective approaches always help an organization to attract qualified and talented manpower. It can be in form of counselling and sensitizing prospective employees for Tier 2/3 cities
25.	Exposure, enhancement of knowledge, skills and updation on the latest development and trend	Some of the basic training conducted in the hospital is 1.Basic life support 2.Advance life support 3.Continuous Medical Enhancement 4.Continuous Nursing Education 5.Employee engagement 6.Clinical protocols adopted by national and international publications and guidances	Super speciality hospital by itself has high-end medical technology as may exist anywhere in the world. CME, CNE and other training such as communication skills, service excellence, stress management, use of world-class high-end medical technology, ensure that the recruitment staff is adequately exposed to the enhancement of knowledge	There is a need for exposure to the latest equipment and technology. Need to train personnel on the latest development, standards and advancements in the field of medicine through the promotion of research and conferences. These need to be conducted locally and budgets need to be allocated to send the person to participate outstation/abroad. For this have to work in close collaboration with major	The staff members are provided with adequate exposure, enhancement of knowledge through on-the-job training programmes such as CME, CNE, service excellence programmes and updating their knowledge and skills through emphasis on the advancement in the field of medicine

(*continued*)

S. no	Description	HR Expert 1	HR Expert 2	HR Expert 3	HR Expert 4
				equipment and diagnostic companies	
26.	Availability of local talent	Limited scope of getting trained staff as they move towards bigger cities and multi-speciality hospitals	The talent is available. Employees are eager and willing to learn. Reputed organization increases their social status suddenly increases and it is a huge motivation in creating right attitude for quick learning	Local talent is limited. Need to pick acceptable cases and train them further	Availability of local talents in Tier 2/3 cities is limited. However, these limited human resources are trained and developed according to the standard of the requirement of the organization
27.	Reluctance to relocate in the corporate hospital sector	Staff is reluctant, so are given promotions, salary hikes or incentives to take responsibilities	There is still reluctance to relocate primarily due to the towns infrastructure issues wrt school, colleges, entertainment and leisure activities, etc	For lower-level staff salary and quality of life is an attraction. For doctors the higher responsibilities, good equipment and quality of life matter	Prospective employees are reluctant to relocate in corporate hospitals from a better facility zone to lower facility zones
28.	Which skills are scalable and how?	Skill sets are scalable with a strong training team in terms of nursing/clinical and general training. The only challenge is with respect to experienced clinicians which remains a challenge as India is simply just not creating as many doctors as the country needs it	The skills are scalable simply due to the fact that people have a learning attitude and one just needs to create a lasting and sustaining training model. Some of the scalable skills are nursing skills, skills of nurse aid through a need-based in service education	Skill sets are all scalable. Starts with an individual who, in turn, teaches and trains a larger group. This process has to be continuous. Attitude to learn, trust and delegate is very critical to this	Nursing skills are scalable. Soft skills can be honed by training both in classrooms and on the job with people with an attitude to learn. Senior doctors can also train junior doctors to improve their skills
29.	Need to reinforce the benefits of lower cost of living and subsistence	The cost of living in Tier 2/3 cities may be less but additional benefits have to	Yes, it is indeed a strong point and an enabler for personnel to look forward to relocating	Yes, this indeed needs to be highlighted and driven home	Reinforcement enables prospective employees to look at the benefits of lower cost of living as an

(*continued*)

Table A1.

Table A1.

S. no	Description	HR Expert 1	HR Expert 2	HR Expert 3	HR Expert 4
		be given as an incentive to make additional savings			incentive to relocating to Tier 2/3 cities
30.	Steps were taken by management to offset and compensate the need to improve the basic infrastructure and amenities locally	According to the available resource, the company can provide accommodation or hostel facilities to admin staff as well technicians	The hospital can cohort a category of staff to provide the same residence facilities are given to clinicians, nurses, admin staff in one locality. We have tied up with schools/creches	Maybe better facilities, better care for families rather than individuals	Better facilities of residence, transport, medical, canteen, schools and creche for children are provided to the members of the organization
31.	Merit of taking the medical services to the masses	It reduces a load of patients in the hospitals – government and private in the big cities A lot of patients need to access treatment on a regular basis as in the case of treatment of cancer. It is not possible for them to travel long distances to access treatment on a weekly or a fortnightly basis	Majority of India, especially Tier 2/3 cities do not have access to quality health-care Rapidity of treatment – the problem becomes more acute when they have to access medical attention immediately and they cannot travel a long distance to access quality health-care	The idea holds merit definitely but striking the sweet spot for the right mix of the facility is very critical and maybe yet to be cracked	There is merit in taking the medical services to the masses as the serious problems that need immediate medical attention can be taken care of. Patients with chronic diseases can get treatment immediately without waste of time
32.	Number of months of cash burn required	Not possible to predict but the EBIDTA positive is possible in 4 to 5 years' time if the set of doctors and staff are stable	By conservative estimates, 24 to 30 months of cash flow should be readily available as a standby cash reserve	Minimum 3 years according to me	Approximately 35 to 40 months of cash burn is required
33.	Ideal time frame revenue able to finance operations	In hindsight, it may take one to two years more than 5 years	After 3 years, the promoters should expect the operations to stabilize and probably from the	Should attempt to achieve this within the second year if not earlier	Ideal time frame is around three years when the revenue will be able to finance operations

(*continued*)

S. no	Description	HR Expert 1	HR Expert 2	HR Expert 3	HR Expert 4
			fourth year, the unit should be able to earn the revenues to meet the salary bills		
34.	Optimum time period to reach the breakeven point	Maybe 7 to 8 years	In total, 5 to 6 years is the optimum time period within which the venture should turn profitable	Should be anywhere between 6 and 8 years in the case of Tier 2/3	It may take around 6 to 7 years to reach the breakeven point
35.	HR expenses vs prospective revenue	It is a herculean task for a stand-alone hospital. A corporate hospital can pull in resources and can spread over its expenses to other units	The HR expenses and salary outgo need to be balanced with respect to prospective revenues. At a Tier 2/3 town, initially wait for the hospital operations to stabilize before activities can be undertaken	This is a tough balancing act. Like any business proposition, this has to balance out for long-term sustainability. In Tier 2/3 the breakeven may be delayed but in the long run, the revenues should be assured once service levels are maintained	An organization has to balance out HR expenses and prospective revenue for the gradual growth of the organization in the future

Table A1.

About the authors

Dr Sanjay Bhattacharya is a practitioner, researcher and instructor specialising in design thinking, strategy and competitiveness. He is a PhD from the Department of Management Studies at IIT Delhi. He holds a Master Degree in construction Management as well as Bachelors in Architecture both from reputed IITs. He has been a faculty teaching Post Graduate Management courses at Delhi Technological University, RICS School of Built Environment and Jaipuria Institute of Management. Further, he has a rich Industry experience of over two decades, involving building projects design and execution in India and Middle East countries. Sanjay Bhattacharya is the corresponding author and can be contacted at: sanjbhat66@yahoo.co.in

Ms Chandrani Bhattacharya is educator actively involved with education of school children. She is an accomplished human resource research enthusiast holding post graduate degrees in Business Management, English and Education. She has over 15 years of public school training, teaching and administrative experience across various cities in India. A highly disciplined and dedicated professional, she is led by her love for her pupils and their holistic growth in being worthy citizens of the next generation.

TCAD Analysis and Simulation of Double Metal Negative Capacitance FET (DM NCFET)

Yash Pathak
Department of Applied Physics
Delhi Technological University
Delhi, India
yash_2k19phdap505@dtu.ac.in

Bansi Dhar Malhotra
Department of Biotechnology
Delhi Technological University
Delhi, India
bansi.malhotra@gmail.com

Rishu Chaujar
Department of Applied Physics
Delhi Technological University
Delhi, India
chaujar.rishu@dtu.ac.in

2021 Devices for Integrated Circuit (DevIC) | 978-1-7281-9955-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/DevIC50843.2021.9455922

Abstract – **In this paper, we investigated the double metal negative capacitance field-effect-transistor (DM NCFET) for improved digital and analog performance in comparison to the conventional negative capacitance field-effect-transistor (NCFET) by using a new concept of ferroelectric layer (HfO_2FE). Visual TCAD software is used to design DM NCFET. The results have been analyzed in form of leakage current (I_{off}), transconductance (G_m), transconductance generation factor (TGF), Early Voltage (V_{EA}), intrinsic gain (A_v), unity gain cutoff frequency (f_T) and improved performance of output conductance (G_d), drain barrier induced lowering (DIBL), switching ratio I_{on} / I_{off}, subthreshold swing (SS). Thus, DM NCFET can be suitable for digital and analog circuit applications.**

Keyword – **Visual TCAD, DM NCFET, SS.**

I. Introduction

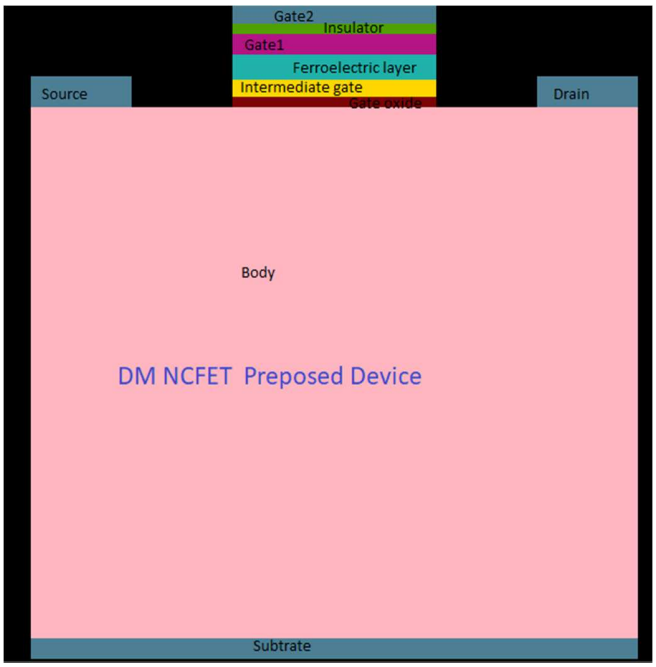
As the size of the device decreases gradually, the performance of the device gets affected and unwanted effects such as a subthreshold swing (SS), drain barrier induced lowering (DIBL) is introduced as a short channel effects (SCEs) [1] [2]. A rule of thumb that has come to dominate computing, Moore's law states that the number of transistors on a microprocessor chip will double every two years or so — which indicating that the chip performance will also perhaps double. The exponential improvement that the law describes transformed the first crude home computers of the 1970s into the sophisticated machines of the 1980s and 1990s, resulting in the rise of the high-speed Internet, smartphones and the wired-up cars, refrigerators and thermostats that are becoming prevalent today [3]. The subthreshold slope of MOSFET determines the ability to turn an on and off the device. At room temperature the ideal value is 60 mV/decade, which is determined by diffusion of carriers from the source to the channel [4]. It is generally accepted that the ongoing scaling of field effect transistors will be eventually limited by the inability to remove the heat generated in the switching process, making it very important to find ways to reduce the power dissipated per switching event. It is also clear that the power dissipation would be lowered significantly if FETs could be operated at lower voltage [5]. Landau field effect transistors promise to lower the power-dissipation of integrated circuits (ICs) by reducing the subthreshold swing (SS) below the Boltzmann limit of 60 mV/dec. The key idea is to replace the classical gate insulator with dielectrics that exhibit negative capacitance FET (NCFET) [6]. When the applied voltage is high, the charge goes down and hence, the name negative capacitance and vice -versa. It is observed from simulated result that DM NCFET have vital advantages such as decrease in leakage current, improved subthreshold swing (SS) high switching ratio, reduction in drain induced barrier lowering (DIBL) etc. with comparison to NCFET. Recently, ferroelectric FETs (Fe-FETs) have attracted enhanced attention due to their promise in both memory and switching applications [7].

II. Device Structure and Methods Simulation

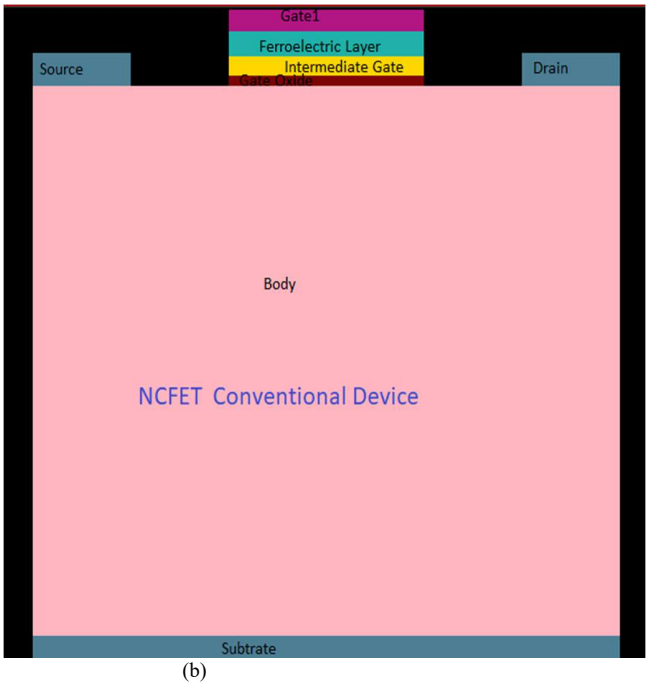
Fig. 1(a) shows the structure of a proposed double metal NCFET and Fig. 1 (b) shows the conventional device NCFET. During the entire process, the structure space and size are the same in both DM NCFET and NCFET except for the two layers. The materials used in both the devices are such that the body is made up of Si with thickness 50nm, Source /Drain are made up of Al with thickness 3nm, substrate is made up of Al with thickness 2nm, Gate oxide is made up of SiO_2 with thickness 21 nm, Intermediate gate is made up of N-poly Si with thickness 2 nm, a new ferroelectric Layer is made up of HfO_2FE with thickness 2 nm, Gate1 is made up of metal. These materials are same in both devices. The difference of proposed device from a conventional device are of two layers such that there is an insulator layer with thickness 1nm and the gate2 is made up of metal with thickness 2nm. A proposed device DM NCFET consists of metal, insulator, metal, ferroelectric, metal, insulator, oxide layer and semiconductor. All the simulations are carried out by using Visual TCAD Software. Table 2 shows the parameter, Mesh size and their value in DM NCFET. In Table 1, the doping profile and concentration of DM NCFET and NCFET device are shown. Doping of Body in DM NCFET and NCFET device is uniform with concentration 1e+16 cm^{-3}.

978-1-7281-9955-9/21/$31.00 ©2021 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:29:33 UTC from IEEE Xplore. Restrictions apply.



(a)



(b)

Fig. 1 Schematic structure of (a) Double metal NCFET Preposed device (b) NCFET conventional device.

Table 1. The doping profile used in designed DM NCFET

Name	Profile	Type	Peak conc./ cm^{-3}	Chara. /um
Body	uniform	acceptor	1e+16	-
vt	Gaussian	acceptor	1e+18	0.001
Ntype1/Ntype2	gaussian	donor	1e+20	0.0005

Table 2. Parameter and their value of DM NCFET

Parameter	Length /Thickness	Mesh size
Source /Drain	3nm	Mesh size =.01µm
Body	50nm	Mesh size =.05µm
Substrate	2nm	Mesh size =.01µm
Gate oxide	1nm	Mesh size =.0003µm
Intermediate gate	2nm	Mesh size =.001µm
Ferroelectric Layer	2nm	Mesh size =.0008µm
Gate1/Gate2	2nm	Mesh size =.01µm
Insulator	1nm	Mesh size =.001µm

III. RESULT AND DISCUSSION

Fig. 2 shows that curve of drain current, I_d (A) and Gate Voltage, V_{gs} (V) for Double Metal NCFET (DM NCFET) and NCFET at Drain Voltage V_d =0.5 V. Fig. 2 shows in log scale at Drain Voltage V_d =0.5V. In DM NCFET, the higher is the I_{on} and the lower is the leakage Current I_{off} of proposed device as compared to the conventional device.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:29:33 UTC from IEEE Xplore. Restrictions apply.

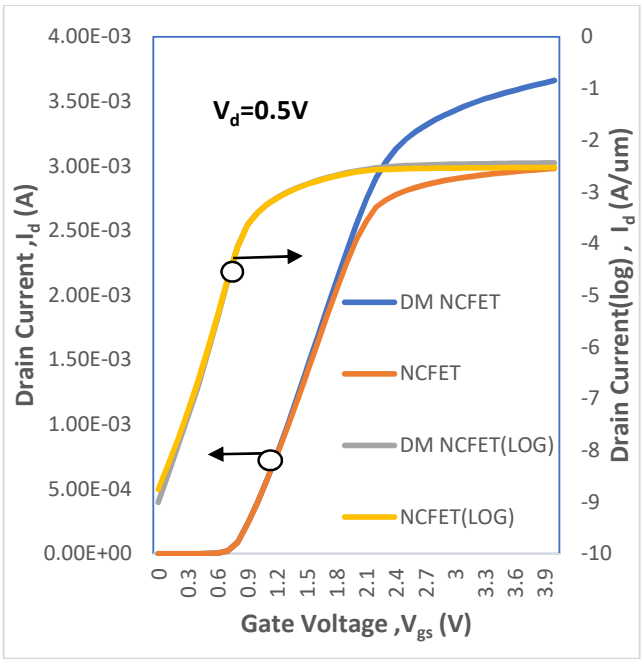


Fig. 2. Drain current and log scale vs gate voltage for DM NCFET and NCFET.

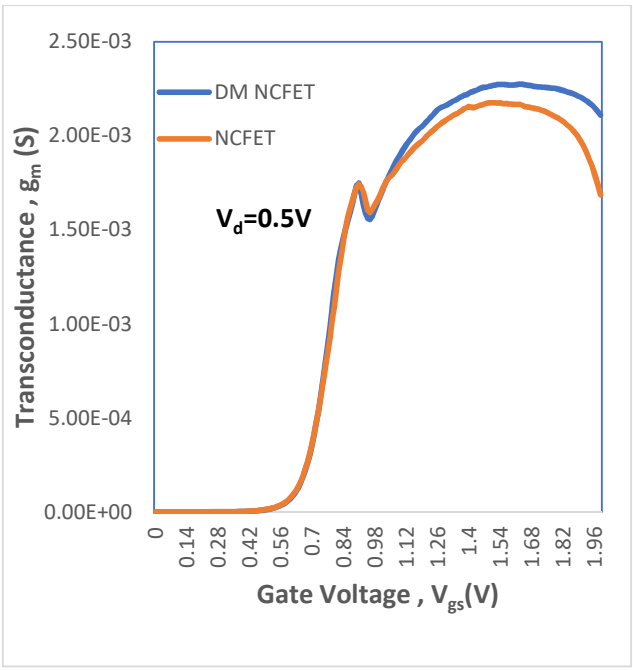


Fig. 3. Curve of Transconductance and gate voltage for DM NCFET and NCFET.

Output conductance decreases due to the shift in the lateral electric field from the gate side to the drain side, which occurs due to the increase in a fringing electric field [10]. Fig. 3. shows the curve of Transconductance and Gate voltage for Double Metal NCFET and NCFET at V_d = 0.5 V. Curve of DM NCFET is higher than NCFET revealing that the transfer of current is higher than NCFET. The transconductance increase of DM NCFET indicates increased in electronic velocity, current density and enhancing electronic mobility.

In Fig. 4 gives the curve of transconductance gain factor (TGF) and Gate voltage for Double Metal NCFET and NCFET at V_d = 0.5V. TGF is the gain generated per unit power loss. TGF perhaps will better work as low power supply. The transconductance is shown in equation 1, transconductance gain factor is shown in equation 2, output conductance shows in equation 3 current ability of device is determined by Output conductance, intrinsic gain shown in equation 4 intrinsic gain is ratio of g_m and G_d, Early voltage shows in equation 5 Early Voltage is high that means reduction in SCEs (short channel effect), Unity gain cutoff frequency is shown in equation 6. These two curves has been calculated using formulae [1] [8].

$$g_m = \frac{\partial Id}{\partial Vgs} \qquad (1)$$

$$TGF = \frac{Gm}{Id} \qquad (2)$$

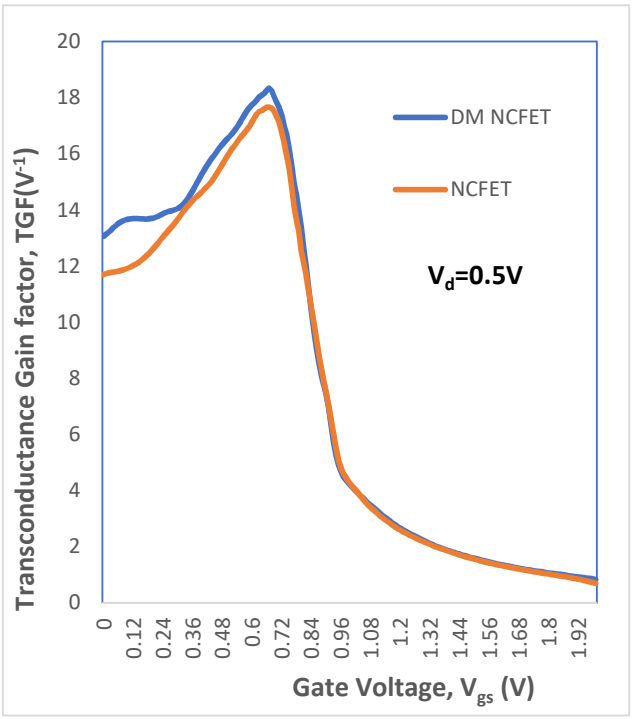


Fig. 4. Curve of TGF and gate voltage for DM NCFET and NCFET.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:29:33 UTC from IEEE Xplore. Restrictions apply.

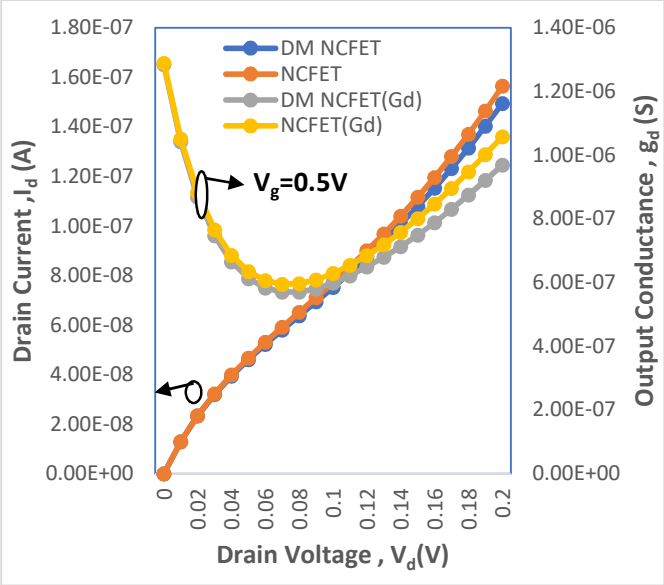


Fig. 5 Curve of Drain current and Output conductance vs Drain voltage for DM NCFET and NCFET

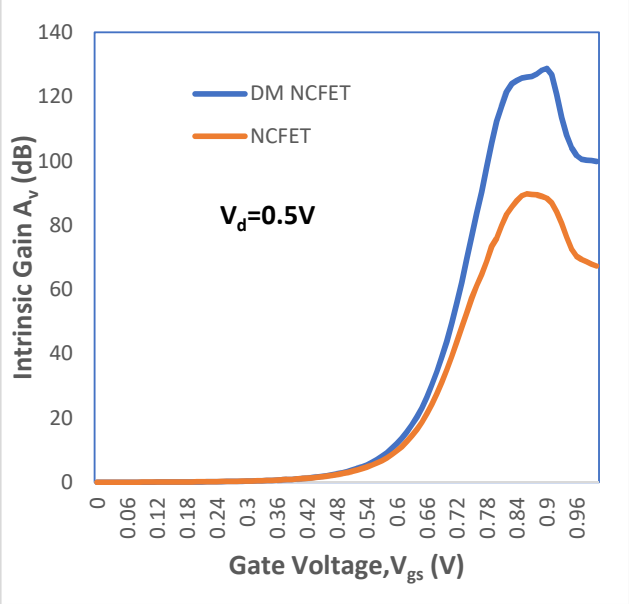


Fig. 6. Curve of Intrinsic gain and gate voltage for NCFET and DM NCFET

Fig. 5. shows the Curve of Drain current and output conductance with Drain voltage for DM NCFET and NCFET at V_{gs}=0.5V. **In** DM NCFET, the Output Conductance (g_d) is lowered than NCFET that reflects a lower Leakage current.

$$g_d = \frac{\partial Id}{\partial Vd} \qquad (3)$$

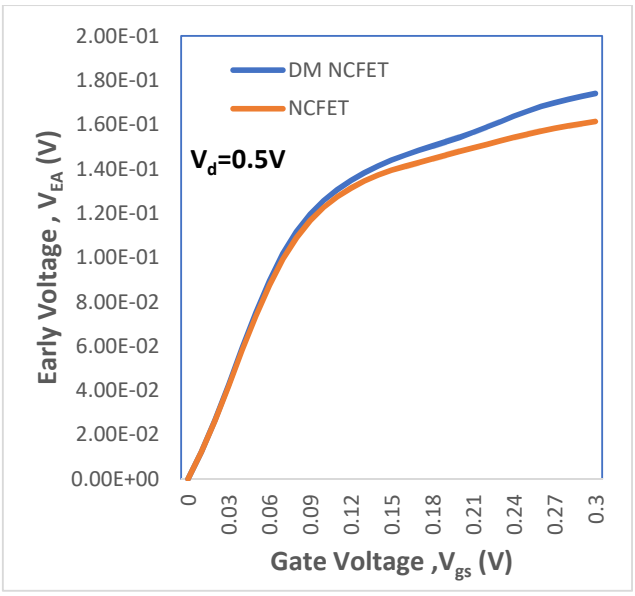


Fig. 7. Curve of Early voltage and gate voltage for DM NCFET and NCFET

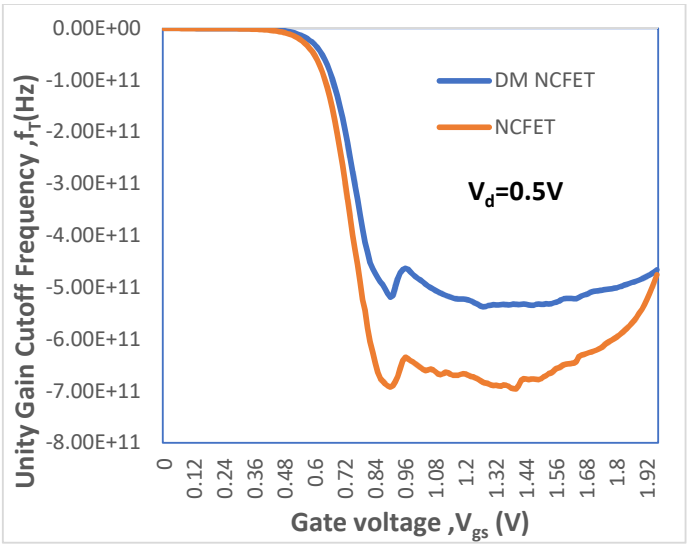


Fig. 8. Curve of Unity gain factor and gate voltage for DM NCFET and NCFET

Fig. 6. shows the curve of intrinsic gain and gate voltage for DM NCFET and NCFET at V_d =0.5V. Intrinsic gain is RF (Radio frequency) performance of device. Intrinsic gain of DM NCFET is higher than NCFET which means a higher conductance gain of device. All the three parameter have been calculated by using the following equations [9].

$$A_v = \frac{g_m}{g_d} \qquad (4)$$

$$V_{EA} = \frac{I_d}{g_d} \qquad (5)$$

$$f_T = \frac{g_m}{2\pi(C_{gs}+C_{ds})} \qquad (6)$$

here, C_{gs} shows capacitance gate to source
C_{gd} shows capacitance gate to drain

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:29:33 UTC from IEEE Xplore. Restrictions apply.

Table 3. Value of Digital and Analog Performance for DM NCFET and NCFET.

S.No.	Device Name	Vth lin(V)	Vth sat(V)	I_{on} (A/um)	I_{off} (A/um)	I_{on}/I_{off}	SS	DIBL
1.	DM-NCFET	0.7105	0.6183	0.002528	$9.28*10^{-10}$	$2.585*10^{6}$	165.954mV/Decade	0.2304
2.	NCFET	0.7108	0.61227	0.002415	$1.76*10^{-9}$	$1.37*10^{6}$	174.869 mV/Decade	0.246

The Fig. 7. variation of early voltage and gate voltage for DM NCFET and NCFET at V_d=0.5V that reflects the RF performance of device. Early voltage is inversely proportional to output conductance and is higher than NCFET that reveals reduction in SCEs. The Fig. 8. exhibits the curve of unity gain cutoff Frequency and Gate voltage for NCFET and DM NCFET at V_d=0.5V, It is inversely proportional to capacitance and also a parameter of RF performance. The curve of unity gain cutoff frequency of DM NCFET is higher than NCFET that reflects lower capacitance. In the Table 3. Represent that reduced subthreshold swing (SS) – control of the channel improved I_{on} /I_{off} ratio and - improved leakage current and decrease in drain induced barrier lowering (DIBL). All the simulation is obtained from Visual TCAD Software. I_{off} current (leakage current) of DM NCFET is lower than that of NCFET.

IV. Conclusion

In this work, the effect of ferroelectric thin layer (HfO_2FE) in negative capacitance Field Effect Transistor (NCFET) is investigated by using Visual TCAD Software. From the simulated result, it has been shown that Double Metal NCFET is improved more than NCFET. The proposed device (DM NCFET) has decreased leakage current (I_{off}) by 47 % than that of the conventional device (NCFET). This device has decreased in sub threshold swing (SS) by 5 % than that of the conventional device (NCFET). the proposed device has also decreased in Drain induced barrier lowering (DIBL) by 6% than that of NCFET. These result clearly reveal that DM NCFET has better and improved the digital and analog parameter than that of NCFET.

References

[1] B. Kumar, A. Kumar and R. Chaujar, "The Effect of Gate Stack and High-κ Spacer on Device," in *IEEE VLSI Device, Circuit and System Conference (VLSI-DCS)*, Kolkata, India, 2020, pp. 159-163.

[2] A. Chabbra, A. Kumar and R. Chaujar, "GaAs Junctionless FinFET using Si3N4 Spacer for High performance Analog Application," in *International Conference On Advances in Communication and Computing Technology (ICACCT)*, 2018, pp. 483-486.

[3] M. M. Waldrop, "The chips are down for Moore's law," *Nature,* vol. 530, pp. 144-147, 2016.

[4] K. E. Moselund, D. Bouvet, V. Pott, C. Meinen, M. Kayal and A. M. Lonescu, "Punch-through impact ionization MOSFET (PIMOS): From device principleto applications," *Solid-State Electronics,* vol. 52(9), pp. 1336-1344, 2008.

[5] S. Salahuddin and S. Datta, "Use of Negative Capacitance to ProvideVoltage Amplification for Low PowerNanoscale Devices," *NANO LETTERS,* vol. 8, pp. 405-410, 2007.

[6] K. Karda, A. Jain, C. Mouli and M. A. Alam, "An anti-ferroelectric gated Landau transistor to achieve sub-60 mV/dec switching at low voltage and high speed," in *Applied Physics Letters 106(16) ,163501*, 2015.

[7] G. A. Salvatore, L. Lattanzio, D. Bouvet, I. Stolichnov, N. Sstter and A. M. Lonescu, "Ferroelectric transistors with improved characteristics at high temperature," in *Applied Physics Letters 97, 053503*, 2010.

[8] S. K. Mohapatra, K. P. Pradhan, P. K. Sahu and M. R. Kumar, "The performance measure of GS-DG MOSFET: An impact of metal gate work function," in *Adv. Nat. Sci. Nanosci. Nanotechnol., vol. 5, no. 2,*, 2014.

[9] B. Kumar and R. Chaujar, "Analog and RF Performance Evaluation of Junctionless Accumulation Mode(JAM) Gate Stack Gate All Around (GS-GAA) FinFET," in *Silicon*, 2021.

[10] A. Nandi, A. K. Saxena and S. Dasgupta, "Impact of dual-k spacer on analog performance of underlap FinFET," *Microelectronics J.,* vol. 43, no. 11, pp. 883-887, 2012.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 07:29:33 UTC from IEEE Xplore. Restrictions apply.

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

The Impact of COVID-19 on the Global Renewable Energy Industry

Sanjay Yadav[1] , Md. Tanweer Alam[1] , Mohammed Sayeed Khan[1]
[1]Department of Mechanical Engineering
Delhi Technological University
New Delhi, India

***Abstract*— The COVID-19 pandemic has created a crisis in the society with wide-ranging impacts not limited to the health and economic sectors. Every aspect of the renewable energy industry has been affected, from the supply chain down to the end consumer. Delays in raw material production, installations and labor shortages are a few of the challenges currently faced by the key stakeholders in the industry. The ability of renewable energy sources to provide clean and cost-effective energy for essential services such as healthcare makes them a vital organ in the fight towards COVID-19. The International Renewable Energy Agency (IRENA) urges governments to leverage the progress achieved with renewable energy to meet the global climate and sustainability objectives in energy security and economic recovery to protect the health and welfare of the members of the society.**

Keywords*— COVID-19; renewable energy; healthcare.

I. INTRODUCTION

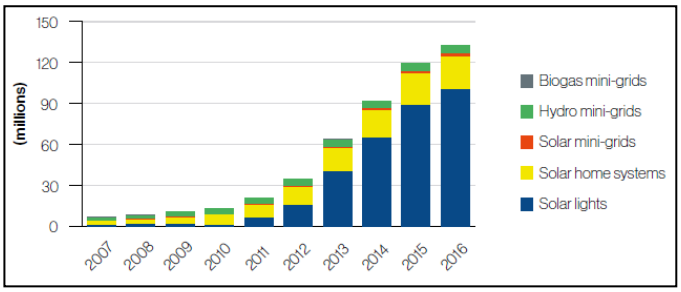
Renewable energy is energy which is obtained from naturally occurring currents of energy or from energy which can be regenerated by naturally occurring energy currents on the time scale of a human life.

Renewable energies cannot generally be transported, but have to be accessed and converted at the location of their occurrence, at typically a much lower density than that of a large power station where the energy potential of the fuels can be concentrated highly. As a result, renewable energy resources typically result in small-scale installations, such as heating and supplementary electrical supply by solar power, or in medium-scale installations, such as wind farms.

Renewable energy is one of the cheapest sources of energy for power generation. Non-existent fuel charges have led to more than two-thirds of the world's population adopting renewable sources to meet their daily energy needs [1]. It also reduces the emission of greenhouse gases and protects the mankind from diseases. A study conducted by Harvard concluded that people living in cities contaminated with air pollution are more likely to die of COVID-19 [2].

Renewable sources of energy have been the most resilient energy source to the COVID-19 lockdown measures. The demand for renewable energy has fallen but the demand for renewable electricity has largely been unaffected. In Quarter 1 (Q1) 2020, global use of renewable energy in all sectors increased by about 1.5% relative to Q1 2019 [3]. Renewable electricity generation increased by almost 3% in the same period due to previously completed solar photovoltaic (PV) and wind project installations. According to Wood Mackenzie estimates, 2020 global solar and energy installations are expected to drop nearly 20% compared to pre-COVID-19 projections [4]. The ongoing pandemic has resulted in 17.8% of the total clean energy workforce of the United States of America (USA) being made redundant [5].

Figure 1: Population served by off-grid renewable energy solutions.



Source: IRENA (2019).

The rapid industrialization and urbanization coupled with population growth in the Asia Pacific region make it the largest market for renewable energy during the forecast period of 2019 – 2021 [6]. According to IRENA, the total solar installed capacity of the region in the year 2019 was 330,786 megawatts (MW) [7]. The adoption of new solar technologies such as floating solar farms and building integrated photovoltaics (BIPV) during the forecast period will lead to solar energy being the largest renewable energy market.

China was the epicenter of the COVID-19 outbreak. The second half of 2020 will see the majority of Chinese renewable energy producers resume their operations to meet the increasing demand and clean energy targets for the forecast period.

II. RESEARCH METHODOLOGY

The authors analyzed the impact of COVID-19 on the renewable energy industry through extensive online research. Multiple sources such as government organization reports and news blogs were utilized to gather data and information on the current and future scenario of the renewable energy industry.

III. RESULTS AND DISCUSSIONS

The renewable energy industry has suffered a major setback in terms of component shortages due to disruptions in global supply chains which have resulted in slowdown of renewable energy projects. A faster recovery from the pandemic would enable projects to be completed at a greater pace. In Australia, a renewables-led economic recovery has been advocated to encourage investment and boost the country's energy transition [8].

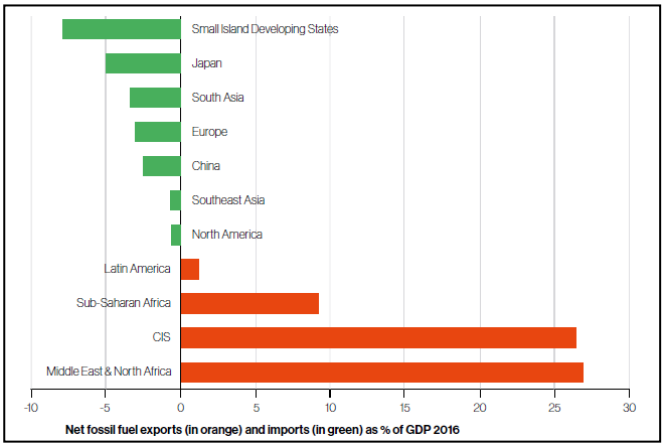
(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

In Q1 2020, the share of renewables in global electricity generation increased to nearly 28% from 26% in Q1 2019 at the expense of the traditional sources of energy, coal and gas [3]. The growth of renewable electricity generation is smaller than anticipated due to the ongoing pandemic.

Before lockdown measures were implemented, shares of variable renewables were similar or higher due to favourable weather conditions, projects completed in 2019 and limited electricity demand growth. After lockdown measures were implemented, a noticeable step up in variable shares of demand was observed due to a decrease in demand for electricity. Belgium, Italy, Germany, Hungary, and USA are a few countries that have recorded high hourly shares of variable renewables during the lockdown [3]. On March 22 2020, Germany imposed strict social distancing measures. During this period, the share of variable renewables has been consistently higher than in the same period in 2019. Overall, electricity systems have been able to accommodate the increasing shares of variable renewables due to prior experiences of increase in solar PV penetration in summer months.

Figure 2: Regional impact of the energy transition.

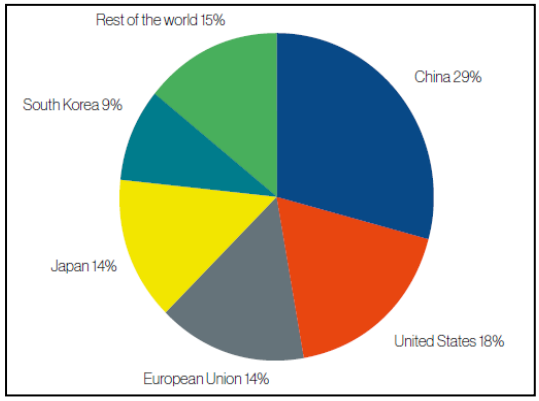


Source: IRENA (2019).

Major suppliers of solar PV and energy storage equipment in China, the Republic of Korea and USA have been affected by COVID-19, resulting in an uncertainty over the capacity growth of distributed solar PV applications in 2020 [9]. In 2019, one-fifth of all renewable capacity deployed consisted of individuals and small-to-medium enterprises installing solar PV panels. Currently, the installation of solar PV has stopped due to lockdown measures implemented to combat the disease.

The impact of COVID-19 on the renewable energy industry could be enormous due to the regions severely affected by pandemic experiencing a reduction in construction activities. The production of renewable electricity depends on the availability of natural resources such as favourable weather conditions. The delay in construction due to supply chain disruption and labor restrictions could decline the growth rate of renewable power capacity addition for the year 2020. The duration and extent of lockdowns, social distancing measures along with the scope and timing of economic stimulus package in response to the dwindling economy of the countries will have a major influence on the functioning of the renewable energy industry.

Figure 3: Cumulative share of renewable energy patents end 2016.

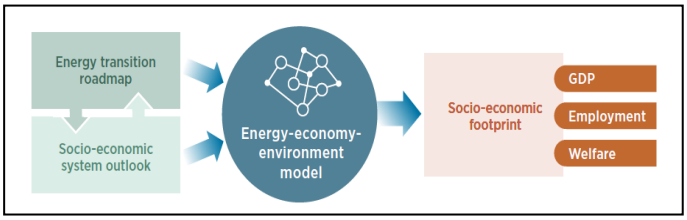


Source: IRENA (2019).

IV. RECOMMENDATIONS

- Revisit deadlines for renewable energy projects that face contractual obligations for near term delivery - To offset the pandemic's impact on the labor force and the global supply chain for renewables, delayed projects should be resumed under the same terms and frameworks [10].
- Affirm and extend policies promoting renewable energy solutions (both centralised and decentralised) - To provide long-term policy certainty in these times of crisis, governments must consider affirming existing and planned support schemes, as well as continuing to implement appropriate market and policy frameworks that support grid development, storage and flexibility, and other infrastructure critical to support a higher penetration of renewable energy [10]. Permitting and siting approvals should be fast-tracked so that the renewable energy industry can plan ahead and protect its workforce.

Figure 4: The energy transition and its socio-economic footprint.



Source: IRENA (2018).

- Prioritise renewable energy in any stimulus measures and commit to phasing out support for fossil fuels - Stimulus packages for any sector should include conditions to reduce emissions and accelerate the transition to a clean, low-carbon economy [10]. Within the energy sector, fossil-fuel subsidies must be steered towards investments in low-carbon infrastructure, as well as towards research and innovation to support higher shares of renewables, particularly in end uses such as transportation, heating and cooling. With higher reliance on renewables, countries and communities can benefit from more stable energy prices and greater energy independence, along with reducing emissions and fulfilling climate objectives.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

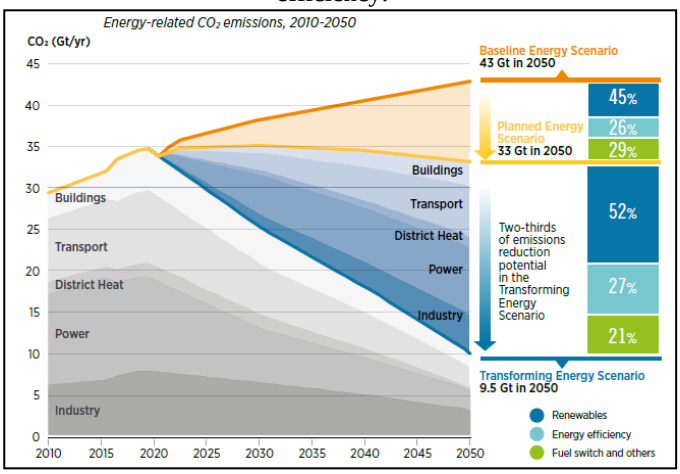
The actions outlined above would support a sustainable economic recovery, make societies more resilient, and help to fulfil the objectives of the Paris Agreement and the 2030 Agenda for Sustainable Development. In response to the crisis, governments are currently in a position to recognise the benefits of renewables, continue building broad public support for the transformative decarbonisation of societies, and pave the way for a clean and low-carbon economy.

V. CONCLUSION

Fossil fuels are still the backbone of the global energy system. Two-thirds of the global emissions are contributed by energy industry through production and consumption [11]. The extent of impact of COVID-19 on the renewable energy industry has not been fully realized. The lockdown measures implemented in various countries to combat COVID-19 have led to a reduction in emissions.

The renewable energy industry is vital for the energy security of the country. It is important that the authorities enact effective policies and promote public-private cooperation for a sustainable energy system. Post COVID-19, countries should focus on renewable energy to meet their energy needs and equip themselves to combat similar situations in the future. The successful energy transition will aid in addressing climate change, combat global warming and promote sustainable development.

Figure 5: The bulk of emission reductions – Renewables and efficiency.



Source: IRENA (2020).

REFERENCES

[1] Layke, J. and Hutchinson, N., (2020). 3 Reasons to Invest in Renewble Energy Now [online]. *World Resources Institute.* Available from: https://www.wri.org/blog/2020/05/coronavirus-renewable-energy-stimulus-packages

[2] Harvard School of Public Health., (2020). Linking Air Pollution to Higher Coronavirus Deaths [online]. *Harvard School of Public Health.* Available from: https://www.hsph.harvard.edu/biostatistics/2020/04/linking-air-pollution-to-higher-coronavirus-death-rates/

[3] IEA., (2020). *Global Energy Review 2020* [online]. Paris: IEA. Available from: https://www.iea.org/reports/global-energy-review-2020

[4] Energy Choice Coalition., (2020). Wood Mackenzie, Industry Associations Report Coronavirus Impact On Energy Transition [online]. *Energy Choice Coalition.* Available from: https://www.energychoicecoalition.org/blog/2020/4/17/wood-mackenzie-industry-associations-report-coronavirus-impact-on-energy-transition

[5] BW Research Partnership., (2020). *Clean Energy Employment Initial Impacts from the COVID-19 Economic Crisis, March 2020* [online]. California: BW Research. Available from: https://e2.org/wp-content/uploads/2020/04/Clean-Energy-Jobs-Initial-COVID-19-Memo-Final.pdf

[6] MarketsandMarkets., (2020). Covid-19 Impact on Renewable Energy Market [online]. MarketsandMarkets. Available from: https://www.marketsandmarkets.com/Market-Reports/covid-19-impact-on-renewable-energy-market-144387621.html

[7] IRENA., (2020). *Renewable Capacity Statistics 2020* [online]. Abu Dhabi: IRENA. Available from: https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2020/Mar/IRENA_RE_Capacity_Statistics_2020.pdf

[8] Filatoff, N., (2020). Time to enable shovel-ready renewable projects to repower Australia's economy [online]. *PV Magazine.* Available from: https://www.pv-magazine-australia.com/2020/05/05/time-to-enable-shovel-ready-renewable-projects-to-repower-australias-economy/

[9] United Nations ESCAP., (2020). COVID-19 Crisis Reinforces The Importance Of The Sustainable Energy Transition [online]. *United Nations ESCAP.* Available from: https://www.unescap.org/blog/covid-19-crisis-reinforces-importance-sustainable-energy-transition#

[10] IRENA., (2020). Call to Action in Response to COVID-19:Renewable Energy is a Key Part of the Solution [online]. *IRENA.* Available from: https://www.irena.org/-/media/Files/IRENA/Coalition-for-Action/Publication/IRENA_Coalition_COVID-19_response.pdf?la=en&hash=BD7A182A347015C49AF624F1200FEE3667EBF396

[11] Loughman, J., (2020). An intentional energy transition after COVID-19 will generate gains for climate and society [online]. *World Economic Forum.* Available from: https://www.weforum.org/agenda/2020/05/renewable-intentional-energy-transition-post-covid19-coronavirus-pandemic/

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

Check for updates

JOURNAL OF THERMAL SCIENCE AND ENGINEERING APPLICATIONS

Thermodynamic and comparative analysis of ERC and ARC integrated wet-ethanol fueled HCCI engine for cogeneration of power and cooling

Mohd Asjad Siddiqui

Department of Mechanical Engineering, Delhi Technological University, Shahbad Daulatpur, Bawana Road, Delhi-110042, India

Email: siddiquimohd.asjad@gmail.com

Abdul Khaliq
Department of Mechanical Engineering, College of Engineering at Yanbu, Taibah University, Yanbu Al Bahr 41911, Saudi Arabia
Email: khaliqsb@gmail.com

Rajesh Kumar[1]
Department of Mechanical Engineering, Delhi Technological University, Shahbad Daulatpur, Bawana Road, Delhi-110042, India

Email: dr.rajeshmits@gmail.com

Abstract

This study attempted for the proposal and analysis of a combined cycle consists of a wet-ethanol fueled and turbocharged HCCI engine coupled to ejector refrigeration cycle (ERC) and absorption refrigeration cycle (ARC) for the simultaneous generation of two distinct outputs namely power and refrigeration. Both first and second laws of thermodynamics were employed to develop a thermodynamic model which has been applied to investigate the performance of combined cycle. Further, performance of the combined cycle for ERC versus ARC was compared and assessed after altering operating parameters (turbocharger pressure ratio, turbocharger compressor efficiency, ambient temperature, and the entrainment ratio of ERC and generator temperature of ARC) to study their effect on engine power output, refrigeration load, exergy of refrigeration, energy and exergy efficiencies of the cooling-power cogeneration cycle. Results show that elevated pressure of turbocharger results in the enhancement of HCCI engine power and increase of the refrigeration of thermal load, simultaneously. However, ambient temperature rising shows the decline of HCCI engine efficiencies and energy efficiency of cogeneration while the cogeneration cycle exergy efficiency is found increasing. Furthermore, the results are reported for the refrigeration performed by LiBr-H_2O operated ARC, and R134a and R290 operated ERC, respectively. Mapping of exergy destruction for the presented cogeneration cycle discovered HCCI engine, boiler of ERC, generator of ARC, and catalytic convertor as the components of significant exergy destruction. Entrainment ratio and type of refrigerant employed in ERC and the generator temperature of ARC shows a marginal impact on the COPs of these cycles.

[1]Corresponding author Email: dr.rajeshmits@gmail.com (Rajesh Kumar)

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

***Keywords**: HCCI engine, energy efficiency, exergy efficiency, Absorption refrigeration cycle, ejector refrigeration cycle,*

1. INTRODUCTION

Establishment of an environmentally benign sustainable society demands the improvement in thermal efficiency of land transportation vehicles. On average two-thirds of the primary energy supplied to internal combustion engines is lost via heat rejection to atmosphere [1-3]. This significant loss of energy to environment asking for the recovery of this rejected heat that would decrease the fuel consumption and reduce the environmental degradation caused because of the fossil fuels consumption. In view of conserving the fossil fuels and to protect the living environment, many investigators attempted the problems related to the design and development of renewable fueled engines [4, 5]. Among the renewable fuels, ethanol has been found as one of the promising options as it has higher flame speed, higher octane number, and it is oxygenated fuel due to which it provides high rate of combustion efficiency and also reduced level of the pollutant emissions and greenhouse gases. Therefore, ethanol fueled engines can be considered as an alternative to meet the aim of achieving higher engine efficiency at lower emissions [6]. Ethanol can be produced from various agricultural products like corn and sugarcane, and studies carried out in regard of the production of ethanol from corn through energy balance reports that significant amount of energy is required to produce ethanol for water removal processes (dehydration and distillation) which consumes 37% of the energy which ethanol contains [7]. Therefore, efforts to utilize wet-ethanol directly as fuel in combustion engines was considered as an integral part of research and from some of the investigations it is revealed that this may lead to considerable increase in net gain of energy from 21% to 55% [8, 9]. Utilization of wet ethanol had its challenges in traditional modes of combustion (SI and CI) and hence requires a unique combustion process

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

which combine the characteristics of both SI (spark ignited) and CI (compression ignited) engines [10]. In this regard, homogeneous charge compression ignition (HCCI) engines have been discovered as an option to burn low grade fuels for land transportation purpose because these engines have the distinction of combining the advantages of both SI and CI combustion processes. HCCI engines utilizes premixed and lean fuel-air mixture at high intake temperature and the mixture is thermally auto-ignited following the higher compression ratio. This results in the drastic reduction of the most harmful emissions of NO_x and PM while providing a higher thermal efficiency [11-13]. Because of the characteristics of utilizing low grade fuels which can be auto-ignited through the process of compression, HCCI engines are expected to exploit wet-ethanol more effectively than the traditional engines and hence can be considered as a suitable mode of combustion for the direct utilization of wet-ethanol as a primary fuel. Analysis joining quality with quantity of energy by combining the first law of thermodynamics with the second law can be used to estimate the quality of energy resource which transform into useful work. This thermodynamic modeling assist in evaluating the exergetic losses and the cycle efficiencies based on energy and exergy [14, 15]. Simultaneous applications of the first and second law applied to investigate the performance of wet-ethanol fueled HCCI engines [16-19] discovered that significant amount of exergy accompanied by engine exhaust heat is expelled which goes unutilized and leads to inefficient and environmentally harmful engine performance. This waste heat can be considered as the potential energy source for driving the heat operated refrigeration and air conditioning systems with a view to reduce the environmental impact and improve the system efficiency. Analysis of these systems from exergy point of views can provide a tool for their true theoretical performance assessment. Ejector refrigeration cycle (ERC) and absorption refrigeration cycle (ARC) are heat operated non-traditional cooling systems which can be employed for converting

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

the engine exhaust heat into the refrigeration effect. ERC's have been in use since many decades for air conditioning of residential buildings. The schematic and details of ERC applied for cold production can be seen in [20, 21]. In the field of land transportation, implementation of ejector technology for refrigerating the thermal load of air conditioner or cabin cooling is scarce and hence only few of the investigations are reported in this regard. Jaruwongwittaya and Chen [22] investigated the feasibility of employing a two-stage ejector refrigeration system in a bus and their results discovered that replacement of conventional vapor compression system with the ejector cooling system can be suitably employed as an alternative. Zengenhagen and Ziegler [23] presented the assessment of an engine exhaust heat driven jet-ejector cooling system integrated to turbocharged gasoline engines. They determined the energy efficiency and charge air temperature for the use of R134a as the refrigerant for ejector cooling. Unal [24] worked on the application of the exhaust heat driven two-phase ejector cooling system for reducing the load and fuel consumption of the engine of passenger bus. The author states that application of two-phase ejector as an expansion valve can enhance the performance of bus air conditioning system. Zhang et al. [25] developed an exhaust heat recovery unit for the automobile integrating to ejector refrigeration cycle. They analyzed the performance of ejector cooling system and recommended a suitable range of operating parameters for the efficient operation of combined system. Another suitable option to recover the engine exhaust heat is the absorption refrigeration system in which the fluid pairs LiBr-H_2O and NH_3-H_2O are normally used to produce cooling through these systems. Since a significant portion of energy supplied to internal combustion engines appear as exhaust waste heat and the use of air conditioners in vehicles arrives at the status of essential commodity for modern transportation. In this context, some theoretical and experimental studies are reported in the literature regarding the use of absorption refrigeration cycle to produce cooling through the

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

recovery of engine exhaust heat. Koehler et al. [26] developed a prototype of ARC for truck refrigeration utilizing the exhaust gas. Based on the description of simulated results they reported that system performance could be improved nearly by double after employing the ARC. Manzela et al. [27] presented the experimental investigations of an ammonia-water absorption refrigeration system driven by engine exhaust gases. They evaluated the effect of exhaust gas enthalpy and the employment of ARC on engine performance and exhaust emissions and discovered that application of ARC results in the decrease of specific fuel consumption and reduction in carbon monoxide emission.

No comparative study on the use of exhaust gas waste heat operated cooling systems bottoming with the HCCI engine fueled by wet-ethanol has been found in the literature. Therefore, it is necessary and meaningful to develop and analyze the systems that can be powered by engine exhaust gases and produce cooling for air conditioning of renewable fueled vehicles. In this study, engine operating in HCCI mode and fueled by wet-ethanol was proposed to be bottomed with the ERC and ARC, separately, to recover the exhaust heat for refrigerating the thermal load of vehicle air conditioning. A comparative thermodynamic analysis of proposed cooling-power cogeneration was conducted by considering the quality and quantity of energy transfers during the energy conversion processes of the cycle. Energetic and exergetic performances of cogeneration cycle was assessed by altering the following parameters; turbocharger pressure ratio, turbocharger compression efficiency, and ambient temperature. In addition, the distribution of fuel energy supplied in terms of energy produced and energy lost as well as the breaking down of exergy of fuel supplied into the exergy produced, exergy destruction in the major components of cogeneration cycle, and the loss of exergy due to thermal exhaust to ambient is also computed, graphed, and discussed. The contribution of exergy destruction is scrutinized and debated in regard

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

of cycle performance improvement. Results are derived for the employment of R134a as the refrigerant of ERC and LiBr-H_2O mixture as the working fluid pair for the ARC. Further, the COP of both ERC and ARC were computed with the variation in the entrainment ratio and generator temperature, respectively.

2. SYSTEM DESCRIPTION

In this paper, investigation on wet-ethanol fuelled and turbocharged HCCI engine coupled to ejector refrigeration cycle (ERC)/absorption refrigeration cycle (ARC) is performed. Figure 1 describes the working of HCCI engine fueled by wet-ethanol and combined with ejector refrigeration cycle. The engine is turbocharged and ethanol fired whose main specifications are reported in Table 1. Ambient air (1) enters the compressor (Comp1) which increases its pressure and temperature. The compressed air is heated (2-3) in regenerator (Reg1) and mixed with Wet-ethanol (4) injected into a fuel vaporizer (FV1) where it vaporizes and forms a homogeneous mixture of ethanol-water-air (5) and enters the HCCI engine. The ethanol water air mixture inducted into the cylinder and mixes with hot residual gases within the cylinder and compressed by the piston (1'-2') as shown in Figure 3. Charge gets heat through compression and this leads to ignition and combustion (2'-3') and generate power (W_{Power}). After combustion engine exhaust gases (6) enter the catalytic converter (CC1) and leave the convertor (7) at a higher temperature due to releasing of heat during combustion of carbon monoxide and unburned hydrocarbon which were not burned in the combustion chamber. Less harmful gases from catalytic converter (CC1) at higher temperature enter the turbine (T1) and generate power (W_T) to drive the turbocharger. The exhaust gases leave the turbine (8) and after having exchanged heat (8-9) with compressed air (2-3) in the regenerator (Reg1) is delivered to the boiler (Boi) for vaporizing (18-11) the refrigerant in ejector refrigeration cycle. The vaporized refrigerant (11) at high pressure and temperature

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

(primary vapor) enters the converging-diverging nozzle of ejector (Eje). The high velocity vapor at the nozzle exit (11a) creates a high vacuum at the entry of mixing chamber where it entrains secondary vapor from the evaporator (16) and mixes with the primary stream. In the mixing chamber both primary (11a) and secondary vapors (16a) mixed together and there it becomes a homogeneous mixture stream. On entering the constant area section, a normal shock wave occurs which results in the rise of pressure. After the shock, the mixed stream (12m) decelerates in the diffuser. The mixed stream (12a) from the ejector enters the condenser (Cond1) where it changes to saturated liquid refrigerant (12-13). One part of this stream as a working fluid goes to the evaporator (Evap1) to produce the desired cooling effect (15-16) through the expansion valve (EV1) (14-15), and the other part of this stream (17) is pumped back (P1) to the boiler. In Figure 2, the turbine (T2) exhaust (8) through the regenerator deliver to the generator (Gen) (9-10) for vaporizing (25-19) the solution of LiBr-H_2O. The water vapor absorbed in the absorber (Abs) by the solution is driven off and flows to the condenser (Cond2) (19) while the LiBr-H_2O solution returns to the absorber (28) through heat exchanger (HE) (26-27) and an expansion valve (EV2b) (27-28). The vaporized refrigerant is condensed in the condenser (19-20) and then flows through an expansion valve (EV2a) (20-21) to receive heat load at the evaporator (Evap2) (21-22). A LiBr-H_2O solution leaves the generator (26) and enters the absorber (28) to absorb low-pressure water vapor from the evaporator. The pump (P2) receives (23) a LiBr-H_2O solution from the absorber, raises the pressure of the strong solution (24) and pumped back to the generator (25) after passing through the heat exchanger (HE).

3. WORKING FLUID THERMODYNAMIC PROPERTIES

In the analysis and design of alternative fueled internal combustion engine, thermodynamic properties of working fluids are necessary to investigate its energetic and exergetic performances.

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

In the ideal engine cycle, the working fluid for processes occurs in the cylinder is considered simply as air. However, in actual cycles, the working fluid consists of several gaseous species.

The specific heat of gases varies with temperature in real engine application, and its variation with pressure is negligible. In the present model of molar specific heat $\bar{C}_p(T)$ variation with temperature, the coefficients (a_1-a_5) for each species, are taken from literature in the range of temperature from 300K to 3000K [28]

$$\bar{C}_p(T) = \bar{R}(a_1 + a_2T + a_3T^2 + a_4T^3 + a_5T^5) \quad (1)$$

The enthalpy and entropy of the gaseous species i in terms of specific heat capacity may be expressed as [29]

$$\bar{h}_{i_{(T)}} = \bar{h}_{i_{(T_0)}} + \int_{T_0}^{T} \bar{C}_{p_i}(T)dT \quad (2)$$

$$\bar{s}_{i\ (T,p)} = \bar{s}_{i_{(T_0,p_0)}} + \int_{T_0}^{T} \frac{\bar{C}_{p_i}(T)}{T} dT - \bar{R}ln\left(\frac{y_i p}{p_0}\right) \quad (3)$$

The intake air, combustion products, and engine exhaust gases are considered to be a mixture of various species. Below are the equations which can be utilized for the computation of molar specific enthalpy, molar specific entropy, and Molecular mass of the mixture [29].

$$\bar{h}_{mix_{(T)}} = \sum_{i=1}^{n} y_i \bar{h}_{i_{(T)}}; \qquad \bar{s}_{mix_{(T,P)}} = \sum_{i=1}^{n} y_i \bar{s}_{i_{(T,P)}}; \qquad M_{mix} = \sum_{i=1}^{n} y_i M_i \quad (4)$$

In the present analysis, fuel-air mixture is stoichiometric because ethanol containing water obtain all the required dilution. The combustion equation for burning of wet-ethanol in the HCCI engine can be considered as follows [18].

$$C_2H_5OH + 6H_2O + 3(O_2 + 3.77N_2) = 1.8\,CO_2 + 0.1CO + 0.05C_2H_5OH + 8.85H_2O + 0.2O_2 + 11.319N_2 \quad (5)$$

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

The procedure for obtaining the balanced reaction equation of an actual reaction where combustion is incomplete is not always straight forward. Actually combustion is the result of a series of very complicated and rapid chemical reactions and the product formed depend on many factors. When fuel is burned in the cylinder of an engine, the product of reaction varies with the temperature and pressure in the cylinder. The appearance of some carbon monoxide and unburned oxygen and fuel is due to incomplete mixing, insufficient time for complete combustion, and other factors. The combustion of wet-ethanol in HCCI engine was taking place by assuming a chemical reaction with 90% combustion efficiency.

Unburned hydrocarbons and carbon monoxide which formed as emissions in HCCI engine can be further oxidized by using Pd-loaded $SiO_2 - Al_2O_3$ as the oxidizing catalyst. Because of having the stoichiometric combustion of fuel-air mixture and its operation at higher compression ratio, the temperature of HCCI engine exhaust gases that enters the catalytic convertor is considered to be enough for oxidizing the carbon monoxide and unburned fuel. Assuming the complete conversion of unburned fuel and carbon monoxide and the application of mass and energy balances, the composition and temperature of gases at the exit of catalytic convertor can be determined

$$\sum_{Product} n_e\left(\bar{h}_f^\circ + \Delta\bar{h}\right)_e = \sum_{Reactant} n_i\left(\bar{h}_f^\circ + \Delta\bar{h}\right)_i \quad (6)$$

Where, i refers the incoming streams entering the catalytic convertor and e refers the exiting stream of products leaving the convertor [19]. The equation for chemical reaction of engine exhaust gases in the catalytic convertor is assumed to be as follows:

$$1.8\,CO_2 + 0.1\,CO + 0.05\,C_2H_5OH + 8.85\,H_2O + 0.2\,O_2 + 11.319\,N_2 = 2\,CO_2 + 9\,H_2O + 11.319\,N_2 \quad (7)$$

3.1 Turbocharger compressor

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

The air inducted is first flows through the compressor to increase its pressure and also the temperature. For a given pressure ratio and considered polytropic efficiency of the compressor, the turbocharger compressor outlet temperature can be determined after applying the equation,

$$T_2 = T_1 \left(\frac{P_2}{P_1}\right)^{\frac{(\gamma-1)}{\gamma \eta_C}} \quad (8)$$

3.2 Regenerator

Air leaving the compressor is delivered to regenerator (an intake air to exhaust gases heat exchanger) where the intake is preheated through exhaust gases. This preheating of air is required to evaporate the wet-ethanol in the fuel vaporizer. The regenerator performance is determined by through its effectiveness which is given by

$$\varepsilon_h = \frac{\left[(\dot{m}.c_p)_2 (T_3 - T_2)\right]}{\left[(\dot{m}.c_p)_{min} (T_8 - T_2)\right]} \quad (9)$$

The assumption, $(\dot{m}.c_p)_2 = (\dot{m}.c_p)_{min}$ is taken to estimate the temperature, T_3, the air temperature at the regenerator exit and with this relation, it is expressed as

$$T_3 = T_8 \varepsilon_h + T_2(1 - \varepsilon_h) \quad (10)$$

The temperature at the regenerator exit is computed by applying the energy balance over it which is reported in Table 2.

3.3 HCCI engine

The premixed ethanol-water mixture leaving the mixer enters the HCCI engine. The charge enters the engine and before compression it is mixed with hot residual gases. Due to operation of engine at higher compression ratio, the charge is heating up to the level of ignition and combustion. But auto-ignition process is observed in HCCI engine for combustion which is controlled by chemical

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

kinetics [9]. The efficiency of combustion process is considered to be 90%. External exhaust gas recirculation is absent in this model of engine [30]. The temperature of the mixture containing the fresh charge and residual gases at the end of engine intake process (i-1') is computed by employing the expression and the ratio of pressure for inlet/exhaust is considered to be 1.4 [31].

$$T_{1'} = \frac{\mathrm{T}_i(1-f)}{1-\frac{1}{(n.\mathrm{r}_c)\left[\frac{\mathrm{p}_e}{\mathrm{p}_i}+(n-1)\right]}} \tag{11}$$

Where 1' signifies the location before the process of compression of fuel-air-residual gas mixture and $P_{1'} = P_{1.}$

Polytropic compression stroke (1'-2')

$$P_{2'} = P_{1'}\left(\frac{V_{1'}}{V_{2'}}\right)^n = P_{1'}(\mathrm{r}_c)^n \tag{12}$$

$$T_{2'} = T_{1'}(\mathrm{r}_c)^{n-1} \tag{13}$$

Assuming the heat addition process (2'-3') as constant volume gives

$$T_{3'} = T_{2'} + q_{in}(1-f)/c_v \tag{14}$$

$$P_{3'} = P_{2'}\left(\frac{T_{3'}}{T_{2'}}\right) \tag{15}$$

Expansion stroke (3'-4') is polytropic and this gives

$$P_{4'} = P_{3'}\left(\frac{1}{\mathrm{r}_c}\right)^n \tag{16}$$

$$T_{4'} = T_{3'}\left(\frac{1}{\mathrm{r}_c}\right)^{n-1} \tag{17}$$

Polytropic blowdown (4'-5')

$$T_{5'} = T_{4'}\left(\frac{P_{4'}}{P_e}\right)^{\frac{(1-n)}{n}} \tag{18}$$

$$P_{5'} = P_e \tag{19}$$

The constant pressure polytropic exhaust stroke (5'-6')

$$T_{5'} = T_e \tag{20}$$

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

$$P_6 = P_{5'} = P_e \tag{21}$$

$$f = \frac{1}{\mathrm{r}_c}\left(\frac{P_e}{P_4}\right)^{\frac{1}{n}} \tag{22}$$

Residual gas fraction (f=0.03) is found to be lower due to large compression ratio.

The heat lost to the ambient during the processes of exhaust and blowdown is computed by assuming 9% loss of heat during these processes [30].

The formulation for calculating the chemical exergy in HCCI engine may be seen as below

$$e_{HCCI}^{ch} = -\Delta g + \bar{R}\mathrm{T}_0\left\{x\mathrm{O}_2 ln\frac{P_{\mathrm{O}_2}^{00}}{P^0} + y\mathrm{H}_2 O ln\frac{P_{H_2O}^{00}}{P^0} - \sum_K \mathrm{x}_k ln\frac{P_k^{00}}{P^0}\right\} \tag{23}$$

The details of the symbols used in above equation can be found in the Ref. [18].

3.4 Turbine

Exhaust from the engine is expanded in the turbine to produce power for driving the compressor.

The temperature at turbine outlet can be determined by [9]

$$\mathrm{T}_8 = \mathrm{T}_7\left(\frac{\mathrm{P}_8}{\mathrm{P}_7}\right)^{\frac{(n-1)\eta_T}{n}} \tag{24}$$

4. EVALUATION CRITERIA AND SYSTEM MODELING

The proposed combined cooling-power cycle is analyzed using Engineering Equation Solver (EES) which is commercially available and provides a built in mathematical and thermodynamic property functions for solving the equations describing the model developed through balances of mass, energy, and exergy [37]. Each component of the system proposed is considered as a steady state control volume, with transfer of energy by heat, work, and mass of fluid enters and leaves. The equations derived after applying the balances of mass, energy, and exergy were employed to formulate the thermodynamic model. The thermodynamic properties of the refrigerant R134a used

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

in ERC was computed after employing the REFPROP V 9.1 and the properties of LiBr-H_2O fluid pair for ARC were taken from the Ref. [38, 39].

Application of second law of thermodynamics along with the traditional first law approach leads to analysis that joins the quality with quantity of energy during the process of energy conversion and it quantifies the destruction and loss of exergy appears at various locations of the system.

For the energetic and exergetic analyses of the cooling-power cogeneration system considered, the thermodynamic model was developed based on the assumptions below [18, 25, 26]:

1- All processes proceed under steady state condition.
2- Effects of kinetic and potential energies is not considered during the energy conversion.
3- Thermodynamic equilibrium is maintained between the inlet and exit of cycle key components.
4- The exchange of heat between the components and environment is ignored.
5- Expansion valves of both ERC and ARC are isenthalpic and pressure reduction devices.
6- The pressure drops in the pipes are not considered.
7- Physical exergies of the ERC fluid and LiBr-H_2O fluid pair for ARC are only considered.
8- State of the working fluid at condenser outlet is saturated liquid.
9- The atmospheric pressure P_0 and temperature T_0 considered for the reference-environment were assumed as (25°C and 101.325 kPa), respectively.

The traditional first law of thermodynamics along with law of conservation of mass applied to any control volume operates at steady state and disregards the change in kinetic and potential energies gives

$$\Sigma \dot{m}_i - \Sigma \dot{m}_o = 0 \tag{25}$$

$$\Sigma \dot{Q} - \Sigma \dot{W} + \Sigma \dot{m}_i h_i - \Sigma \dot{m}_o h_o = 0 \tag{26}$$

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

After applying the Eqs. (12) and (13), the energy balance equations are formulated for the wet-ethanol fuelled HCCI engine corresponding to its P-V diagram shown in Figure 3, and also for the ERC and ARC which are employed for engine exhaust heat recovery. The formulated equations are displayed in Table 2 &3.

Exergy can be defined as the maximum useful work that can be obtained through a process which brings the system into equilibrium with a thermal reservoir, usually the environment.

The exergy rate balance equation that can be applied to any control volume under steady state operating condition can be presented as

$$\Sigma\left(1-\frac{T_0}{T}\right)\dot{Q}_j - \Sigma\dot{W} + \Sigma\dot{m}_i e_{x_i} - \Sigma\dot{m}_o e_{x_o} - \dot{E}_{x,D} = 0 \tag{27}$$

Where, $\dot{m}$ is the mass flow rate of a substance. First two terms on the left-hand side of Eq. (14) indicates the exergy transfer via heat, and work during a process.

$\dot{E}_{x,D}$ is the exergy destruction rate during the process. e_x is the flow exergy per unit mass of a substance which can be break down into the chemical and physical exergy and is given by;

$$e_x = e_x^{ph} + e_x^{ch} \tag{28}$$

Physical specific exergy (e_x^{ph}) accounts for change in enthalpy of a gaseous species and loss in entropy during its transition from a specified state to restricted dead state. It is given by the following expression

$$e_x^{ph} = (h - h_0) - T_0(s - s_0) \tag{29}$$

The chemical exergy (e_x^{ch}) takes into account the deviations in composition of the substance from the substances occupied the reference environment. It may be expressed as [19]

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

$$e_x^{ch} = \sum x_i(\mu_i^* - \mu_0) = -\Delta g_i^0 + \sum \bar{R}\, T_0 x_i ln\left(\frac{p}{p_{ref}}\right) \quad (30)$$

Where μ_i is the chemical potential and g_i is the Gibbs function of species "i", and they can be evaluated at a specified temperature T and pressure p. T_0 and p_{ref} are the temperature and pressure at the reference environment, respectively. For engine applications, the reference environment temperature and pressure are taken to be as $T_0 = 298\ K$ and $p_{ref} = 1.013\ bar$. For chemical exergy computation in engine applications, the molar composition of the environment is to be taken as: 20.35% O_2, 75.67% N_2, 0.03% CO_2, 3.03% H_2O, and 0.92% other substances [32].

Considering the definition of exergy under the assumptions listed, and proposed balances, the equations for the calculation of exergy destroyed during the process occurring in the main components of the proposed configuration of combined energy system are developed and shown in Table 4&5.

Evaluation criteria:

Habitually, the performance of energy conversion system is determined based on first law efficiency which is a measure of the percentage of the energy input that is transformed to useful energy output. The energy efficiency of the proposed cooling power-cogeneration system which can be evaluated by applying first law, may be presented as:

$$\eta_{en,comb} = \frac{\dot{W}_{BT-}\dot{W}_{Eje,P}+\dot{Q}_{ref}}{\dot{m}_f LHV \eta_{cc}} \quad (31)$$

where, $\dot{W}_{BT}$ is the brake thermal power, and $\dot{W}_{Eje,P}$ is the power consumed by the refrigerant pump, respectively. $\dot{m}_f$ is the rate of fuel consumption, and LHV is the lower heating value of fuel, $\dot{Q}_{ref}$ is the rate of cooling capacity, and η_{cc} is the combustion efficiency.

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

The network produced in the HCCI engine which will be available as a useful work may be obtained after using the expression [30]

$$WD = \left[\left(\frac{P_3'V_3' - P_4'V_4'}{n-1}\right) - \left(\frac{P_2'V_2' - P_1'V_1'}{n-1}\right)\right] \quad (32)$$

$$W_{BT} = Actual\ WD = 0.8 * WD \quad (33)$$

Where W_{BT} is the actual work done which is used to calculate brake thermal efficiency of HCCI engine

Since the two energy outputs of the proposed cycle; power generation and rate of cooling produced are different from quality point of view because power is free from entropy and cooling is associated with entropy, therefore, thermal efficiency based on traditional first law approach is not suitable for systems which produce two different outputs of energy. In this context, a relatively new term called exergy efficiency was introduced which provides a more accurate measure of thermodynamic performance of the cycle and it is given by

$$\eta_{ex,comb} = \frac{\dot{W}_{BT} - \dot{W}_{Eje,P} + \dot{E}_{ref}}{\dot{m}_f LHV \eta_{cc} \varphi} \quad (34)$$

where, $\dot{E}_{ref}$ is the amount of exergy associated with the cooling capacity $\dot{Q}_{ref}$ and is given by

$$\dot{E}_{ref} = \dot{Q}_{ref}\left(\frac{T_0 - T_{Evap}}{T_{Evap}}\right) \quad (35)$$

Where, T_{Evap} is the evaporator temperature, T_0 is the ambient temperature, and ϕ is the ratio of exergy of the fuel to its energy [19].

5. RESULTS AND DISCUSSION

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

Below are the computational results revealed after conducting the energetic and exergetic performances of the HCCI engine fueled by wet-ethanol and bottoming with the ERC/ARC for cooling-power cogeneration. Results are computed for the use of R134a and R290 as the refrigerants of ERC, and LiBr-H_2O as working fluid pair for ARC. In the analysis, the effects of altering the turbocharger pressure ratio, turbocharger compression ratio, ambient temperature is investigated on the efficiencies of combined cycle. The effect of entrainment ratio and generator temperature variation was also examined on the COPs of ERC and ARC, respectively.

Fig. 4 represents the change in energy efficiency of HCCI engine based combined system after altering the turbocharger pressure ratio. The HCCI engine efficiency evaluated on first law basis is found to be increasing with the rise of pressure ratio across the turbocharger. This is due to the reason higher turbocharger pressure improves the fuel-air mixing along with the increase of charge density which raises the temperature of mixture and hence the engine power output. Further, for the same turbine and compressor efficiencies and regenerator effectiveness any rise in turbocharger pressure ratio results in the increase of engine exhaust temperature which is an inlet temperature to the boiler and this will increase the heat transfer to refrigerant passes through the boiler of ERC or the generator of ARC and this in turn improves the cooling capacity of the evaporator of cycle concerned. This simultaneous increase in the output of HCCI engine and the cooling capacity of refrigeration cycle contributes to rise in the efficiency of cooling-power cogeneration cycle. Change of refrigerant in the ERC from R134a to R290 reduces the energy efficiency of cogeneration cycle because of the considerable difference in boiling point of (-26.1℃) for R134a and (-42.11^{0}C) for R290. Further, lower refrigeration output for R290 (propane) operated system attributes to the low critical temperature of propane (96.74^{0}C) compared to higher critical temperature of R134a (101.1^{0}C). Relatively a high motive flow

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

temperature is required to produce the decent refrigeration output and hence the lower critical temperature of refrigerant is a hinderance in obtaining the higher output. It is further noticed that an increase in pressure ratio across the turbocharger from 2.5 to 3.5 raises the HCCI engine first law efficiency from 44.09% to 46.32%, and for cogeneration it is increased from 47% to 49.21% when R134a is used as ERC refrigerant and from 46.06% to 48.29% when R290 is the ERC refrigerant, and in case of ARC bottoming it is increased from 55.08% to 57.25%.

Fig. 5 displays the influence of turbocharger pressure ratio on the exergy efficiency of HCCI engine and the cogeneration system for both ERC and ARC, respectively. The exergy efficiency of HCCI engine is shown to be distinct than its energy efficiency and this difference is found because of the considerable difference between the exergy and energy of fuel supplied. Since exergy efficiency of HCCI engine is the ratio of power exergy to the fuel exergy, therefore, exergy efficiency is higher than the corresponding energy efficiency of HCCI engine. The exergy efficiency of combined cycle (power and cooling) is found to be lower than its energy efficiency. The exergy of refrigeration is thermodynamically lower than the refrigeration capacity and since this difference dominates over the difference between the exergy and energy of ethanol, therefore, exergy efficiency of cogeneration which is the exergetic output to the exergy of fuel input is appeared to be lower than the energy efficiency of cogeneration. Similar to the trend of energy efficiency of cogeneration, the exergy efficiency of cogeneration cycle where cooling is produced by the bottoming of ARC is found to be larger than the exergy efficiency of cogeneration bottoming with ERC.

Fig. 6 displays the impact of turbocharger compressor efficiency on the energy efficiency of HCCI engine and cogeneration with both ERC and ARC. The energy efficiency of HCCI engine is found to be increased at the rise of turbocharger compressor efficiency. This is because higher

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

turbocharger efficiency decreases the compressor power input and the turbine power as well which results in the decrease of engine exhaust temperature and the higher energy efficiency of the engine. Further, it is to be noted that increase in turbocharger compressor efficiency results in the decrease of energy contents of gases enters the boiler/generator of ERC/ARC which decreases the mass flow rate of the refrigerant and hence the refrigeration capacity of the bottoming cycle. Since increase in HCCI engine power dominates over the decrease of cooling capacity of ERC/ARC, therefore, the energy efficiency of cooling-power cogeneration cycle is also increased when the turbocharger compressor efficiency increases. At a given turbocharger compressor efficiency, the energy efficiency of cogeneration cycle ARC is employed as the bottoming cycle is considerably higher than ERC employed cycle for cooling. It is shown that raising of turbocharger efficiency from 0.7 to 0.9 increases the HCCI engine efficiency from 44.51% to 45.93%, and the energy efficiency of R134a operated ERC cogeneration increases from 47.41% to 48.83% and of the ARC cogeneration increases from 46.48% to 47.9%, respectively.

The variation of the exergy efficiency of HCCI engine and the cogeneration cycle for ERC and ARC with the turbocharger compressor efficiency is depicted in Fig. 7. It is observed that as the turbocharger compressor efficiency increases the exergy efficiency of HCCI engine increases. This is because at the higher turbocharger compressor efficiency the irreversibility decreases as in case of compressor, irreversibility is the difference between the actual power and reversible power input. Since the exergy of refrigeration is significantly less than the refrigeration effect, therefore, the exergy efficiency of cogeneration cycle is little higher than the exergy efficiency of HCCI engine. It is recorded that increase in turbocharger compressor efficiency from 0.7 to 0.9 raises the exergy efficiency of HCCI engine from 52.72% to 54.56%, and for the ARC cogeneration cycle it is increased from 53.75% to 55.58%, respectively.

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

Fig. 8 shows the effect of changing the ambient temperature on the energy efficiency of HCCI engine and cooling-power cogeneration cycle. The energy efficiency of HCCI engine is found to be decreased with the rise of ambient temperature because of the reduction in volumetric efficiency and increase in fuel consumption in the combustion chamber at higher ambient temperature. It is further to be noted that the rate of decrease in energy efficiency of cogeneration cycle is less than the rate of decrease in the energy efficiency of HCCI engine. This is because the refrigeration capacity of ERC/ARC start increasing with the increase of ambient temperature due to the reduction of energy losses at the heat exchanger components of ERC and ARC but the decrease in the engine output dominates over the increase in capacity of refrigeration. Therefore, the efficiency of cogeneration diminishes gradually when the ambient temperature rises. Since the refrigeration capacity of ARC at a given temperature is greater than the refrigeration capacity of ERC, therefore, the efficiency of HCCI engine combined with the ARC is considerably greater than the case of ERC bottoming. Further, it is observed that at a given ambient temperature the exergy destroyed in the R290 operated ERC at the condenser and expansion valves is greater than the exergy destroyed in the R134a operated ERC, therefore, the efficiency of R143a operated ERC cogeneration is found to be higher than the efficiency of R290 operated cogeneration cycle.

The influence of ambient temperature on the exergy efficiency of HCCI engine and the cogeneration cycle was also computed and depicted in Fig. 9. The exergy efficiency of HCCI engine behave similar to its energy efficiency but diminishes more gradually because of decrease in physical exergy of fuel at higher ambient temperature. On the other hand, the exergy efficiency of cogeneration cycle combined with ARC/ERC start increasing because of the increase of exergy of refrigeration with the rise of ambient temperature. By definition, exergy of refrigeration is the ratio of the refrigeration capacity to the Carnot COP operates between the refrigeration temperature

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

and ambient temperature. Since cooling capacity of ARC for the same heat input is greater than the cooling capacity of ERC, therefore, the rate of increase in the exergy efficiency of cogeneration (HCCI engine combined with ARC) is significantly higher than the rate of increase in the exergy efficiency of cogeneration (HCCI engine combined with ERC) for the same increase of ambient temperature.

The effect of change in the entrainment ratio and refrigerant type on the energetic and exergetic COPs of ERC was also examined and is shown below in Fig. 10. Both energetic and exergetic COPs are found to be decreased gradually when the entrainment ratio is increased. This is because entrainment ratio (refrigerant mass flow rate via evaporator to the flowing mass entering the ejector) increases with the increase of boiler temperature and for the constant heat load of the boiler this leads to the reduced mass flow of refrigerant entering the ejector and hence a lower entrained mass flow rate from the evaporator. Since the decrease in entrained mass flow rate of refrigerant is less than the decrease of its motive mass flow rate, therefore, amount of cooling produced which diminishes gradually which shows a slight decrement in the COP of ERC. The magnitude of the exergy of refrigeration is always less than the refrigeration effect but both depend strongly on mass of refrigerant flowing through the evaporator to generate cooling, therefore, the exergetic COP of ERC is also decreases gradually with the rise of entrainment ratio.

Fig. 11 shows the variation in COP of the absorption refrigeration cycle (ARC) with the rise in generator temperature. At a constant evaporative cooling, temperature of heat transfer fluid has been increased which results in the increase of generator temperature. This rise in temperature also increase the pressure. This in turn decrease the pressure at the outlet of expansion valve, which determines the evaporator temperature and refrigeration capacity. Therefore, the heat input to generator is increased, while keeping the cooling output as constant in the ARC. Since the energy

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

and exergy supplied to the generator plays an important role in the determination of COP, therefore, the energetic and exergetic COP of ARC are decreased at the rise of generator temperature. A rise in generator temperature from 359K to 367K, declines the energetic COP of ARC from 0.7333 to 0.7299 and its exergetic COP from 0.1100 to 0.1094, respectively.

The breaking down of fuel energy supplies to combined cycle (HCCI engine-ERC) at its base line operation is illustrated in Fig. 12. From the ERC employed combined cycle where R134a is the refrigerant, it is determined that out of 333.36 kW of fuel energy supplied, 130.1 kW is the produced power and 9.767 kW is the refrigeration of thermal load. About 101.34 kW of energy is lost via engine heat transfer and 45.72 kW is the energy lost via thermal exhaust to atmosphere, and 59.07 kW is the energy losses at the refrigerant condenser. In case of R290 operated ERC, because of the lower boiling point, refrigeration effect produced by R290 refrigerant is considerably less than R134a and found as 6.706 kW and energy losses at the refrigerant condenser also reduces from 59.07 kW to 56.04 kW.

The breaking down of fuel energy supplies to combined cycle (HCCI engine-ARC) at its base line operation is illustrated in Fig. 13. The distribution of fuel energy supplied to HCCI engine-ARC is showing quite different figure for the refrigeration effect and the energy losses at the refrigerant condenser. The refrigeration produced by ARC is significantly larger than ERC because of the higher boiling point of water which act as refrigerant in the LiBr-H_2O mixture and it is found as 35.62 kW. The energy losses at the condenser and absorber of ARC are 38.3 kW and 46.29 kW, respectively. The distribution of fuel energy in the components of HCCI engine is unaltered by changing the type of bottoming cycle from ERC to ARC.

Employment of the principle of the decrease of exergy enunciated from the combined application of the first and second laws of thermodynamics to the developed wet-ethanol fuelled combined

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

power and cooling cycle provides a breakdown of the amount of fuel exergy supplied and is converted into engine power exergy and the refrigeration exergy, and loss of exergy as well as the exergy degraded due to generation of entropy during energy transformation in the cycle components. To this effect, the breaking down of input exergy into the exergy destroyed, loss of exergy, and the exergy produced was computed and is depicted in Fig.14.

In case of ERC employed bottoming cycle for cold production, from the 275.5 kW of exergy supplied to combined cycle, 130.1 kW is the power exergy and 0.811 kW is the exergy of refrigeration (R134a refrigerant). Following are seen as the components of high exergy destruction; HCCI engine (61.19 kW), boiler (20.46 kW), catalytic convertor (17.89 kW), and the exergy destruction in rest of the components of combined cycle is significantly less compared to above figures and hence not mentioned by shown clearly in Fig. 14. The losses of exergy via engine heat transfer and thermal exhaust are found to be as 37.21 kW and 7.818 kW, respectively. Replacement of R134a with R290 as the ERC refrigerant changes the exergy of refrigeration from 0.811 kW to 0.5568 kW, exergy destruction in the boiler from 20.46 kW to 21.29 kW, exergy destruction in the ejector from 3.643 kW to 2.876 kW, exergy losses at the condenser from 1.044 kW to 1.243 kW, respectively. These differences of refrigeration exergy and exergy destruction in the boiler, ejector, and condenser arises due to the considerable difference of boiling point of two refrigerants.

The breaking of fuel exergy into the exergy destroyed, exergy produced, and losses of exergy via exhaust to environment for the combined cycle of HCCI engine-ARC configuration is displayed in Fig. 15. A significantly high exergy of refrigeration is found 2.819 kW compared to less than 1 found in case of ERC. Exergy destruction in the generator is 18.69 kW and in the condenser is 1.18 kW. These figures are different from ERC because of having a considerable difference of

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

boiling points of R134a, R290, and water which are utilized as refrigerants in cycles considered for analysis. The exergy destruction in the evaporator of ARC is 0.105 kW compared to 0.805 Kw in case of R134a ERC. This distinction arises because of the difference of the LiBr-H_2O ARC evaporator temperature above than 0^0C and can be much less than 0^0C in case of ERC evaporator. The exergy destruction in the absorber is also found to be considerable 2.115 kW because of the mixing (source of entropy generation) of absorbent and refrigerant in the absorber. Similar to energy distribution, exergy distribution of fuel in the HCCI engine components is also found unaltered after changing the bottoming cycle from ERC to ARC.

6. CONCLUSION

This investigation presents a comparative study of the implementation of low-grade ethanol fuel in HCCI engine combined with two different configurations of cooling cycles for combined production of engine power and refrigeration. HCCI engine was used for power generation and the ERC/ARC powered by the engine exhaust heat were employed as bottoming cycles for the refrigeration of a thermal load. Cogeneration cycle energetic and exergetic performances were analysed after altering the turbocharger compressor ratio, ambient temperature, compressor isentropic efficiency, entrainment ratio of ERC, and generator temperature of ARC. Breaking down of the fuel energy/exergy supplied to the combined cycle in terms of the energy/exergy produced as output, energy/exergy loss to ambient, and the exergy destruction due to irreversibilities in the cycle components is displayed. Furthermore, comparison of computed results for the ARC (LiBr-H_2O) and ERC (refrigerants; R134a and R290) are presented and debated. The results revealed that operating parameters defined above, significantly influences the cogeneration performance. The key findings obtained from this investigation can be summarized as follows:

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

- Both energetic and exergetic efficiencies of HCCI engine as well as of cogeneration cycle are found slightly increasing when the turbocharger compressor efficiency is increased.
- Elevation of turbocharger pressure resulting in the simultaneous increase of HCCI engine power and refrigeration of thermal load.
- The increase of ambient temperature involves the decline of energy and exergy efficiencies of HCCI engine and also the energy efficiency of cogeneration while the exergy efficiency of cogeneration is found increasing.
- Variation of entrainment ratio and type of refrigerant employed shows a marginal decline in the energetic and exergetic COPs of ERC.
- Bottoming of cooling cycles reveals that at the system base line operation, the refrigeration performed by ARC, R134a ERC, and R290 ERC were to be as 35.62 kW, 9.767 kW, and 6.706 kW, respectively.
- HCCI engine (61.19 kW), boiler of ERC (20.46 kW), generator of ARC (18.69 kW), catalytic convertor (17.89 kW) are discovered as the components of significant exergy destruction.
- Replacement of R134a with R290 as the ERC refrigerant changes the exergy of refrigeration from 0.811 kW to 0.5568 kW, exergy destroyed in the boiler from 20.46 kW to 21.29 kW, exergy destroyed in the ejector from 3.643 kW to 2.876 kW, exergy losses at the condenser from 1.044 kW to 1.243 kW, respectively.
- The exergy destruction in the evaporator of ARC is 0.105 kW compared to 0.805 kW in case of R134a ERC.

Findings of current study clearly reveals the benefits of exergy analysis over the traditional energy analysis. The comparison between the destruction of exergy at the individual component

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

determined that HCCI engine carries a potential of major improvement in view of improving the overall efficiency of proposed cogeneration cycle. Estimation of the fuel's energy and exergy distribution in the proposed configuration of power-cooling cogeneration behaves distinctly.

Nomenclature

c_p	specific heat at constant pressure, kJ/kg
$\dot{E}x$	exergy rate, kW
e_x^{ch}	chemical exergy, kJ/kg
e_x^{ph}	physical exergy, kJ/kg
e_x	specific flow exergy, kJ/kg
g	specific Gibbs function, kJ/kg
$\bar{h}_f^\circ$	enthalpy of formation on molar basis, kJ/kmol
h	specific enthalpy, kJ/kg
$\dot{m}$	rate of flowing mass, kg/s
M	molecular mass, kg/kmol
n	polytropic index, number of moles
p	pressure, kPa
$\dot{Q}$	heat transfer per unit time, kW
R	molar gas constant, J/K mol
s	entropy per unit mass, kJ/kg-K

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

T	temperature, K
$\dot{W}$	output power, kW
x	mass fraction
y	mole fraction

Greek symbols

η_{en}	energy efficiency, %
η	efficiency, %
η_{ex}	exergy efficiency, %

Subscript

a	air
Abs	absorber
Boi	boiler
BT	brake thermal
CC1	catalytic converter of ERC
CC2	catalytic converter of ARC
ch	chemical
Comp1	compressor of ERC
Comp2	compressor of ARC
Cond1	condenser of ERC
Cond2	condenser of ARC
D	destruction

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

Eje	ejector
EV	expansion valve
Evap1	evaporator of ERC
Evap2	evaporator of ARC
exh	exhaust
exp	expansion
f	fuel
FV1	fuel vaporizer of ERC
FV2	fuel vaporizer of ARC
Gen	generator
HE	heat exchanger
i	inlet
o	Outlet
P1	pump of ERC
P2	pump of ARC
ph	physical
Reg1	regenerator of ERC
Reg2	regenerator of ARC
rg	residual gas
sol	solution
surr	surrounding

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

T1 turbine of ERC

T2 turbine of ARC

Abbreviations

ARC Absorption Refrigeration Cycle

ERC Ejector Refrigeration Cycle

HCCI Homogeneous Charge Compression Ignition

LHV Lower Heating Value

LiBr-H_2O Lithium Bromide-Water

REFERENCES

[1] Armstead, J.R., Miers, S.A., 2014, "Review of waste heat recovery mechanisms for internal combustion engine," ASME J. Thermal Sci. Eng. Appl., 6 (1), pp. 014001. https://doi.org/10.1115/1.4024882

[2] Srinivas, K.K., Mago, P.J., and Krishnan, S.R., 2010, "Analysis of exhaust waste heat recovery from a dual fuel low temperature combustion engine using an organic Rankine cycle," Energy, 35(6), pp. 2387-2399, https://doi.org/10.1016/j.energy.2010.02.018

[3] Sara, H., Chalet, D., Cormerais, M., 2018, "Different configurations of exhaust gas heat recovery in internal combustion engine: evaluation on different driving cycles using numerical simulations," ASME J. Thermal Sci. Eng. Appl., 10 (4), pp. 041010. https://doi.org/10.1115/1.4039304

[4] Hill, J., 2007, "Environmental costs and benefits of transportation biofuel production from food and lignocelluloses-based energy crops: a review," Agron. Sustain. Dev., 27 (1), pp. 1-12, https://doi.org/10.1051/agro:2007006

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

[5] Ma, F., and Wang, Y., 2008, "Study on the extension of lean operation limit through hydrogen enrichment in a natural gas spark-ignition engine," Int. J. Hydrogen Energy, 33 (4) pp. 1416-1424. https://doi.org/10.1016/j.ijhydene.2007.12.040

[6] Prashant, D.K., Lata, D.B., and Joshi, P.C., 2016, "Investigation on the effect of ethanol blend on the combustion parameters of dual fuel diesel engine," Appl. Therm. Eng., 96 pp. 623-631. https://doi.org/10.1016/j.applthermaleng.2015.11.051

[7] Fagundez, J.L.S., Sari, R.L., Mayer, F.D., Martins, M.E.S., and Salau, N.P.G., 2017, "Determination of optimal wet ethanol composition as a fuel in spark ignited engine," Appl. Therm. Eng., 112 pp. 317-325. https://doi.org/10.1016/j.applthermaleng.2016.10.099

[8] Patzek, T.W., 2004, "Thermodynamics of corn-ethanol bio-fuel cycle," Crit. Rev. Plant Sci., 23(6) pp. 519-567. https://doi.org/10.1080/07352680490886905

[9] Frias, J.M., Aceves, S.M., and Flowers, D.L., 2007, "Improving ethanol life cycle energy efficiency by direct utilization of wet-ethanol in HCCI engines," ASME J. Energy Resour. Technol., 129 (4) pp. 332-337. https://doi.org/10.1115/1.2794768

[10] Mack, J.H., Aceves, S.M., and Dibble, R.W., 2009, "Demonstrating direct use of wet-ethanol in a homogeneous charge compression ignition engine," Energy 34 (6) pp. 782-787. https://doi.org/10.1016/j.energy.2009.02.010

[11] Yao, M.F., Zhang, Z., and Liu, H., 2009, "Progress and recent trends in homogeneous charge compression ignition (HCCI) engines," Prog. Energy Combust. Sci., 35 (5) pp. 398-437. https://doi.org/10.1016/j.pecs.2009.05.001

[12] Christensen, M., Johansson, B., and Einwall, P., 1997, "Homogeneous charge compression ignition (HCCI) using iso-octane, ethanol and natural gas – a comparison with

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

spark ignition operation," SAE Technical Paper No. 972874. https://doi.org/10.4271/972874

[13] Mack, J.H., Dibble, R.W., Buchholz, B.A, and Flowers, D.L., 2005, "The effect of the di-tertiary butyl peroxide (DTBP) additive on HCCI combustion of fuel blends of ethanol and di-ethyl ether," SAE Technical Paper No. 2005-01-2135. https://doi.org/10.4271/2005-01-2135

[14] Rokopoulos, C.D., and Giakoumis, E.G., 2004, "Parametric study of transient turbocharged diesel engine operation from the second law perspective," SAE Technical Paper No. 2004-01-1679. https://doi.org/10.4271/2004-01-1679

[15] Caton, J.A., 2012, "Exergy destruction during the combustion process as functions of operating and design parameters for a spark ignition engine, "Int J. Energy Res., 36(3) pp. 368-384. https://doi.org/10.1002/er.1807

[16] Boldaji, M.R., Gainey, B., and Lawler, B., 2019, "Thermally stratified compression ignition enabled by wet ethanol with a split injection strategy: A CFD simulation study," Appl. Energy, 235 pp. 813-826. https://doi.org/10.1016/j.apenergy.2018.11.009

[17] Saxena, S., Schneider, S., Aceves, S., and Dibble, R., 2012, "Wet ethanol in HCCI engines with exhaust heat recovery to improve the energy balance of ethanol fuels," Appl. Energy, 98 pp. 448-457, https://doi.org/10.1016/j.apenergy.2012.04.007

[18] Khaliq, A., Trivedi, S.K., and Dincer, I., 2011, "Investigation of a wet ethanol fuelled HCCI engine based on first and second law analysis," Appl. Therm. Eng., 31(10) pp. 1621-1629.

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

[19] Khaliq, A., and Trivedi, S.K., 2012, "Second law assessment of a wet-ethanol fueled HCCI engine combined with organic Rankine cycle," ASME J. Energy Resour. Technol., 134(2) pp. 022201. https://doi.org/10.1115/1.4005698

[20] Roman, R., and Hernandez, J.I., 2011, "Performance of ejector cooling systems using low ecological impact refrigerants," Int. J. Refrigeration, 34 (7) pp. 1707-1716. https://doi.org/10.1016/j.ijrefrig.2011.03.006

[21] Galindo, J., Gil, A., Dolz, V., Ponce-Mora, A., 2020, "Numerical optimization of an ejector for waste heat recovery used to cool down the intake air in an internal combustion engine," ASME J. Thermal Sci. Eng. Appl., 12 (5), pp. 051024. https://doi.org/10.1115/1.4046906

[22] Jaruwongwittaya, T., and Chen, G., 2012, "Application of two stage ejector cooling system in a bus," Energy Procedia, 14 pp. 187-197. https://doi.org/10.1016/j.egypro.2011.12.916

[23] Zegenhagen, M.T., and Ziegler, F., 2015, "Feasibility analysis of an exhaust gas waste heat driven jet-ejector cooling system for charge air cooling of turbocharged gasoline engines," Appl. Energy, 160 pp 221-230. https://doi.org/10.1016/j.apenergy.2015.09.057

[24] Unal, S., 2015, "Determination of the ejector dimensions of a bus air-conditioning system using analytical and numerical methods," Appl. Therm. Eng., 90 pp. 110-119. https://doi.org/10.1016/j.applthermaleng.2015.06.090

[25] Zhang, H., Wang, L., Jia, L., and Wang, X., 2018, "Performance investigation of automobile waste heat recovery system for ejector refrigeration cycle" 13th IEEE Conference on Industrial Electronics and Applications (ICIEA): pp. 400-405. Wuhan, China, May 31-June 02. DOI: 10.1109/ICIEA 2018.8397750.

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

[26] Koehler, J., Tegethoff, W.J., Westphalen, D., and Sonnekalb, M., 1997, "Absorption refrigeration system for mobile applications utilizing exhaust gases," Heat and Mass Transfer 32 pp. 333-340. https://doi.org/10.1007/s002310050130

[27] Manzela, A.A., Hanriot, S.M., Cabezas-Gomez, L., and Sodre, J.R., 2010, "Using engine exhaust gas as energy source for an absorption refrigeration system," Appl. Energy, 87 (4) pp. 1141-1148. https://doi.org/10.1016/j.apenergy.2009.07.018

[28] Turns, S.R., and Kraige, D.R., 2007, "Properties Tables Booklet for Thermal Fluids Engineering," Cambridge University Press, New York.

[29] Khaliq, A., Habib, M.A., and Choudhary, K.A., 2019, "Thermo-environmental evaluation of a modified combustion gas turbine plant," ASME J. Energy Resour. Technol., 141(4) pp. 042004. https://doi.org/10.1115/1.4041898

[30] Heywood, J.B., 1998, "Internal Combustion Engines Fundamentals," McGraw Hill, New York.

[31] Steinhilber, T., and Sattelmayer, T., 2006, "The effect of water addition on HCCI diesel combustion," SAE Technical Paper No. 2006-01-3321.

[32] Saxena, S., Bedoya, I.D., Shah, N., and Phadke A., 2013, "Understanding loss mechanisms and identifying areas of improvement for HCCI engines using detailed exergy analysis," ASME J. Eng. Gas Turbines Power, 135 (9) pp. 091505. https://doi.org/10.1115/1.4024589

[33] Khaliq, A., 2017, "Energetic and exergetic performance investigation of a solar based integrated system for cogeneration of power and cooling," Appl. Therm. Eng., 112 pp. 1305-1316. https://doi.org/10.1016/j.applthermaleng.2016.10.127

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

[34] Seckin, C., 2019, "Effect of operational parameters on a novel combined cycle of ejector refrigeration cycle and Kalina cycle, ASME J. Energy Resour. Technol., 142 (1) pp. 012001. https://doi.org/10.1115/1.4044220

[35] Shokati, N., Ranjbar, F., and Yari, M. A., 2018, "Comprehensive exergoeconomic analysis of absorption power and cooling cogeneration cycles based on Kalina part I: Simulation," Energy Convers. Manage., 158 pp. 437-459. https://doi.org/10.1016/j.enconman.2017.12.086

[36] Dhahad, H.A., Hussen, H.M., Nguyen, P.T., Ghaebi, H., and Ashraf, M.A., 2020, "Thermodynamic and thermoeconomic analysis of innovative integration of Kalina and absorption refrigeration cycles for simultaneously cooling and power generation," Energy Convers. Manage., 203 pp. 112241. https://doi.org/10.1016/j.enconman.2019.112241

[37] Klein, S.A., 2012, "Engineering Equation Solver (EES) for Microsoft Windows operating systems: Academic professional version: F-chart software," Madison, W.I. (Available online http://www.fchart.com)

[38] REFPROP, 2013, "NIST Reference Thermodynamic and Transport Properties," version 9.1.

[39] Chua, H.T., Toh, H.K., Malek, A., Ng K.C., Srinivasan, K., 2000, "Improved thermodynamic property fields of LiBr-H_2O solutions," Int. J. Refrigeration, 23 (6), pp. 412-429. https://doi.org/10.1016/S0140-7007(99)00076-6

Accepted Manuscript Not Copyedited

FIGURE CAPTIONS:

Fig. 1: System layout of wet ethanol operated HCCI engine combined with ejector refrigeration cycle

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

Fig. 2: System layout of wet ethanol operated HCCI engine integrated to absorption refrigeration cycle

Fig. 3: P-V diagram for the HCCI engine whose thermodynamic cycle is modelled by a turbocharged Otto cycle

Fig. 4: Variation of energy efficiency of HCCI, HCCI-Ejector system, HCCI-Absorption system with turbocharger pressure ratio

Fig. 5: Variation of exergy efficiency of HCCI, HCCI-Ejector system, HCCI-Absorption system with turbocharger pressure ratio

Fig.6: Variation of energy efficiency of HCCI, HCCI-Ejector system, HCCI-Absorption system with turbocharger compressor efficiency

Fig.7: Variation of exergy efficiency of HCCI, HCCI-Ejector system, HCCI-Absorption system with turbocharger compressor efficiency

Fig.8: Variation of energy efficiency of HCCI, HCCI-Ejector system, HCCI-Absorption system with ambient temperature (K)

Fig.9: Variation of exergy efficiency of HCCI, HCCI-Ejector system, HCCI-Absorption system with ambient temperature (K)

Fig.10: Variation of energetic and exergetic COPs with entrainment ratio

Fig. 11: Variation of energetic and exergetic COPs with generator temperature (K)

Fig. 12: Distribution of fuel energy in the HCCI engine-ERC combined cycle of cogeneration

Fig. 13: Distribution of fuel energy in the HCCI engine-ARC combined cycle of cogeneration

Fig. 14: Distribution of fuel exergy in the combined cycle (HCCI engine-ERC) configuration

Fig. 15: Distribution of fuel exergy in the combined cycle (HCCI engine-ARC) configuration

TABLE CAPTIONS:

Table 1: Engine specifications [9]

Table 2: Component wise energy balance equations for the combined HCCI engine-ERC [18, 34]

Table 3: Equations of energy balance for the components of ARC [36]

Table 4: Component wise exergy balance equations for the combined HCCI engine-ERC [18, 33]

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

Table 5: Equations of exergy balance for the components of ARC cooling cycle [35]

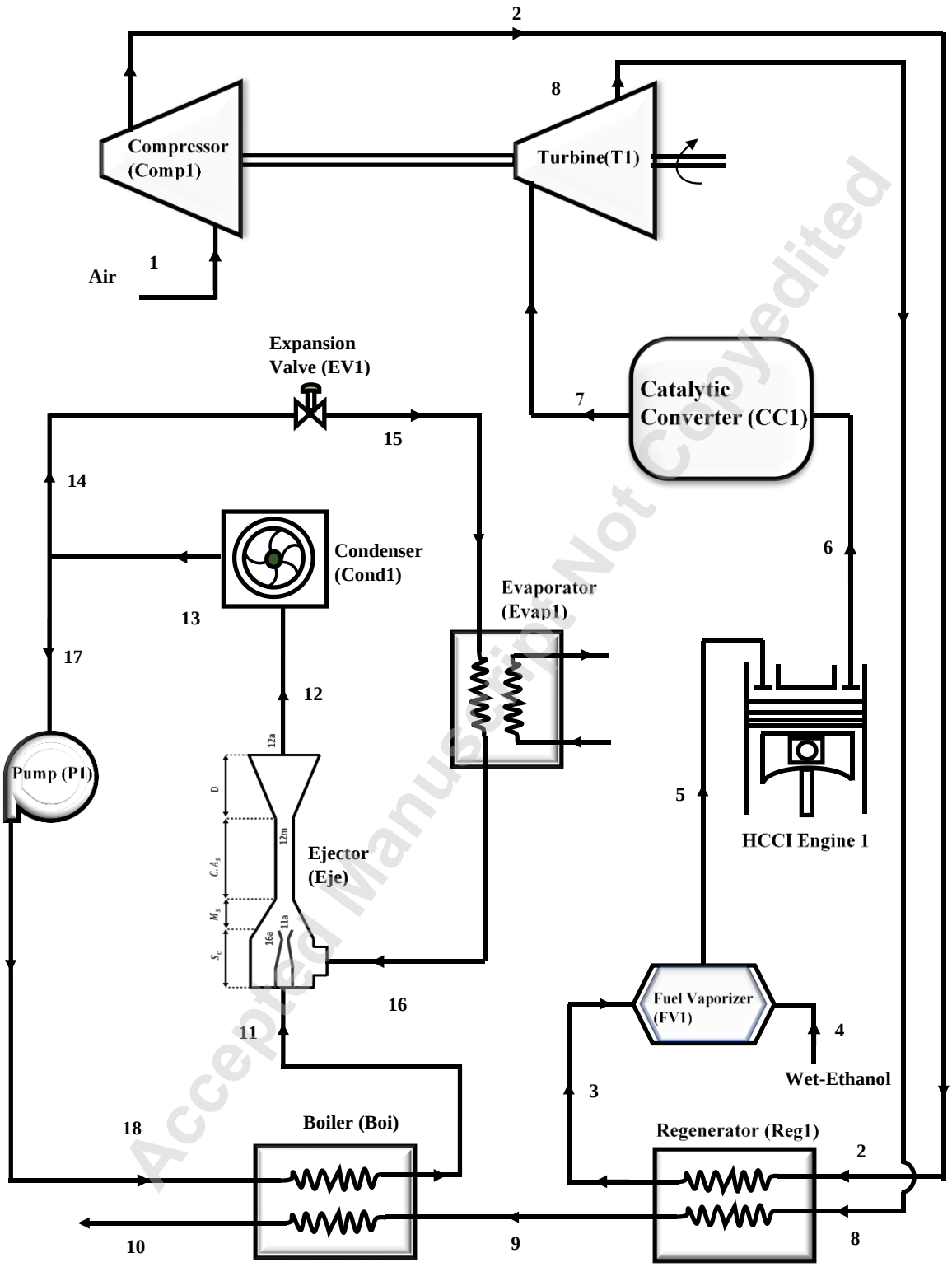


Fig. 1: System layout of wet ethanol operated HCCI engine integrated ejector refrigeration cycle

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

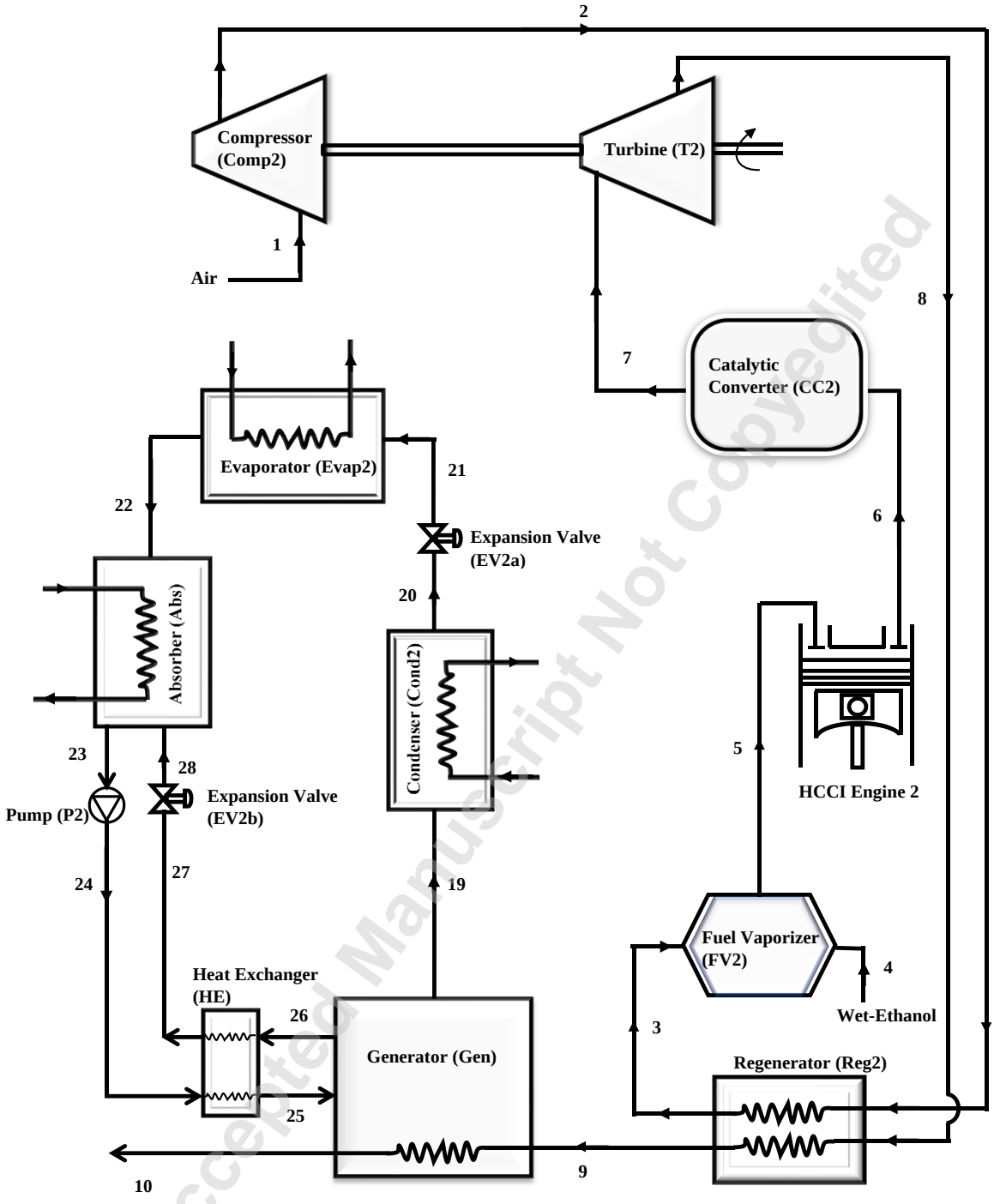


Fig. 2: System layout of wet ethanol operated HCCI engine integrated to absorption refrigeration cycle

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

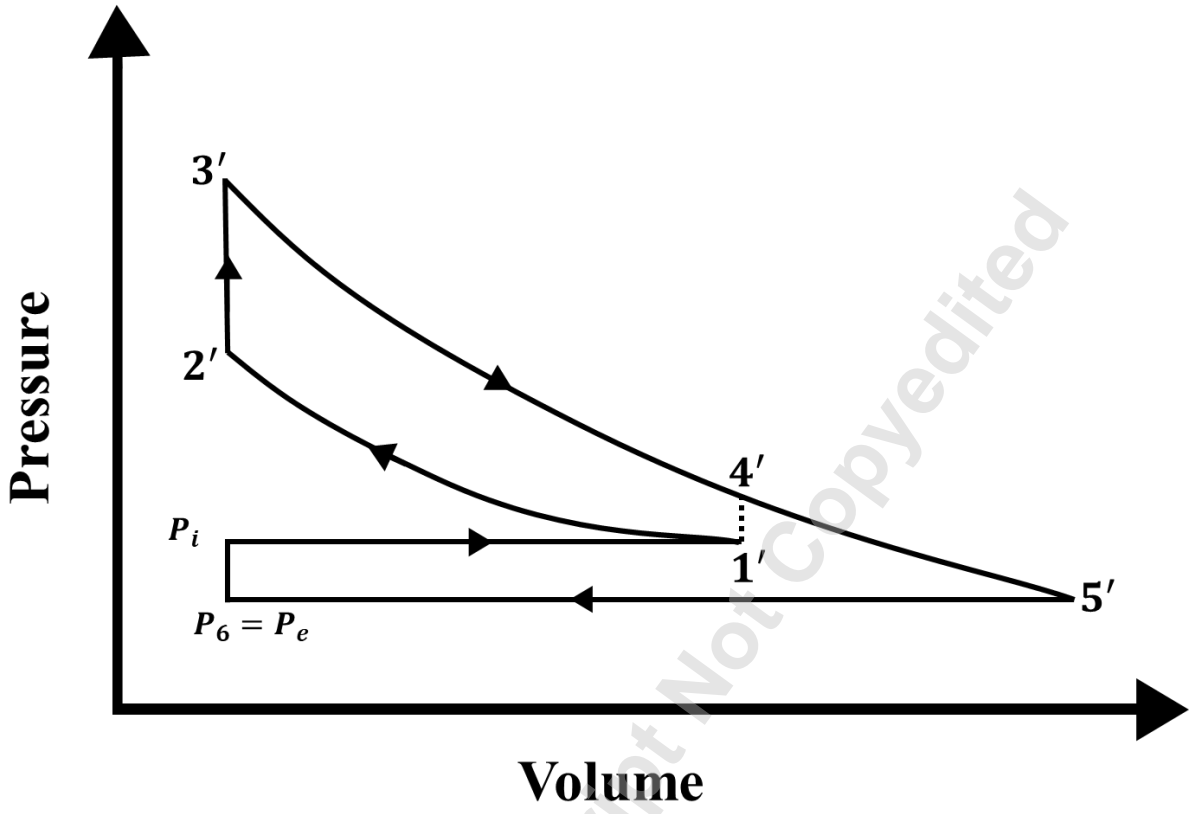


Fig. 3: P-V diagram for the HCCI engine whose thermodynamic cycle is modelled by a turbocharged Otto cycle

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

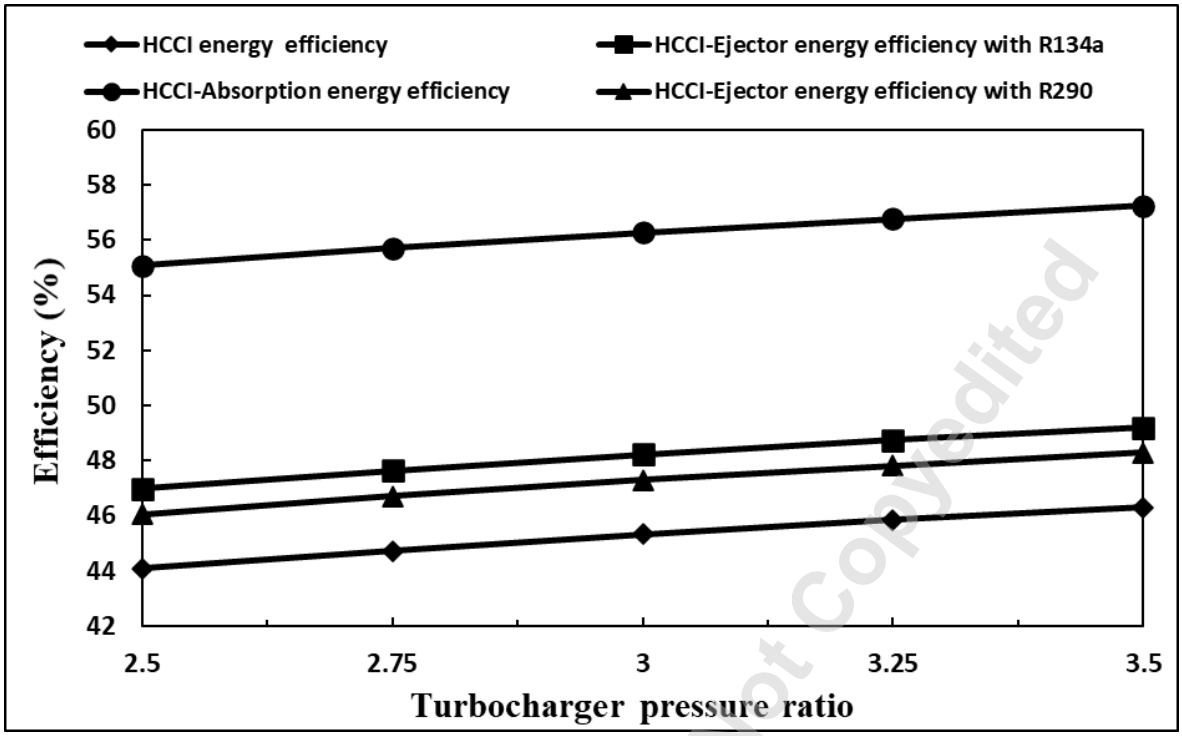
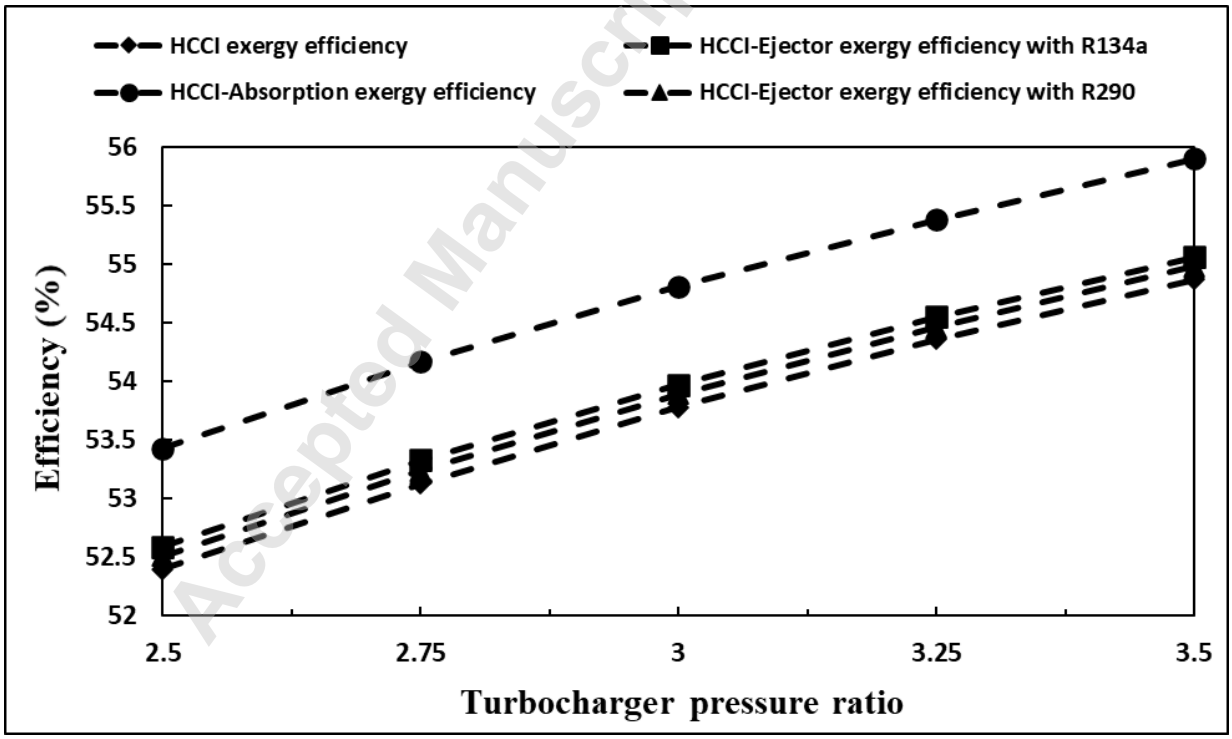


Fig. 4: Variation of energy efficiency of HCCI, HCCI-Ejector system, HCCI-Absorption system with turbocharger pressure ratio



Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

Fig. 5: Variation of exergy efficiency of HCCI, HCCI-Ejector system, HCCI-Absorption system with turbocharger pressure ratio

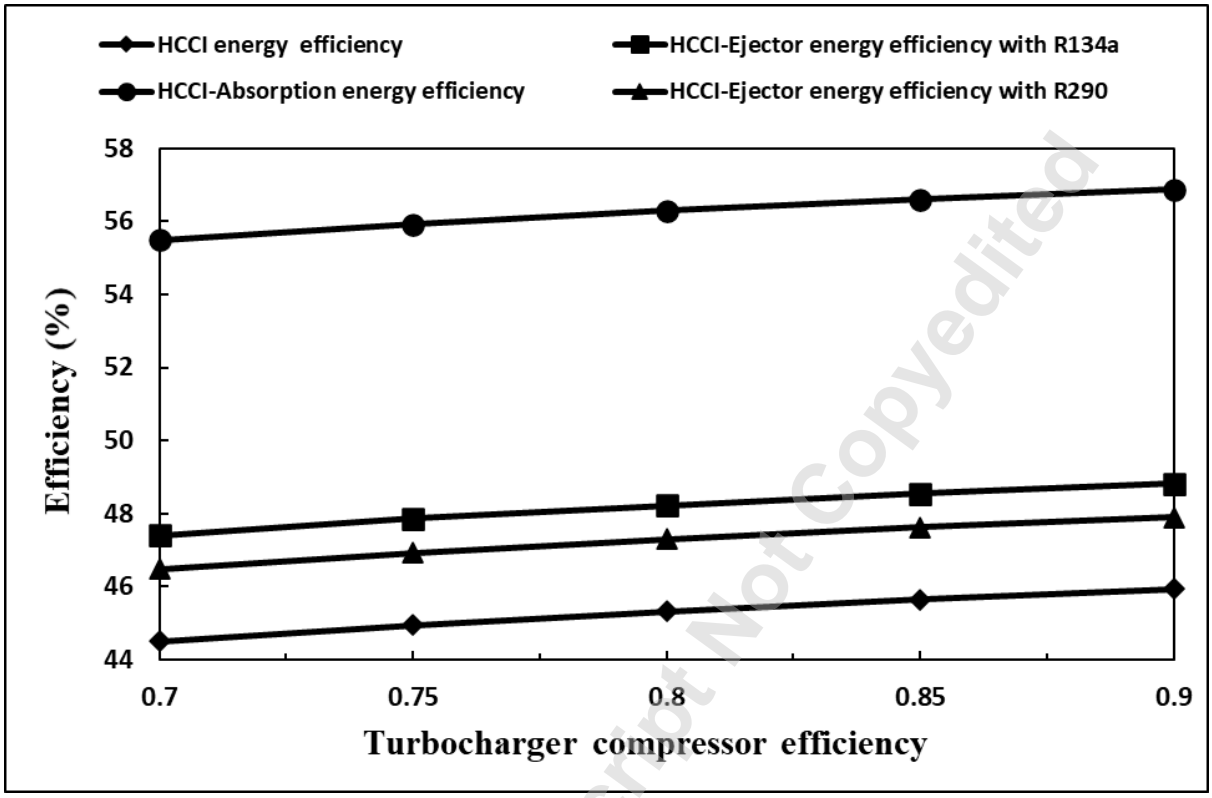


Fig.6: Variation of energy efficiency of HCCI, HCCI-Ejector system, HCCI-Absorption system with turbocharger compressor efficiency

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

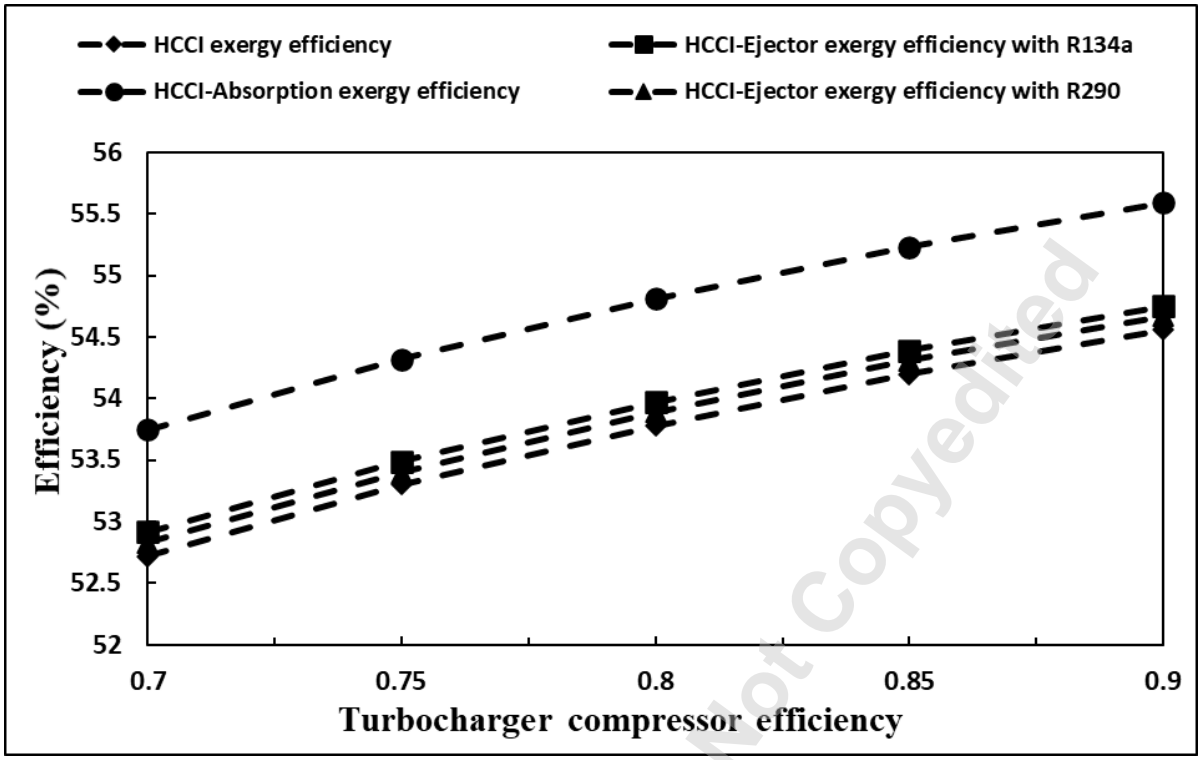
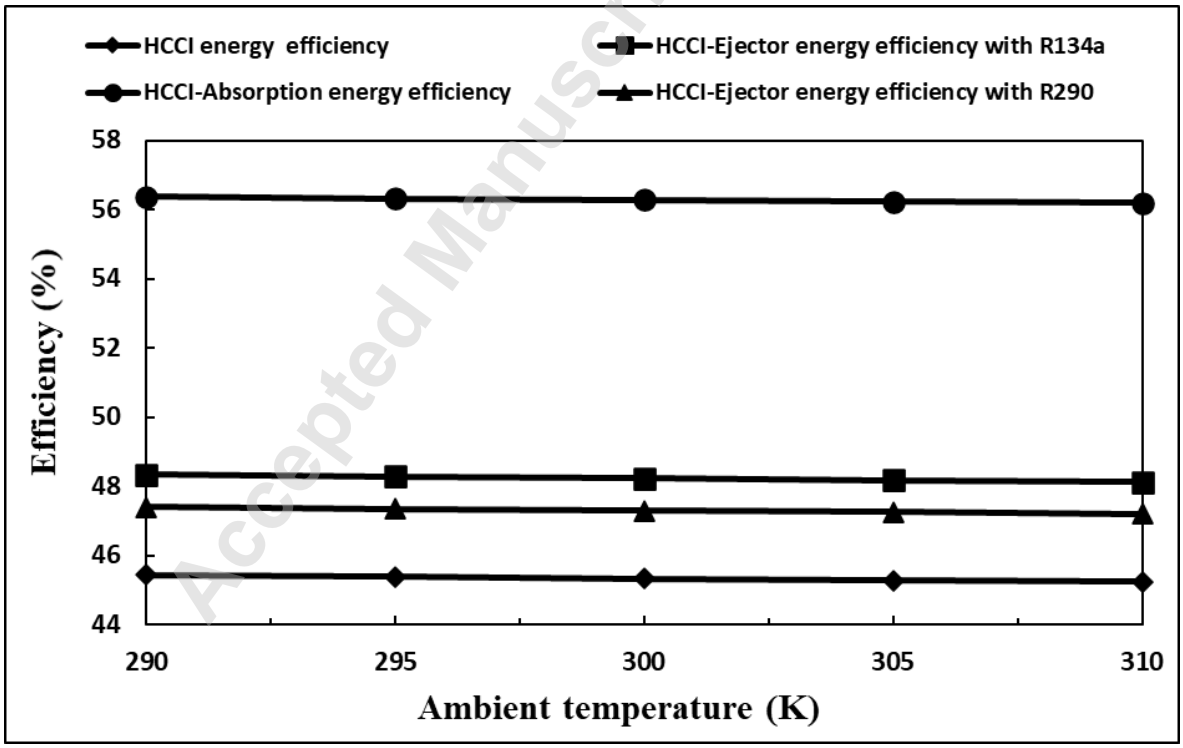


Fig.7: Variation of exergy efficiency of HCCI, HCCI-Ejector system, HCCI-Absorption system with turbocharger compressor efficiency



Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

Fig.8: Variation of energy efficiency of HCCI, HCCI-Ejector system, HCCI-Absorption system with ambient temperature (K)

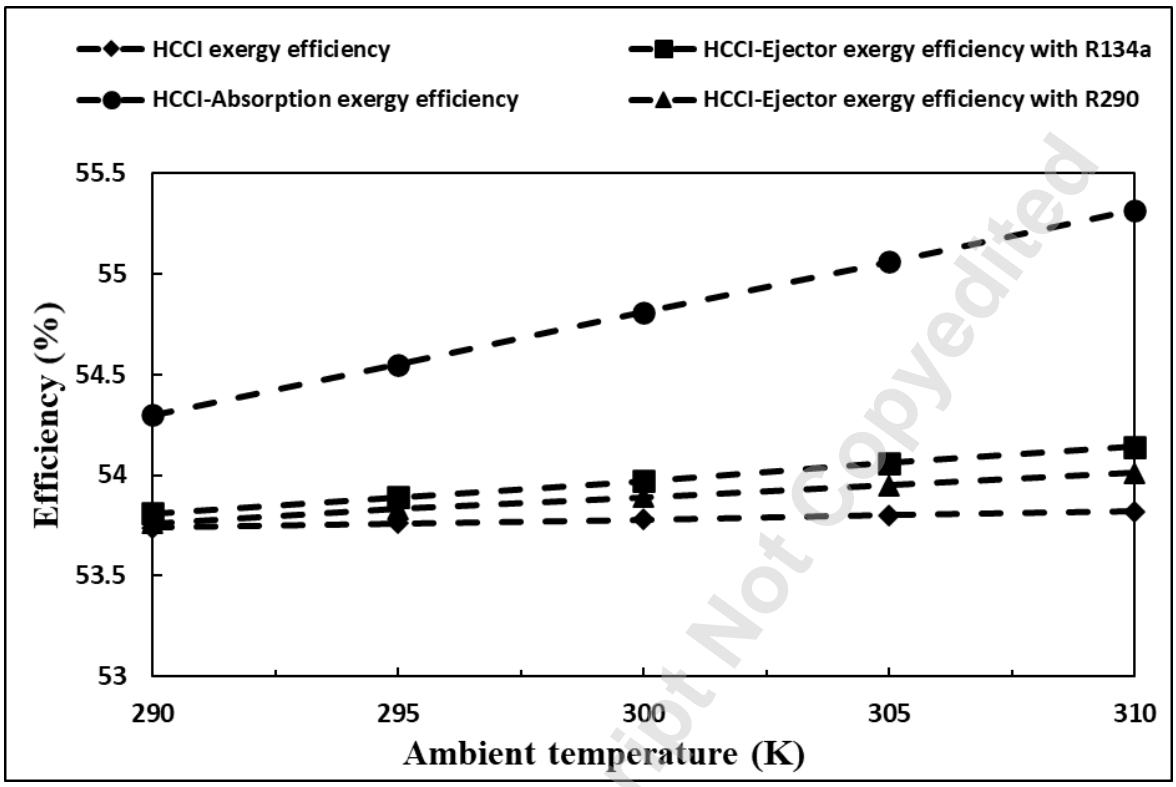


Fig.9: Variation of exergy efficiency of HCCI, HCCI-Ejector system, HCCI-Absorption system with ambient temperature (K)

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

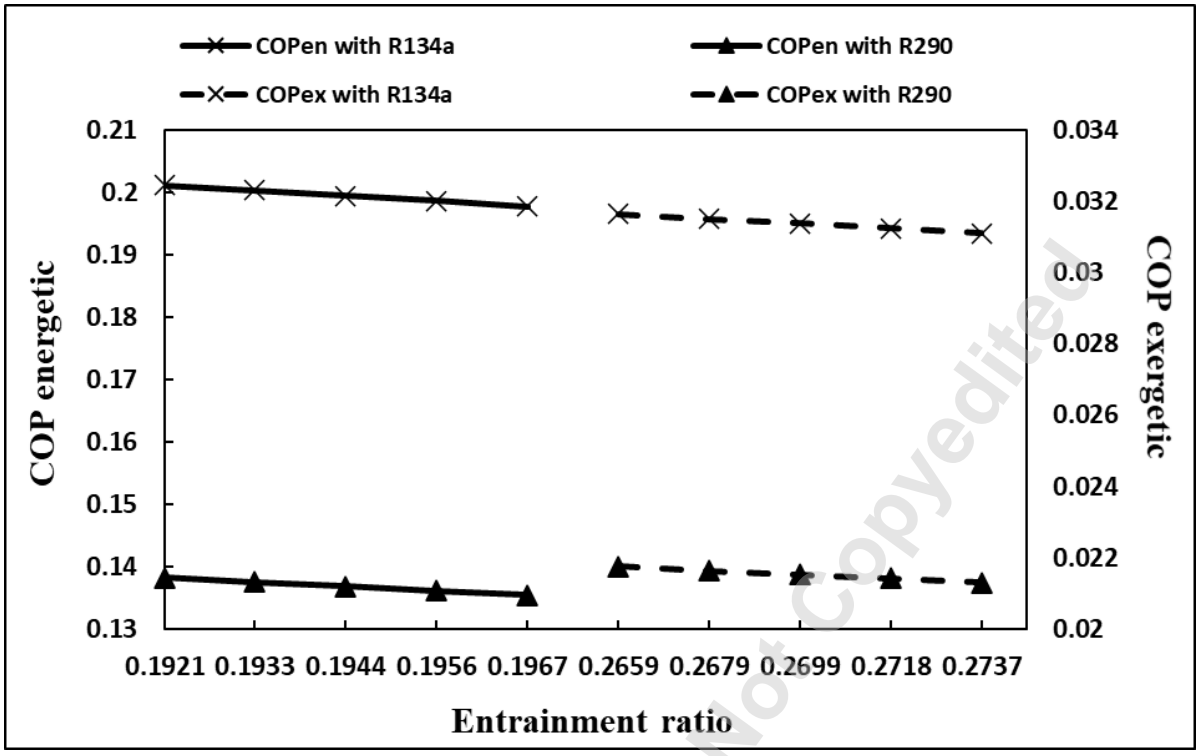


Fig.10: Variation of energetic and exergetic COPs with entrainment ratio

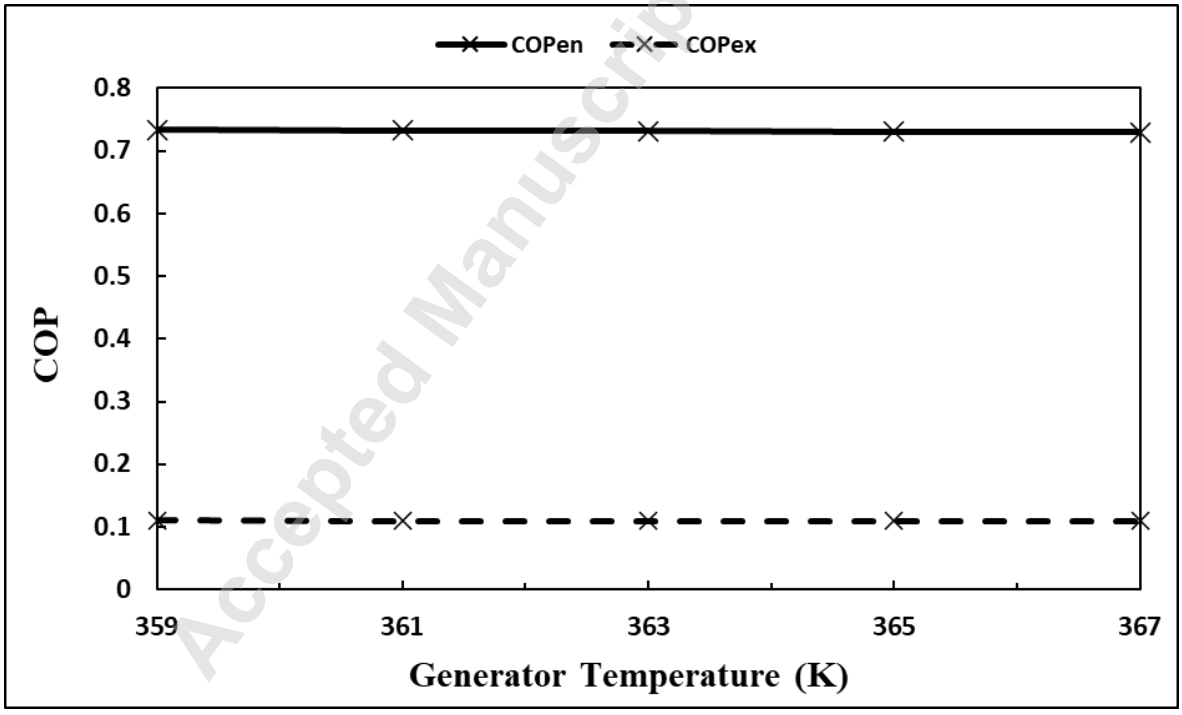


Fig. 11: Variation of energetic and exergetic COPs with generator temperature (K)

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

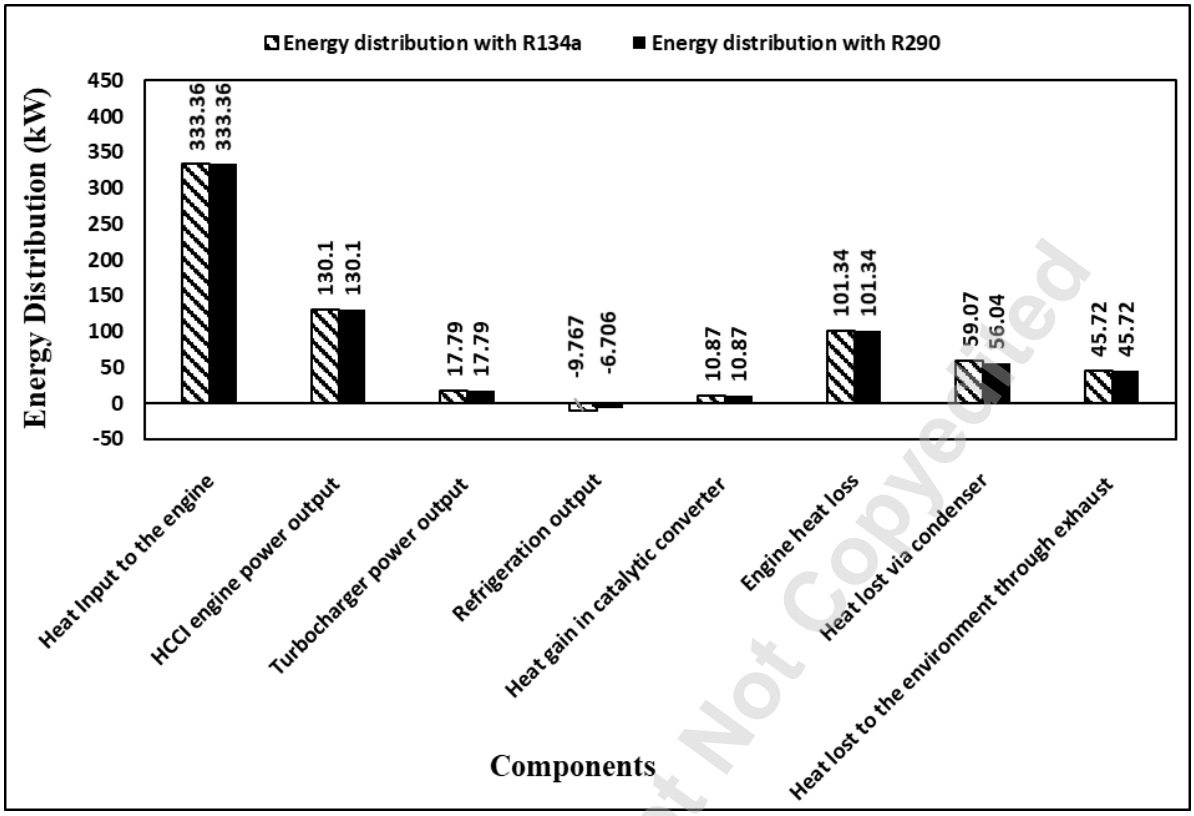


Fig. 12: Distribution of fuel energy in the HCCI engine-ERC combined cycle of cogeneration

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

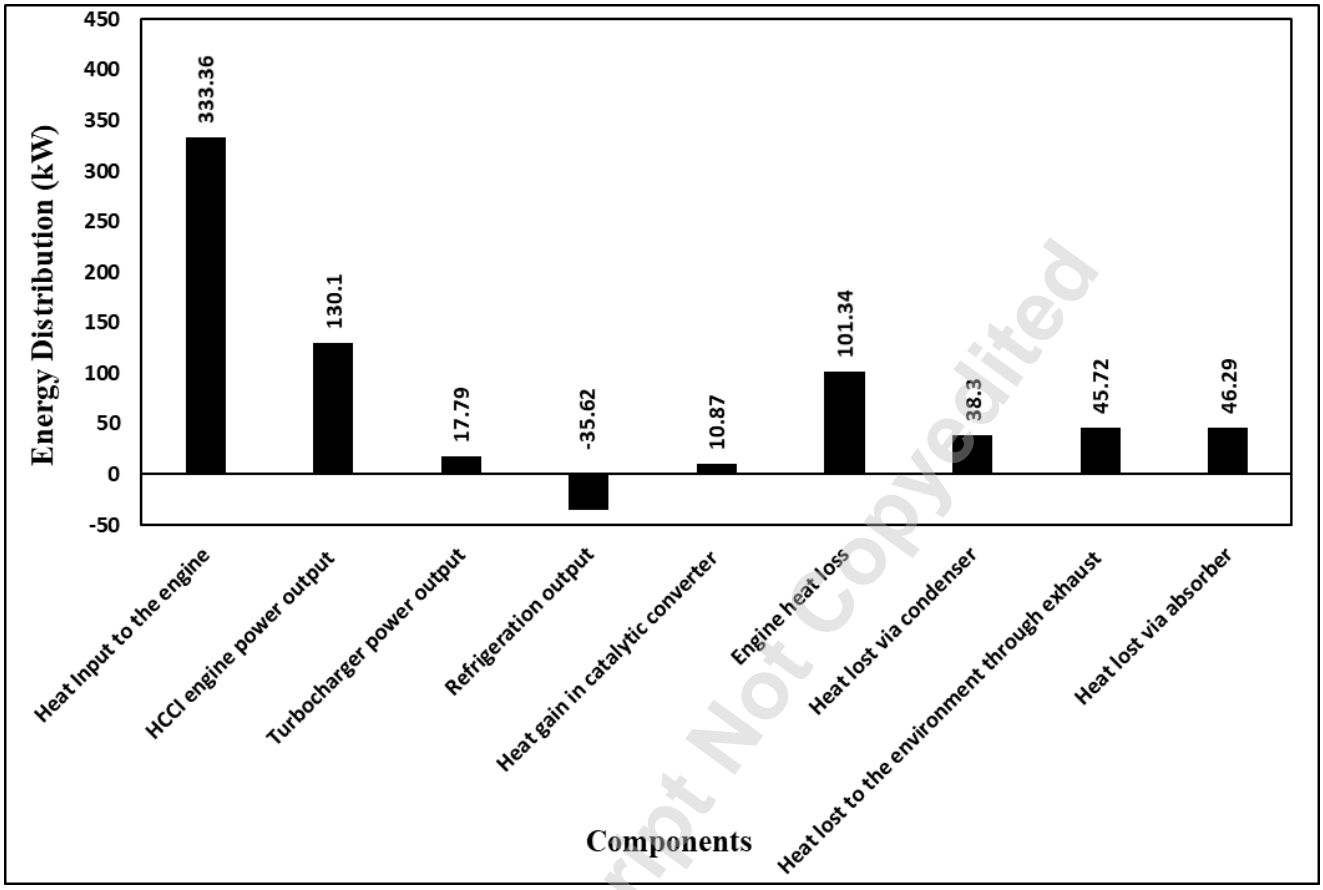


Fig. 13: Distribution of fuel energy in the HCCI engine-ARC combined cycle of cogeneration

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

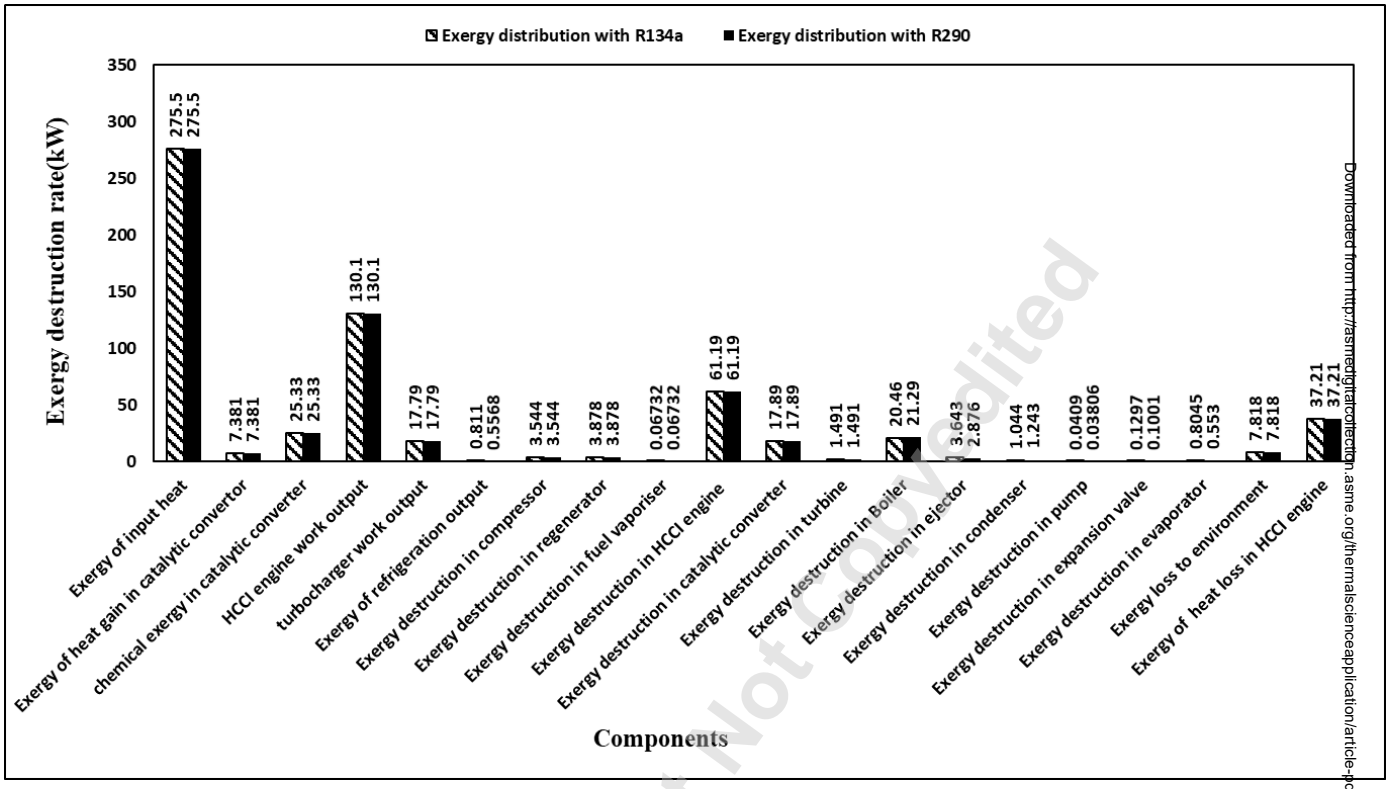
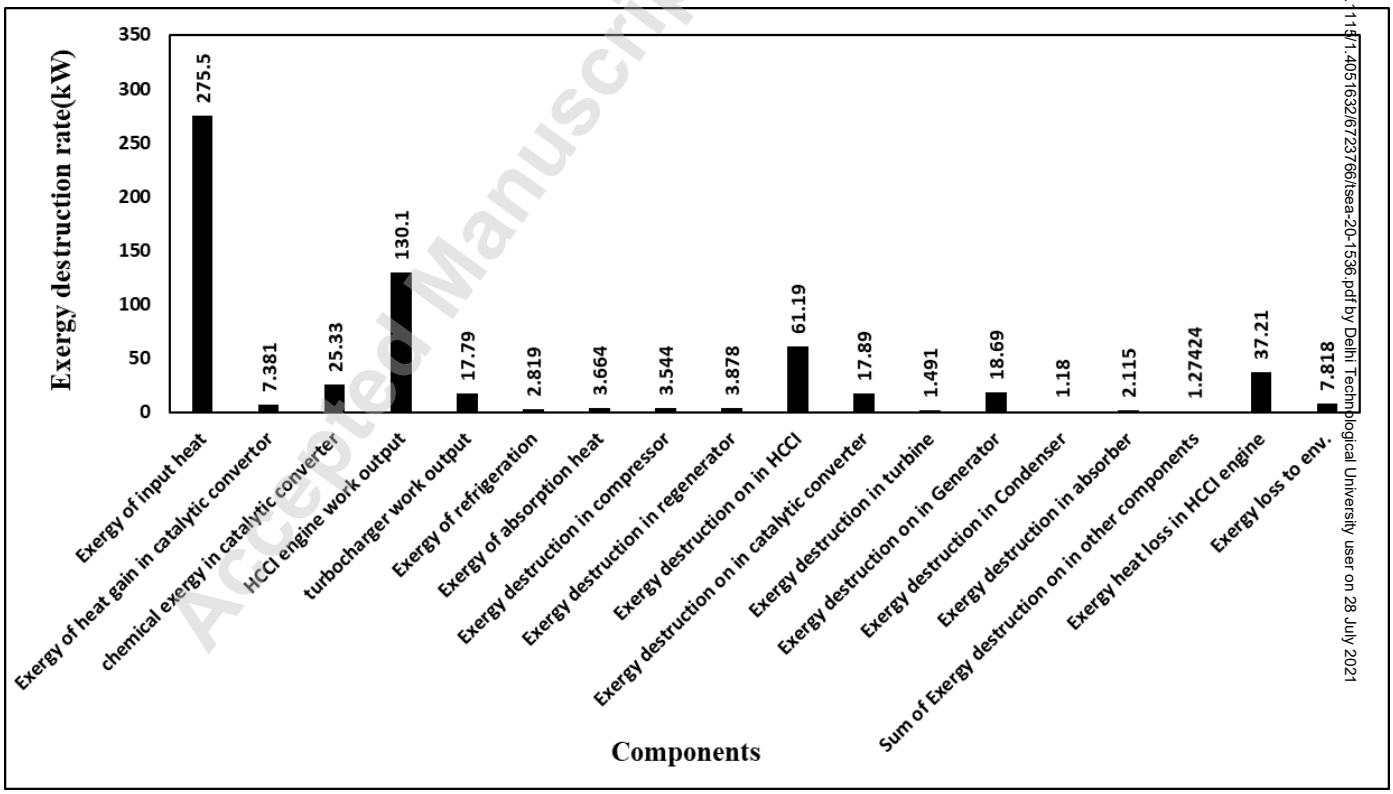


Fig. 14: Distribution of fuel exergy in the combined cycle (HCCI engine-ERC) configuration



Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

Fig. 15: Distribution of fuel exergy in the combined cycle (HCCI engine-ARC) configuration

Table 1: Engine specifications [9]

Volume swept per cylinder	2400 cm^3
Cylinder diameter	13.7 cm
Piston displacement	16.5 cm
Length of connecting rod	26.2 cm
Ratio of intake to clearance volume	16:1
Speed of engine	30 rps
Engine volumetric efficiency	100%
Ratio of actual fuel-air mixture to chemically correct fuel-air mixture	<0.5
Combustion process in HCCI engine	Chemical kinetics applied to control the bulk auto ignition process

Table 2: Component wise energy balance equations for the combined HCCI engine-ERC [18, 34]

Component	**Balance Equations**
Compressor (Comp1)	$\dot{W}_{Comp1} = \dot{m}_a(h_2 - h_1)$
Regenerator (Reg1)	$\dot{m}_a\,(h_3 - h_2) = \dot{m}_{exh}\,(h_8 - h_9)$
Fuel vaporizer (FV1)	$\dot{m}_a\,h_3 + \dot{m}_f h_4 = (\dot{m}_a\, + \dot{m}_f)h_5$
HCCI engine (Compression) process,1′- 2′	$(\dot{m}_a\, + \dot{m}_f + \dot{m}_{rg})(h_{2'} - h_{1'}) = \dot{W}_{comp} - \dot{Q}_{surr}$

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

Heat addition process, 2′- 3′	$(\dot{m}_a + \dot{m}_f + \dot{m}_{rg})(h_{3'} - h_{2'}) = \dot{Q}_H$
Expansion process,3′- 4′	$(\dot{m}_a + \dot{m}_f + \dot{m}_{rg})(h_{3'} - h_{4'}) = \dot{W}_{exp} + \dot{Q}_{surr}$
Blowdown Process, 4′-5′	$\dot{m}_{exh}\, h_{4'} = \dot{m}_{exh}\, h_{5'} + \dot{Q}_{surr}$
Catalytic convertor (CC1)	$\sum_{Product} n_e\left(\bar{h}_f^{\circ} + \Delta\bar{h}\right)_e = \sum_{Reactant} n_i\left(\bar{h}_f^{\circ} + \Delta\bar{h}\right)_i$
Turbine (T1)	$\dot{W}_{T1} = \dot{m}_{exh}\,(h_7 - h_8)$
Boiler (Boi)	$\dot{Q}_{boil} = \dot{m}_{11}(h_{11} - h_{18})$
Pump (P1)	$\dot{W}_{P1} = \dot{m}_{18}\,(h_{18} - h_{17})$
Ejector (Eje)	$\dot{m}_{11}h_{11} + \dot{m}_{16}h_{16} = \dot{m}_{12}h_{12}$
Condenser (Cond1)	$\dot{m}_{12}(\,h_{12} - \,h_{13}) = \dot{m}_{Cond1}(\,h_{Cond1,o} - \,h_{Cond1,i})$
Expansion valve (EV1)	$h_{14} = h_{15}$
Evaporator (Evap1)	$\dot{Q}_{Evap1} = \dot{m}_{16}\,(h_{16} - h_{15})$

Table 3*:* Equations of energy balance for the components of ARC [36]

Component	**Balance Equations**
Generator (Gen)	$\dot{m}_9 h_9 + \dot{m}_{25}h_{25} = \dot{m}_{10}h_{10} + \dot{m}_{ref}h_{19} + (\dot{m}_{sol} - \dot{m}_{ref})\,h_{26}$
Condenser (Cond2)	$\dot{m}_{ref}(h_{19} - \,h_{20}) = \dot{m}_{Cond2}(h_{Cond2,o} - \,h_{Cond2,i})$
Expansion Valve (EV2a)	$h_{20} = \,h_{21}$
Evaporator (Evap2)	$\dot{Q}_{Evap2} = \dot{m}_{ref}(h_{22} - h_{21})$
Expansion Valve (EV2b)	$h_{28} = h_{27}$
Absorber (Abs)	$(\dot{m}_{sol} - \dot{m}_{ref})h_{28} + \dot{m}_{Abs}h_{Abs,i} + \dot{m}_{ref}h_{22} + = \dot{m}_{sol}h_{23} + \dot{m}_{Abs}h_{Abs,o}$

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

Pump (P2)	$\dot{W}_{P2} = \dot{m}_{sol}(h_{24} - h_{23})$
Heat Exchanger (HE)	$\dot{m}_{sol}h_{24} + (\dot{m}_{sol} - \dot{m}_{ref})h_{26} = (\dot{m}_{sol} - \dot{m}_{ref})h_{27} + \dot{m}_{sol}h_{25}$

Table 4: Component wise exergy balance equations for the combined HCCI engine-ERC [18, 33]

Component	**Balance Equations**
Compressor (Comp1)	$\dot{E}_{x,D,Comp1} = \dot{m}_1(e_{x,1} - e_{x,2}) + \dot{W}_{Comp1}$
Regenerator (Reg1)	$\dot{E}_{x,D,Reg1} = \dot{m}_2(e_{x,2} - e_{x,3}) + \dot{m}_8(e_{x,8} - e_{x,9})$
Fuel vaporizer (FV1)	$\dot{E}_{x,D,FV1} = \dot{m}_3 e_{x,3} + \dot{m}_4 e_{x,4} - \dot{m}_5 e_{x,5}$
HCCI engine (Compression) process,1′-2′	$\dot{E}_{x,D,Comp} = \dot{m}_5(e_{x,1'} - e_{x,2'}) + \dot{W}_{comp} - \dot{Q}_{surr}\left(1 - \frac{T_0}{T}\right)$
Heat addition process, 2′-3′	$\dot{E}_{x,D,Heat\ addition} = \dot{m}_5(e_{x,2'} - e_{x,3'}) + \dot{Q}_H\left(1 - \frac{T_0}{T_H}\right)$
Expansion process, 3′- 4′	$\dot{E}_{x,D,exp} = \dot{m}_5(e_{x,3'} - e_{x,4'}) - \dot{W}_{exp} - \dot{Q}_{surr}\left(1 - \frac{T_0}{T}\right)$
Blowdown Process, 4′-5′	$\dot{E}_{x,D,exh\ \&\ blowdown} = \dot{m}_5(e_{x,4'} - e_{x,5'}) - \dot{Q}_{surr}\left(1 - \frac{T_0}{T}\right)$
Catalytic convertor (CC1)	$\dot{E}_{x,D,CC1} = \dot{m}_6 e_{x,6} - \dot{m}_7 e_{x,7} + \dot{m}_6 e_{ch}$
Turbine (T1)	$\dot{E}_{x,D,T1} = \dot{m}_7(e_{x,7} - e_{x,8}) - w_{T1}$
Boiler (Boi)	$\dot{m}_{18}ex_{18} + \dot{Q}_{boi}\left(1 - \frac{T_0}{T_{ave,boi}}\right) = \dot{m}_{11}ex_{11} + \dot{m}_{boi}ex_{d,boi}$
Pump (P1)	$\dot{E}_{x,D,P1} = \dot{m}_{17}(e_{x,17} - e_{x,18}) + \dot{W}_{P1}$
Ejector (Eje)	$\dot{E}_{x,D,Eje} = \dot{m}_{11}e_{x,11} + \dot{m}_{16}e_{x,16} - \dot{m}_{12}e_{x,12}$
Condenser (Cond1)	$\dot{E}_{x,D,Cond1} = \dot{m}_{12}e_{x,12} + \dot{m}_{Cond1}(e_{x,Cond1,i} - e_{x,Cond1,o}) - \dot{m}_{13}e_{x,13}$
Expansion valve (EV1)	$\dot{E}_{x,D,EV1} = \dot{m}_{14}(e_{x,14} - e_{x,15})$
Evaporator (Evap1)	$\dot{E}_{x,D,Evap1} = \dot{m}_{15}(e_{x,15} - e_{x,16}) + \dot{m}_{Evap1}(e_{x,Evap1,i} - e_{x,Evap1,o})$

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

Journal of Thermal Science and Engineering Applications. Received September 02, 2020;
Accepted manuscript posted June 20, 2021. doi:10.1115/1.4051632
Copyright © 2021 by ASME

Table 5: Equations of exergy balance for the components of ARC cooling cycle [35]

Component	Balance Equations
Generator (Gen)	$\dot{E}_{x,D,Gen} = \dot{m}_9(e_{x,9} - e_{x,10}) + \dot{m}_{sol}e_{x,25} - (\dot{m}_{sol} - \dot{m}_{ref})e_{x,26} - \dot{m}_{ref}e_{x,19}$
Condenser (Cond2)	$\dot{E}_{x,D,Cond2} = \dot{m}_{ref}(e_{x,19} - e_{x,20}) + \dot{m}_{Cond2}(e_{x,Cond2,i} - e_{x,Cond2,o})$
Expansion Valve (EV2a)	$\dot{E}_{x,D,EV,abs} = \dot{m}_{ref}(e_{x,20} - e_{x,21})$
Evaporator (Evap2)	$\dot{E}_{x,D,Evap2} = \dot{m}_{ref}e_{x,21} + \dot{m}_{Evap2}e_{x,Evap2,i} - \dot{m}_r e_{x,22} - \dot{m}_{Evap2}e_{x,Evap2,o}$
Expansion Valve (EV2b)	$\dot{E}_{x,D,EV2b} = (\dot{m}_{sol} - \dot{m}_r)(e_{x,27} - e_{x,28})$
Absorber (Abs)	$\dot{E}_{x,D,ab} = (\dot{m}_{sol} - \dot{m}_{ref})e_{x,28} + \dot{m}_{ref}e_{x,22} + \dot{m}_{Abs}e_{x,Abs,i} - \dot{m}_{sol}e_{x,23} - \dot{m}_{Abs}e_{x,Abs,o}$
Pump (P2)	$\dot{E}_{x,D,P2} = \dot{m}_{sol}(e_{x,23} - e_{x,24}) + \dot{W}_{P2}$
Heat Exchanger (HE)	$\dot{E}_{x,D,HE} = (\dot{m}_{sol} - \dot{m}_{ref})(e_{x,26} - e_{x,27}) + \dot{m}_{sol}(e_{x,24} - e_{x,25})$

Accepted Manuscript Not Copyedited

Downloaded from http://asmedigitalcollection.asme.org/thermalscienceapplication/article-pdf/doi/10.1115/1.4051632/6723766/tsea-20-1536.pdf by Delhi Technological University user on 28 July 2021

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/352836140

Thidiazuron outpaces 6-benzylamino purine and kinetin in delaying flower senescence in Gladiolus grandiflora by alleviating physiological and biochemical responses

Article *in* Journal of Applied Biology & Biotechnology · June 2021

DOI: 10.7324/JABB.2021.9407

CITATIONS
0

READS
35

2 authors, including:

Madhulika Singh
Swami Shraddhanand College, University of Delhi
2 PUBLICATIONS **1** CITATION

SEE PROFILE

All content following this page was uploaded by Madhulika Singh on 30 June 2021.

The user has requested enhancement of the downloaded file.

Journal of Applied Biology & Biotechnology Vol. 0(00), pp. 1-7, June, 2021
Available online at http://www.jabonline.in
DOI: 10.7324/JABB.2019.70103

Thidiazuron outpaces 6-benzylamino purine and kinetin in delaying flower senescence in *Gladiolus grandiflora* by alleviating physiological and biochemical responses

Madhulika Singh[1*], Neha Tiwari[2]

[1]Department of Botany, Swami Shraddhanand College, University of Delhi, Delhi, India
[2]Department of Biotechnology, Delhi Technological University, Delhi, India

ARTICLE INFO

Article history:
Received on: April 16, 2021
Accepted on: May 25, 2021
Available online: XX

Key words:
Gladiolus grandiflora, lipid peroxidation, lipoxygenase, membrane stability index, Senescence, vase life

ABSTRACT

An experiment was carried out to elucidate the role of thidiazuron (TDZ), 6-benzylamino purine (BAP), and kinetin on the vase life of *Gladiolus grandiflora* cut flower. Harvested flowers were treated with BAP, kinetin, and TDZ of various concentrations, viz., 100 µM, 0.5 mM, and 1 mM. Our findings revealed that there was a significant increase in flower longevity in treatment with BAP, kinetin, and TDZ as compared to untreated flowers. Vase solutions which contain BAP (0.5 mM), kinetin (1 mM), and TDZ (100 µM) were most effective in improving vase life (14, 13, and 15 days, respectively) of cut *Gladiolus* floral spike. Flower vase life was extended with increase in fresh weight, vase solution uptake, high membrane integrity, besides reduced pH of vase solution, malondialdehyde content, and lipoxygenase activity in the flowers. However, TDZ outpaced BAP and kinetin in improving the vase life of *G. grandiflora*.

1. INTRODUCTION

Flowers are the excellent model system to unveil the physiological and biochemical processes during senescence as the senescence process in flowers is rapid and foreseeable. Senescence is defined as an age-dependent degenerative process in plants leading to death [1]. The senescence process involves halt of various biochemical pathways and up-regulation of many degradative pathways, finally leading to the cell, tissue, or whole plant death [2]. Genetically, senescence is determined and it is governed by the various plant hormones and environmental factors during plant development process. During senescence process, a number of changes such as reduced fresh weight, vase solution uptake, loss of membrane integrity, increase in lipid peroxidation, production of reactive oxygen species (ROS), increase in membrane lipid peroxidation, degradation of fatty acids, proteins, DNA, RNA, and sugars are induced [3–5]. Morphological changes which are visible during petal senescence are change of petal color and rolling of petals [6], and anatomical changes include degradation of mesophyll cells before morphological changes are apparent [7,8].

It is the life span and quality of the cut flower that determine the commercial value of flowers [9]. To increase flower life span, it is important to understand physiological and biochemical processes associated with petal senescence [10]. Cytokinins act as senescence retarding hormones that increase the life span in both the ethene-sensitive and ethene-insensitive flowers [1,11]. Because of cytokinin delays senescence process, it is used in postharvest technology to enhance the postharvest life of flowers [12]. Cytokinin delays flower senescence by increasing uptake of water, fresh weight sink activity, metabolites contents in flower petals, besides preventing membrane lipid peroxidation [11,13]. Cytokinin both adenine type 6-benzylamino purine (BAP, kinetin) and diphenyl urea derived thidiazuron (TDZ) [14] retard senescence process by increasing the content of various metabolites, vase solution uptake, besides preventing peroxidation of membrane lipids [12]. Exogenous application of cytokinin has important

*Corresponding Author
Madhulika Singh, Department of Botany, Swami Shraddhanand College, University of Delhi, Delhi, India. E-mail: madhulikasingh@ss.du.ac.in

© 2021 Singh and Tiwari. This is an open access article distributed under the terms of the Creative Commons Attribution License -NonCommercial-ShareAlike Unported License (http://creativecommons.org/licenses/by-nc-sa/3.0/).

role in enhancing vase life of flowers like lotus, Gerbera, Iris and daylily [15–17]. Although both classes of cytokinin are effective in enhancing vase life, however, the role of TDZ is more promising in cut flowers.*Gladiolus grandiflora* "*White Prosperity"* cultivar is a mid-season bloomer which is grown in India as an ornamental plant and has a great role in floriculture market [18]. However, short span of vase life of flower reduces its potential commercial value. *Gladiolus* is ethylene insensitive; therefore, the reduced vase life is because of oxidative stress [19] and role of cytokinin's in ameliorating oxidative stress during flower senescence in *G. grandiflora* "*White Prosperity" cultivar* has not been explored yet. In this respect, an experiment was depicted to study the effect of BAP, kinetin, and TDZ in alleviating flower senescence in *G. grandiflora* by assessing physiological and biochemical parameters to improve its postharvest life span.

2. MATERIALS AND METHODS

All the experiments were performed in the Department of Botany, SSN College, University of Delhi.

2.1. Plant Material

Gladiolus grandiflora, i.e., "*White Prosperity." Gladiolus* spikes used in the experiment were collected from Indian Agricultural Research Institute, New Delhi.

2.2. Experimental Treatments

Uniform length (15 cm) of the floral spikes was cut and it is divided into four sets. Three sets of spikes were subjected to different concentrations of BAP, kinetin, and TDZ, viz., 100 µM, 0.5 mM, 1 mM and another set of floral spikes were kept in Milli Q and referred as control. Each treatment was represented by three replicates and the experiment was performed at temperature of 22°C ± 3°C and relative humidity 67% ±_3% and light intensity 1000 lux.

2.3. Estimation of Vase Life

The flower spikes fresh weight was determined on a daily basis till the end of the vase life of flower. The vase life of floral spikes was defined as when the fresh weight of floral spikes was less than its initials fresh weight.

2.4. Vase Solution Uptake Estimation

The floral spikes were kept undisturbed in the test tubes containing 25 ml of solution. The uptake of vase solution by spikes was measured and expressed as cumulative uptake of vase solution in ml/spike.

2.5. pH of Vase Solution

The vase solution pH was analyzed using the pH meter (HACH analyzer, HQ440d multi).

2.6. Membrane Stability Index (MSI)

The MSI of *Gladiolus* flower petals was measured using the experimental protocol of Bailly *et al.* [20]. Petal tissue (200 mg) was immersed in the tube containing 10 ml Milli Q water. After 5 hours of incubation, the electrical conductivity (C_1) of the solution was determined using conductometer (HACH analyzer, HQ440d multi). The petal tissue was then heated for 30 minutes and the conductivity (C_2) was measured.

$$MSI = [1-(C_1/C_{-2})] \times 100$$

2.7. Malondialdehyde (MDA) Content

The lipid peroxidation was measured in terms of MDA content by following the experimental protocol of Heath and Packer [21]. Petal (500 mg) was homogenized in 0.1% trichloroacetic acid (TCA). The homogenate was centrifuged for 5 minutes at 15,000 rpm (Beckman refrigerated centrifuge, Model J2-21). To the supernatant, 4 ml of 0.5% thiobarbituric acid in 20% TCA was added [21]. The content was boiled for 30 minutes and then quickly cooled. Then centrifuged for 10 minutes at 10,000 rpm (Beckman refrigerated centrifuge, Model J2-21). The absorbance data of the supernatant was recorded at 532 nm using Eppendorf UV-VIS Spectrophotometer, 5810 R. The MDA content of the sample was calculated using its absorption coefficient of 155 m mol^{-1} cm^{-1}.

2.8. Lipoxygenase (LOX) Activity

The LOX enzyme activity was measured using the experimental protocol given by Doderer *et al.* [22]. Flower petals (500 mg) were homogenized in phosphate buffer solution (0.1 M) containing ethylenediaminetetraacetic acid (EDTA) (0.5 mM). The homogenate was centrifuged (Beckman refrigerated centrifuge, Model J2-21) for 15 minutes at 15,000 rpm. The supernatant was referred as enzyme extract. LOX activity was measured by adding enzyme (50 µl) to substrate solution made by adding 70 µl linoleic acid to 10 ml Milli Q having 100 µl tween 20, final volume adjusted to 200 ml using 0.1 M phosphate buffer. Absorbance data was recorded at 234 nm and the enzyme activity was measured as change in ΔA_{234} $minutes^{-1}$ g^{-1} protein [22].

2.9. Statistical Analysis

We used a completely randomized design with three replicates. The experimental data was examined using Statistical Package for the Social Sciences (SPSS) 21 Statistical program (IBM SPSS Statistics 21) by one way analysis analysis of variance. Comparison of the means was determined by post hoc Duncan's test ($p < 0.05$).

3. RESULTS

The standardization of BAP, kinetin, and TDZ concentration in extending postharvest life of cut flowers of *G. grandiflora* was carried out. Various concentrations of BAP, kinetin, and TDZ, viz. 100 µM, 0.5 mM, and 1 mM were used. Application of 0.5 mM BAP and 1 mM kinetin and 100 µM TDZ exhibited maximum enhancement of vase life of *Gladiolus* (Fig. 1) due to increase in various physiological parameters (fresh weight of flowers, uptake of vase solution, membrane integrity) and decrease in biochemical parameters (lipid peroxidation and LOX activity). However, among all the cytokinin treatments (BAP, kinetin, and TDZ), TDZ was found as most effective in enhancing the vase life of *Gladiolus*.

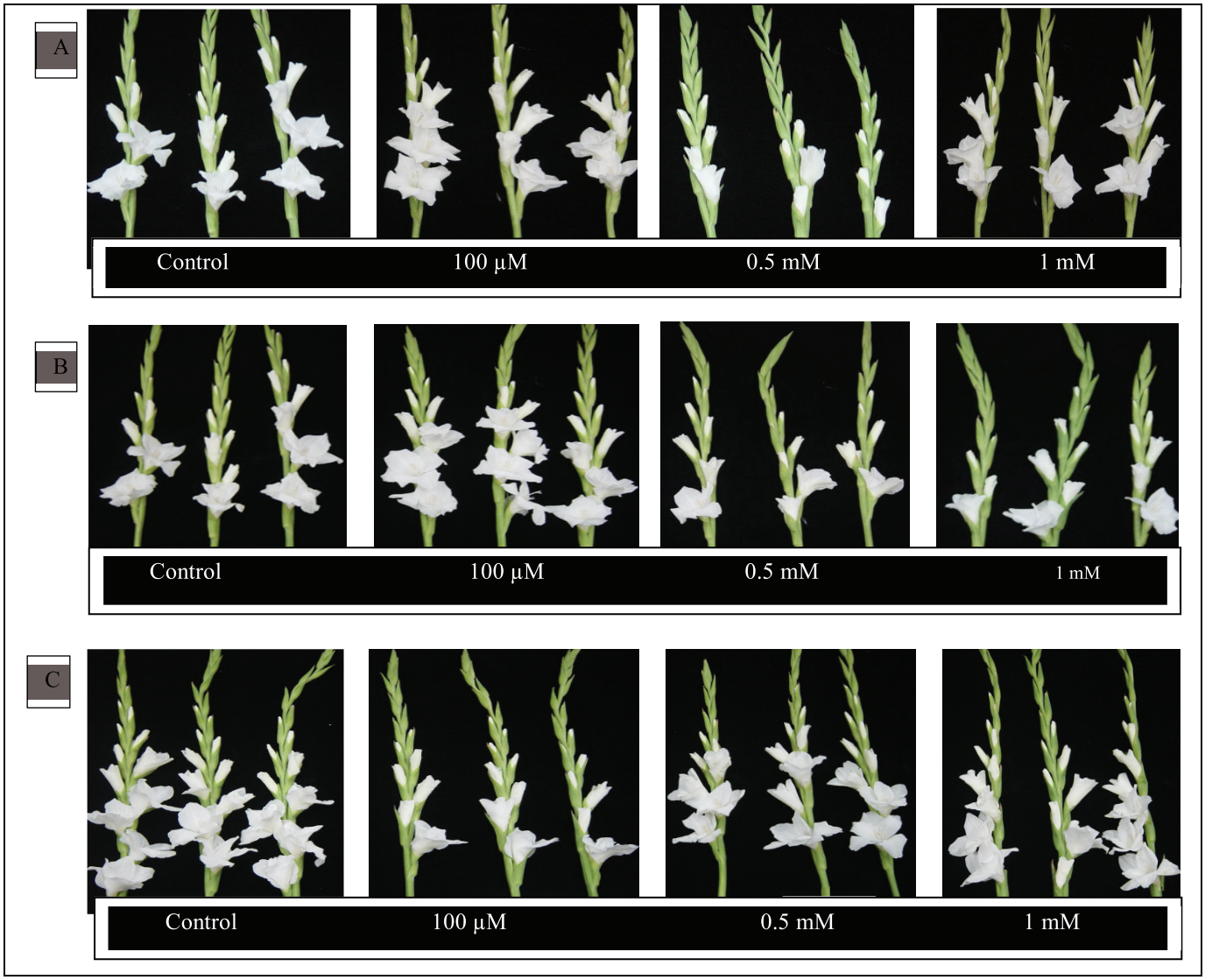


Figure 1: Influence of different concentrations of (A) BAP (B) kinetin, and (C) TDZ during flower senescence in comparison to control.

3.1. Vase Life

The cut flower kept in Milli Q water (control) had a vase life of 11 days only; however, treatment with cytokinin increased the vase life of cut *Gladiolus* flowers evidently. The vase life of flowers kept in 0.5 mM BAP, 1 mM kinetin, and 100 µM TDZ solutions increased to14, 13, and 15 days, respectively, indicating that TDZ was the most efficient cytokinin in enhancing the vase life of *Gladiolus* (Fig. 2).

3.2. Vase Solution Uptake

Various concentrations of BAP, kinetin, and TDZ increased the cumulative uptake of vase solution as compared to untreated flowers. It was found that the 100 µM TDZ concentration was most effective in increasing uptake of vase solution (118 ml/spike) by flower spike followed by 0.5 mM BAP (98.66 ml/spike) and 1 mM kinetin (95.00 ml/spike) as compared to control (53.99 ml/spike) (Fig. 3).

3.3. pH of Vase Solution

Different concentrations of BAP, kinetin, and TDZ decreased the pH of vase solution as compared to control. It is revealed from Figure 4 that 100 µM TDZ concentration was most effective in reducing pH of vase solution (pH 3.00) followed by 0.5 mM BAP (pH 3.17) and 1 mM kinetin (pH 3.45) as compared to control (pH 4.95).From the aforesaid results, it is clear that the 0.5 mM BAP, 1mM kinetin, and 100 µM TDZ treatments had maximum effect on improving vase life of *Gladiolus*. Further, studies were made to understand the effect of 0.5 mM BAP, 1mM kinetin, and 100 µM TDZ on membrane stability index, MDA content, and LOX activity in *Gladiolus* during senescence.

3.4. Membrane Stability Index (MSI)

A linear decline in MSI was observed in both cytokinin treated and untreated flowers during the flower development process (Fig. 5). There was a significant difference in MSI cytokinin treated flowers

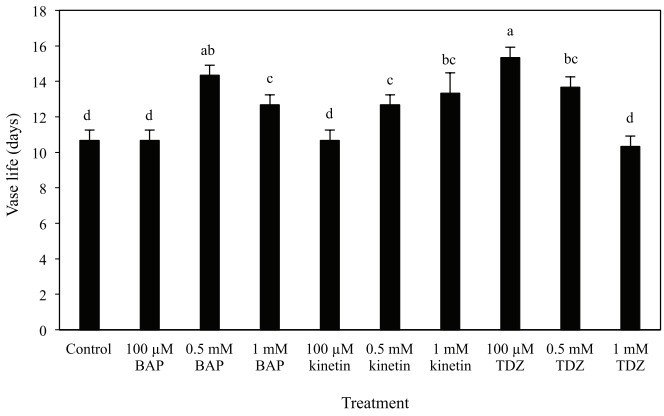


Figure 2: Influence of different concentrations of BAP, kinetin, and TDZ on vase life of *Gladiolus*.

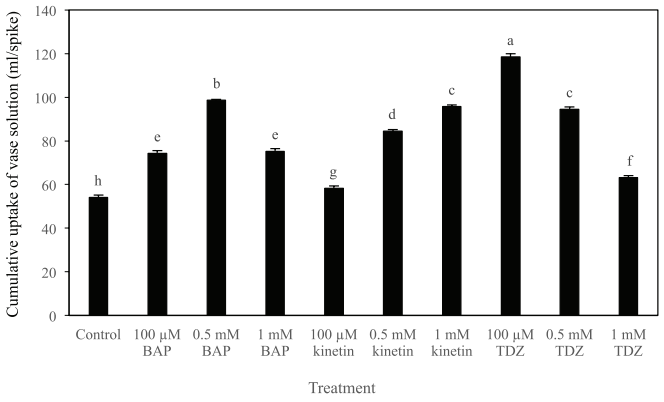


Figure 3: Influence of different concentrations of BAP, kinetin, and TDZ on vase solution uptake of *Gladiolus*.

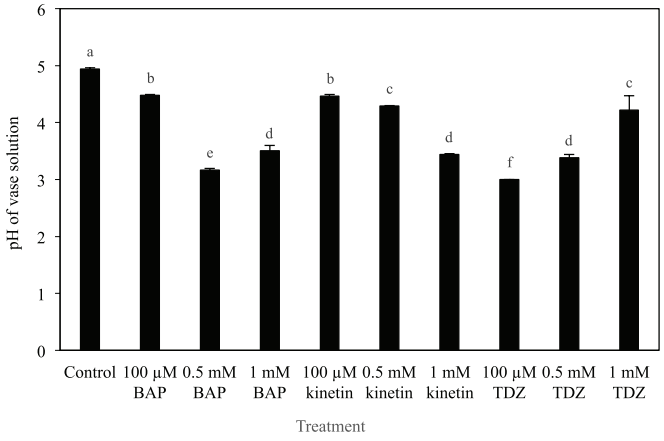


Figure 4: Influence of different concentrations of BAP, kinetin, and TDZ on the pH of vase solution of *Gladiolus*.

and control. The decrease in MSI was significantly alleviated at stage V by treatment with 100 µM TDZ (62.07%) followed by 0.5 mM BAP (36.76%) and 1 mM kinetin (25.75%) in comparison to control (Fig. 6).

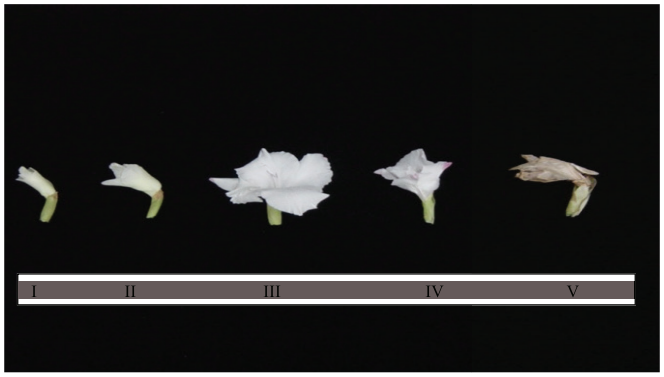


Figure 5: Different stages of flower development of *Gladiolus*. I stage—Slightly opened/point brush stage; II stage—Half opened stage; III stage—Fully opened stage; IV stage—Incipient senescence stage; V stage—Senescent stage.

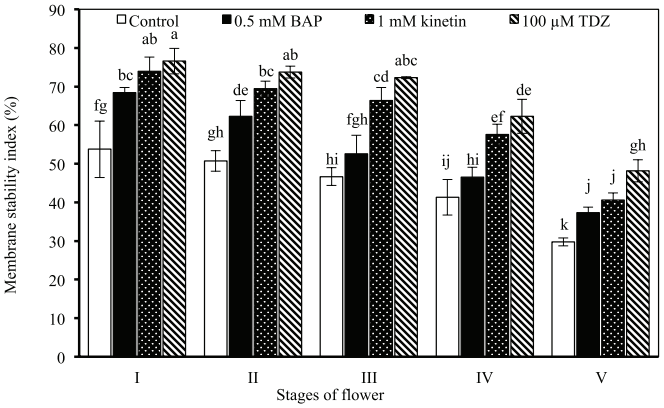


Figure 6: Effect of 1 mM kinetin, 0.5 mM BAP, and 100 µ M TDZ on membrane stability index (%) as compared to control in *Gladiolus* flower during senescence.

3.5. MDA Content

Lipid peroxidation level was studied in terms of MDA content and decrease in the MDA content was observed till stage II, then gradual increase was observed till stage V in both cytokinin treated flowers and in control. Cytokinin-treated flowers had decreased amount of MDA content throughout flower development as compared to the untreated flowers. MDA content displayed a decrease by 63.70% in flowers treated with TDZ that is followed by BAP and kinetin where the MDA content was reduced by 36.72% and 22.53%, respectively, as compared to untreated flowers where MDA level was maximum (Fig. 7).

3.6. LOX Activity

The LOX activity decreased gradually till stage II and then activity increased till stage V in both control and cytokinin treated flowers (Fig. 8). At any given stage, the LOX activity was remarkably less in cytokinin-treated flowers as compared to controlled one. LOX activity significantly decreased by 44.20%, 36.10%, and 12.15% in the flowers treated with 100 µM TDZ, 0.5 mM BAP, and 1 mM

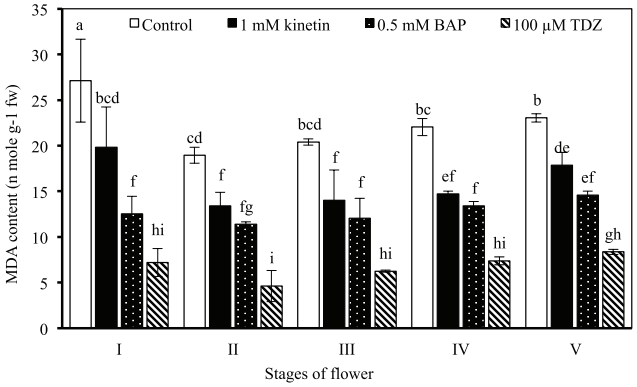


Figure 7: Effect of 1 mM kinetin, 0.5 mM BAP, and 100 µM TDZ on MDA content (n mole g^{-1} fw) as compared to control in *Gladiolus* flower during senescence.

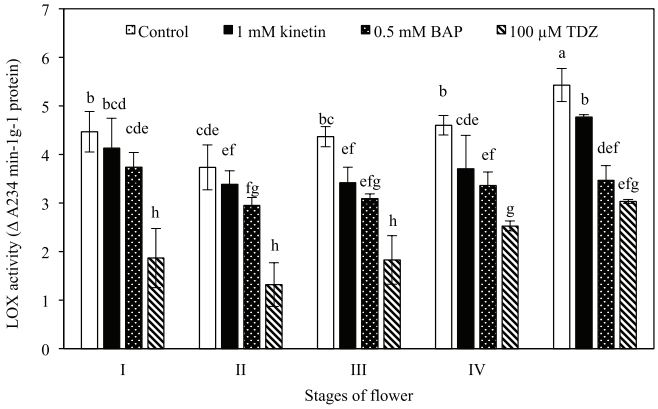


Figure 8: Effect of 1 mM kinetin, 0.5 mM BAP, and 100 µ M TDZ on LOX activity (ΔA_{234} $minute^{-1}g^{-1}$ protein) as compared to control in *Gladiolus* flower during senescence.

kinetin as compared to control. Minimum LOX activity was seen in the flowers subjected to 100 µM TDZ treatment.

4. DISCUSSION

Senescence is defined as age-dependent degenerative process leading to plant death [1]. Indeed, flowers have the showiest petals but have the shortest life span too. The senescence of flowers is fast and foreseeable; therefore, flowers are the perfect model for the study of the senescence [23]. In the market, the life of petals determines the life span of the cut flower [9]. The life span of cut flowers is influenced by number of factors which vary in different ornamental plants [24]. Therefore, the determination of petal senescence helps us to enhance the life span of flowers, as well as in understanding the mechanisms underlying petal senescence [25].

The difference between vase life of cytokinin treated flowers was remarkable than the control. The vase life was increased by treatment with kinetin (1 mM), TDZ (100 µM), and BAP (0.5 mM) (Fig. 2). However, the vase life of TDZ (100 µM) treated flowers was more pronounced than BAP (0.5 mM) followed by kinetin (1 mM). The vase life was enhanced due to increased fresh weight, vase solution uptake, and decreased ion leakage. The vase life of cut rose flowers was increased due to kinetin treatment [26]. Chamani *et al.* [27] reported that TDZ significantly enhanced the fresh biomass and life span of cut carnation flowers. Gulzar et al. [28] reported that kinetin and cycloheximide increased the vase life of daylily flowers by maintaining higher fresh mass. Marandi *et al.* [29] showed that essential oils, salicylic acid increased the vase life of cut Gladiolus flowers by increasing fresh weight and vase solution uptake. BAP, kinetin, and TDZ increased the life span of cut *Gladiolus* flowers by influencing membrane stability and water balance [30,31]. Cytokinin application delays senescence in cut flowers such as *Chrysanthemum, Hemerocallis, Eustoma, Anthurium,* and *Nicotiana* [1,17,31–33]. Application of cytokinins might improve flower vase life by providing more photosynthates to the developing flower bud which creates a water potential gradient, leading to more uptake of water making petal tissue more turgid [17]. This reduction in vase life of control may be due to decreased water uptake [34] and increased permeability of cell membrane.

There was a significant difference between vase solution uptake by cytokinin treated flowers and control. The vase solution uptake was increased by treatment with kinetin (1 mM), TDZ (100 µM), and BAP (0.5 mM) (Fig. 3). However, the vase uptake by TDZ (100 µM) treated flowers was more pronounced than BAP (0.5 mM) followed by kinetin (1 mM). This may be due to decreased pH of vase solution. When the pH of vase solution was lowered, the vase solution traveled faster in the water conducting system. Similar work was reported [35], where salicylic acid enhanced the uptake of vase solution of cut flowers by reducing the pH of vase solution.

There was a remarkable difference in pH value of cytokinin treated flowers and control. The BAP (0.5 mM), kinetin (1 mM), and TDZ (100 µM) acidified (pH 3.17, 3.45, 3.00, respectively) (Fig. 4) the vase solution. When the pH of vase solution was lowered the vase solution travelled faster in the water conducting system.At this pH, the bacterial growth might be reduced, thus improving water uptake. Gowda and Gowda [36] also showed that aluminum sulfate treated flowers lowered the pH of vase solution and reduced the growth of microbes in vase solution during senescence.

The membrane stability index was significantly increased by treatment with 0.5 mM BAP, 1 mM kinetin, and 100 µM TDZ in comparison to control (Fig. 6). However, 100 µM TDZ treatment significantly increased the membrane stability index than 0.5 mM BAP treatment followed by 1 mM kinetin treatment. The change in membrane stability index demonstrates the rate of ion leakage from petals. Leakage of electrolyte has been observed in *Alstroemeria peruviana* [37] during flower senescence. The life span of flowers can be modulated by free radial scavengers [38] suggesting involvement of free radicals during senescence process. Kinetin delayed senescence in cut roses by maintaining cell membrane integrity [26]. The daylily flowers treated with cycloheximide and kinetin maintained the membrane integrity [14,28]. Mortazavi *et al.*[14] reported that TDZ and nitric oxide delayed senescence in cut rose flower by maintain membrane integrity. It is now reasonable to propose that cytokinin may act as free radical scavenger, thus maintaining membrane integrity for extended life span. Cytokinin maintains integrity of cellular membrane by

enhanced water uptake and reduced lipid peroxidation in plants during senescence [39] Kaur *et al.* [40] reported that BAP could enhance membrane stability index in *Calendula officinalis* during senescence. Cytokinin treatment enhanced the membrane integrity due to its participation in maintenance of highwater content, and decreased lipid peroxidation rates in petals, which are affected by senescence [32,39].

It was observed that the MDA content was increasing from the stage III to stage V in both treated and control flowers. However, 0.5 mM BAP, 1 mM kinetin, and 100 µM TDZ treated flowers showed reduced level of MDA during flower development as compared to untreated flowers (Fig. 7). Moreover, 100 µM TDZ treated flowers showed reduced level of lipid peroxidation as compared to 0.5 mM BAP followed by 1 mM kinetin. The MDA content via lipid peroxidation is enhanced by ROS [41]; therefore, 0.5 mM BAP, 1 mM kinetin, and 100 µM TDZ might play a pivotal role in abating the levels of free radicals by decreasing LOX activity. TDZ and BAP treatment increased the cell membrane integrity because of the decreased LOX activity which reduced peroxidation of membrane lipids [32]. Moreover, BAP treated flowers showed a remarkable decrease in MDA content and increase in MSI in cauliflower and carnations during senescence [42,43]. The treatment has reduced the lipid peroxidation; therefore, the role of cytokinin as free radical scavenger can be confirmed.

Gladiolus is ethylene insensitive; therefore, oxidative stress is the main reason of senescence in *Gladiolus*. There is a significant difference in LOX activity in cytokinin treated flowers and control. The LOX activity showed a significant decrease in 0.5 mM BAP, 1 mM kinetin, and 100 µM TDZ treatments as compared to control (Fig. 8). Moreover, TDZ was highly effective in maintaining reduced level LOX activity than 0.5 mM BAP followed by 1 mM kinetin, which is apparent from lower level of MDA and increased membrane stability index. The decrease in the LOX activity in cytokinin treated plants resulted in preventing the phospholipids and proteins structure by reducing the secretion of proteases into the cell cytoplasm [44,45]. Similar work was also reported by where TDZ, kinetin, and BAP application improved the postharvest life of *Iris germanica* by decreasing LOX activity and preventing membrane lipid peroxidation [32]. The decreased LOX activity maintained the membrane integrity by preventing the release of proteases into the cytoplasm from vacuoles thereby maintaining the structure of phospholipids, proteins, and thiols [44,45].

5. CONCLUSION

In conclusion, BAP, kinetin, and TDZ reduced the senescence induced oxidative damage in *G. grandiflora* which is evident from improved flower life span. TDZ was found most effective and outpaced BAP and kinetin in increasing vase life of cut flowers. Cytokinin treatment resulted in maintaining higher fresh weight, vase life membrane stability, besides maintaining lower pH, lipid peroxidation, and LOX activity. The possible role of cytokinins (BAP, kinetin, and TDZ) in controlling senescence process in *G. grandiflora* "White prosperity" cultivar is yet to be explored. Detailed study on a crucial role of cytokinin in retarding of senescence and increasing postharvest life of flowers need to be done at physiological and biochemical level. Furthermore, the mode of action of TDZ should be studied in detail, as it has been showing best results delaying senescence at comparatively lower concentrations. Understanding the mechanistic way of flower senescence in *Gladiolus* will unveil strategies to improve its post-harvest performance.

ACKNOWLEDGMENTS

MS is grateful to the SSN College, University of Delhi, Delhi, India for providing the financial assistance. The author is also thankful to the Department of floriculture, Indian Agricultural Research Institute, New Delhi for providing *Gladiolus* flowers.

CONFLICT OF INTEREST

The authors do not have any conflict of interest in this research.

FUNDING

There is no funding to this research.

REFERENCES

1. Ahmad SS, Tahir I. Increased oxidative stress, lipid peroxidation and protein degradation trigger senescence in *Iris versicolor* L. flowers. Physiol Mol Biol Plants 2016;22(4):507–14; doi:10.1007/s12298-016-0392-9
2. Sarwat M, Tuteja N. Chapter 13 – flower senescence: present status and future aspects. In: Sarwat M, Tuteja N (ed.). Senescence signalling and control in plants, Academic Press, Cambridge, MA, pp 211–25, 2019; doi:10.1016/B978-0-12-813187-9.00013-5.
3. Salleh FM, Mariotti L, Spadafora ND, Price AM, Picciarelli P, Wagstaff C, et al. Interaction of plant growth regulators and reactive oxygen species to regulate petal senescence in wallflowers (*Erysimum linifolium*). BMC Plant Biol 2016;16:77; doi:10.1186/s12870-016-0766-8
4. Chen C, Zeng L, Ye Q. Proteomic and biochemical changes during senescence of phalaenopsis 'Red Dragon' petals. Int J Mol Sci 2018;19:1317; doi:10.3390/ijms19051317
5. Hemati E, Daneshvar MH, Heidari M. The roles of sodium nitroprusside, salicylic acid, and methyl jasmonate as hold solutions on vase life of Gerbera jamesonii 'Sun Spot.' Adv Hortic Sci 2019;33:187–95.
6. Bagniewska-Zadworna A, Stelmasik A, Minicka J. From birth to death – *Populus trichocarpa* fibrous roots functional anatomy. Biolgia Plant 2014;58(3):551–60; doi:10.1007/s10535-014-0433-6
7. Wagstaff C, Malcolm P, Rafiq A, Leverentz M, Griffiths G, Thomas B, et al. Programmed cell death (PCD) processes begin extremely early in *Alstroemeria* petal senescence. New Phytol 2003;160:49–59; doi:10.1046/j.1469-8137.2003.00853.x
8. Shibuya K, Yamada T, Ichimura K. Morphological changes in senescing petal cells and the regulatory mechanism of petal senescence. J Exp Bot 2016;679(20):5909–918; doi:10.1093/jxb/erw337
9. Hegazi MA. Evaluation of Pre- or postharvest application of some minerals and organic agents on the growth, flowering and vase life of *Rudbeckia hirta*, L. J Agric Sci 2016;8:226; doi:10.5539/jas.v8n9p226
10. Wojciechowska N, Sobieszczuk-Nowicka E, Bagniewska-Zadworna A. Plant organ senescence–regulation by manifold pathways. Plant Biol 2018,20:167–81.
11. Iqbal N, Khan NA, Ferrante A, Trivellini A, Francini A, Khan MIR. Ethylene role in plant growth, development and senescence: interaction with other phytohormones. Front Plant Sci 2017;8:475; doi:10.3389/fpls.2017.00475

12. Gupta J, Dubey R. Factors affecting post-harvest life of flower crops. Int J Curr Microbiol Appl Sci 2018;7:548–57; doi:10.20546/ijcmas.2018.701.065
13. Radio MC, Arrom L, Puig S, Munne-Bosch S. Hormonal sensitivity decreases during the progression of flower senescence in *Lilium longiflorum*. J Plant Growth Regul 2017;36(2):402–12; doi:10.1007/s00344-016-9648-4
14. Mortazavi SN, Talebi SF, Naderi RA, Sharafi Y. Regulation of ethylene biosynthesis by nitric acid and thidiazuron during postharvest of rose. J Med Plant Res 2011;5(20):5177–83.
15. Macnish AJ, Jiang CZ, Negre-Zakharov F, Reid MS. Physiological and molecular changes during opening and senescence of *Nicotiana* mutabilis flowers. Plant Sci 2010;179:267–72; doi:10.1016/j.plantsci.2010.05.011
16. Imsabai W, van Doorn WG. Effects of auxin, gibberellin, and cytokinin on petal blackening and flower opening in cut lotus flowers (*Nelumbo nucifera*). Postharvest Biol Technol 2013;75:54–57; doi:10.1016/j.postharvbio.2012.05.015
17. Reid MS, WU MJ. Ethylene in flower development and senescence. In: Mattoo AK, Suttle JC (eds.). The plant hormone ethylene. CRC Press Taylor and Francis Group, Boca Raton, FL, pp 4–32, 2018; doi:10.1201/9781351075763-12
18. Azimi M. Progeny test of crosses among different cultivars of *Gladiolus*. *J Plant Prod* 2019;41(4):29–44; doi:10.22055/ppd.2018.21501.1460
19. Rahmani I, Ahmadi N, Ghanati F, Sadeghi M. Effects of salicylic acid applied pre- or post-transport on post-harvest characteristics and antioxidant enzyme activity of *Gladiolus* cut flower spikes. N Z J Crop Hortic Sci 2015;43:294–305; doi:10.1080/01140671.2015.1096799
20. Bailly C, Benamar A, Corbineau F, Come D. Changes in malondialdehyde content and in superoxide dismutase, catalase and glutathione reductase activities in sunflower seeds as related to deterioration during accelerated aging. Physiol Plant 1996;97:104–10; doi:10.1111/j.1399-3054.1996.tb00485.x.
21. Heath RL, Packer L. Photoperoxidation in isolated chloroplasts: I. Kinetics and stoichiometry of fatty acid peroxidation. Arch Biochem Biophys 1968;125:189–98; doi:10.1016/0003-9861(68)90654-1
22. Doderer A, Kokkelink I, van der Veen S, Valk BE, Schram André W, Douma AC. Purification and characterization of two lipoxygenase isoenzymes from germinating barley. Biochim Biophys Acta 1992;1120:97–104; doi:10.1016/0167-4838(92)90429-H
23. Tripathi SK, Tuteja N. Integrated signaling in flower senescence. Plant Signal Behav 2007;2:437–45; doi:10.4161/psb.2.6.4991
24. Vehniwal SS, Abbey L. Cut flower vase life-influential factors, metabolism and organic formulation. Hortic Int J 2019;3(6):275–81; doi:10.15406/hij.2019.03.00142
25. Van Doorn WG, Woltering EJ. Physiology and molecular biology of petal senescence. J Exp Bot 2008;59:453–80; doi:10.1093/jxb/erm356
26. Mayak S, Halevy AH. The action of kinetin in improving the water balance and delaying senescence processes of cut rose flowers. Physiol Plant 1974;32:330–6; doi:10.1111/j.1399-3054.1974.tb03146.x
27. Chamani E, Irving DE, Joyce DC, Arshad M. Studies with thidiazuron on the vase life of cut rose flowers. J Appl Hortic 2006;8:42–4.
28. Gulzar S, Tahir I, Amin I, Farooq S, Sultan SM. Effect of cytokinins on the senescence and longevity of isolated flowers of day lily (Hemerocallis fulva) cv. royal crown sprayed with cycloheximide. Acta Hortic 2005;669:395–404; doi:10.17660/ActaHortic.2005.669.52
29. Marandi RJ, Hassani A, Abdollahi A, Hanafi S. Improvement of the vase life of cut Gladiolus flowers by essential oils, salicylic acid and silver thiosulfate. J Med Plant Res 2011;5:5039–43.
30. Lama B, Ghosal M, Gupta SK, Mandal P. Assessment of different preservative solutions on vase life of cut roses. J Ornamental Plants 2013;3(3):171–81.
31. Tahir I, Nisar S, Dar RA. Gibberellin and cytokinins modulate flower senescence and longevity in *Nicotiana plumbaginifolia*. Acta Hortic 2019;1263:469–76; doi:10.17660/actahortic.2019.1263.61
32. Ahmad SS, Tahir I, Wani AS, Dar RA, Nisar S. Adenine type and diphenyl urea derived cytokinins improve the postharvest performance of Iris germanica L. cut scapes. Physiol Mol Biol Plants 2018;24:1127–37; doi:10.1007/s12298-018-0554-z
33. Bergmann BA, Ahmad I, Dole JM. Benzyladenine and gibberellic acid pulses improve flower quality and extend vase life of cut dahlias. Can J Plant Sci 2018;99(1):97–101; doi:10.1139/cjps-2018-0126
34. Gul F, Tahir I, Shahri W. Flower senescence and some postharvest considerations of *Amaryllis belladonna* cut scapes. Plant Physiol Rep 2020;25:315–24; doi:10.1007/s40502-020-00506-8
35. Bayat H, Aminifard MH. Salicylic acid treatment extends the vase life of five commercial cut flowers. Electron J 2017;13:1.
36. Gowda J, Gowda VN. Effect of calcium, aluminium and sucrose on vase life of Gladiolus. Crop Res (Hisar) 1990;3:105–6.
37. Leverentz MK, Wagstaff C, Rogers HJ, Stead AD, Chanasut U, Silkowski H, et al. Characterization of a novel lipoxygenase-independent senescence mechanism in *Alstroemeria peruviana* floral tissue. Plant Physiol 2002;130:273–83; doi:10.1104/pp.000919.
38. Saeed T, Hassan I, Abbasi NA, Jilani G. Effect of gibberellic acid on the vase life and oxidative activities in senescing cut Gladiolus flowers. Plant Growth Regul 2014;72:89–95.
39. Honig M, Plíhalova L, Husickova A, Nisler J, Dolezal K. Role of cytokinins in senescence, antioxidant defence and photosynthesis. Int J Mol Sci 2018;19(12):4045; doi:10.3390/ijms19124045
40. Kaur P, Singh N, Mukherjee D. Regulation of membrane leakage and activities of some antioxidant enzymes in petals of cut flowers of Calendula officinalis and Salvia splendens with metabolites and plant growth regulators. J Appl Hortic 2015;17:31–9.
41. Kellogg DE. The role of phyletic change in the evolution of Pseudocubus vema (Radiolaria). Paleobiology 1975;1:359–70.
42. Siddiqui MW, Singh JP, Nayyer MdA, Barman K, Ahmad MS, Kumar V. 6-Benzylaminopurine affects lipid peroxidation and membrane permeability and thereby preserves curd quality and antioxidants during storage of cauliflower. Acta Physiol Plant 2015;37:96. doi:10.1007/s11738-015-1848-1
43. Ramtin A, Naderi R, Kalatejari S, Matinizadeh M. Comparison of plant growth regulators and exogenous ethylene effects on two types of cut carnation (Dianthus caryophyllus L.). J Ornamental Plants 2019;9:55–64.
44. Liu T, Longhurst AD, Talavera-Rauh F, Hokin SA, Barton MK. The Arabidopsis transcription factor ABIG1 relays ABA signaled growth inhibition and drought induced senescence. ELife 2016;5:e13768; doi:10.7554/eLife.13768.
45. Dek MSP, Padmanabhan P, Sherif S, Subramanian J, Paliyath AG. Upregulation of phosphatidylinositol 3-Kinase (PI3K) enhances ethylene biosynthesis and accelerates flower senescence in transgenic *Nicotiana tabacum* L. Int J Mol Sci 2017;18:1533; doi:10.3390/ijms18071533.

How to cite this article:
Singh M, Tiwari N. Thidiazuron outpaces 6-benzylamino purine and kinetin in delaying flower senescence in *Gladiolus grandiflora* by alleviating physiological and biochemical responses. J Appl Biol Biotech XXXX; XX(XX):XX–XX.

View publication stats

2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP - 2021)

UndarkGAN: Low-light Image Enhancement with Cycle-consistent Adversarial Networks

Anil Singh Parihar
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
parihar.anil@gmail.com

Pratik Anand
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
pratikanand99@gmail.com

Aditya Sharma
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
adityas29@outlook.com

Amol Grover
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
amolgrover2@gmail.com

2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP) | 978-1-6654-3277-1/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCCSP52374.2021.9465521

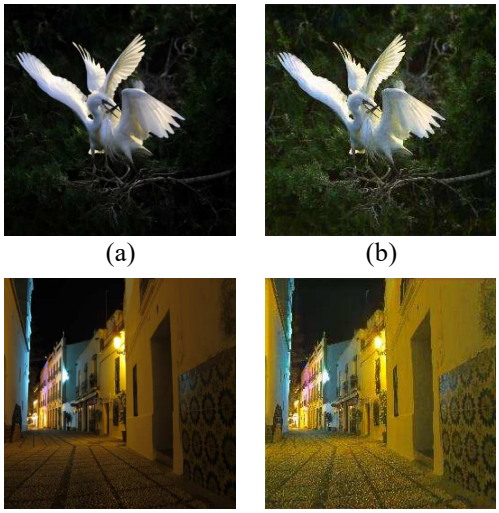
***Abstract*—The paper presents a novel method for enhancing low-light images with a cycle-consistent adversarial network. Images are an essential medium for expressing information and capturing memories. However, the environment always influences the quality of images and the results may be less than ideal, with poorly lit images a common occurrence. The network we propose here processes these low-light images and enhances them while keeping true to their structure and preserving their original details. Unlike several other methodologies, our technique does not introduce any odd contrasts, visual artifacts, or noise in the enhanced images. Our proposed network enhances the dark regions without needlessly overexposing or oversaturating the well-lit portions of the input image. While the predominant share of low-light enhancement techniques requires a paired set of well-lit and poorly-lit images to learn from, the technique we present here does not. We perform experiments and compare our proposed methodology against several recent techniques to demonstrate how the network outpaces them, both perceptually and quantitatively.**

Keywords—computer vision, low-light image enhancement, cycle-consistent adversarial networks, generative adversarial networks.

I. INTRODUCTION

The process of capturing images is virtually indivisible from its physical environment and the surrounding conditions. While good lighting conditions are generally suitable for capturing images due to the ample light; low-light conditions have always had an adverse effect on the quality of an image. Not only do low-light conditions degrade the aesthetic quality of images, but such images also suffer from low brightness, low contrast, color distortion, and considerable noise. As a consequence of this loss of information, these images are not suitable for computer vision tasks including digital image processing, detecting faces, autonomous driving, remote sensing and monitoring, etc.

Through low-light enhancement, it is possible to improve the lighting conditions of images. This not only improves their aesthetics, but it also makes them useful for subsequent processing. There are multiple ways to achieve low-light image enhancement. Earlier approaches use mathematical techniques based on histogram equalization [1, 2] and illumination techniques [3, 4, 5]. More modern approaches utilize CNNs and GANs. These techniques are either applied directly or augment the aforementioned classical techniques.



(a) (b)

(a) (b)

Fig. 1. *(a) Low-light input images (b) Enhanced output images*

Although these above are the different *approaches* towards achieving enhanced low-light images, all of them have processes that can be further binned into two distinct types: (a) paired, and (b) unpaired.

Paired approaches require coupled inputs. These pairs are made up of the same image captured in two visual conditions: the ground truth or reference image in an ideal visual condition; and the corresponding image in a dimly-lit setting. The network takes this paired input and attempts to gain an understanding between the two to eventually attempt to enhance an unpaired dimly lit image. This approach suffers from the lack of data: it is difficult to create large datasets with the same scenery captured in such varying conditions. They often then resort to synthesizing dark images via

978-1-6654-3277-1/21/$31.00 ©2021 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:32:44 UTC from IEEE Xplore. Restrictions apply.

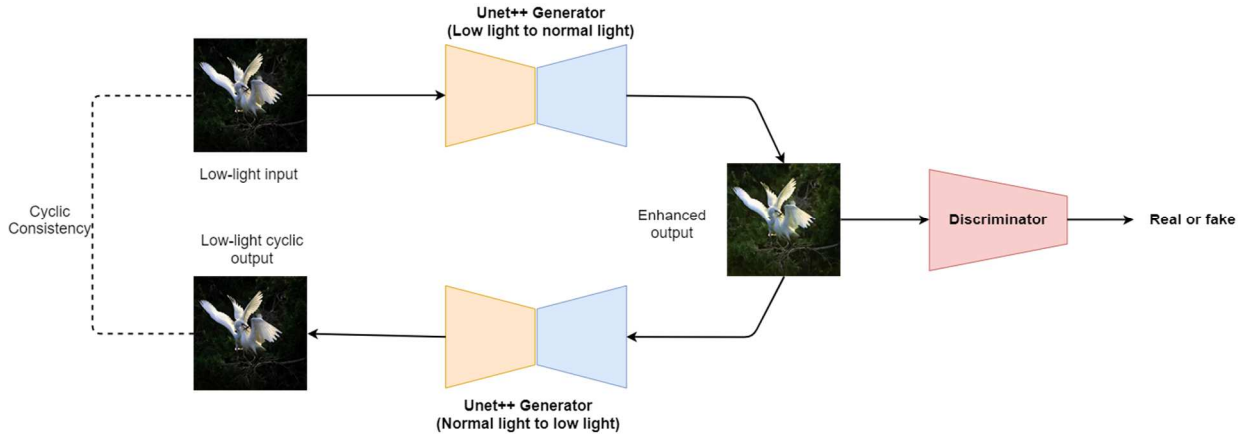


Fig. 2. *Cyclic training process of our network*

postprocessing but these synthesized images are often not ideal for real-world applications. Such approaches include DALE [27] and RetinexNet [16]. On the other side of the spectrum, unpaired techniques do not require coupled images. These techniques rely on two separate datasets of well-lit images and dimly-lit images. The unpaired approaches include EnlightenGAN [24]. There are still other enhancement approaches like ZeroDCE [28] which don't require a reference ground-truth image, but instead estimate requisite finely tuned image curves.

In this paper, we present a novel unpaired technique for low-light image enhancement. To achieve this, we create a specialized cycle-consistent adversarial network. Unlike most common methodologies, our proposed network does not require any paired image datasets to create enhanced images. Conventional cycle-consistent adversarial networks are often used to perform image-to-image translations between two domains. We adapt the network to perform translation between dim- and well-lit domains to achieve enhancement.

We can highlight the major points of our contributions as follows:

1. The network restates the enhancement process as image-to-image translations. In an image with both bright and dark portions, the translation process gives more attention to the dark portions and does not needlessly affect the lighter regions.
2. The novel adaptation that extends and builds upon the cycle-consistent adversarial network architecture. This adapted network is suitable for performing low-light enhancement. Unlike basic methods, this network enhances dark images without introducing any visual artifacts or odd contrasts in the process.

II. Related Work

We have several approaches towards enhancing low-light images. Starting from the basics, we focus on classical methods which are based on Adaptive Histogram Equalizations [1]; Variational Methods which improves the HE [2]; Retinex Theory [3, 4, 5]; and Gamma Correction [6] which overcomes the issue of under and over enhancement. These are mathematical models which do not dynamically learn from new inputs but instead try to change the properties of a low-lit image to correspond to what the properties of a well-lit image look like. SRIE [7] estimates the illumination and reflectance map using a vibrational model. MF [8] enhances the illumination map by fusing multiple derivations of the latter [22]. NPE [9] uses the image naturalness for preventing over-enhancement of the illumination map. LIME [10] used an illumination map, estimated using pixel maxima in the RGB channels and imposed with structure prior, for the low-light enhancement task.

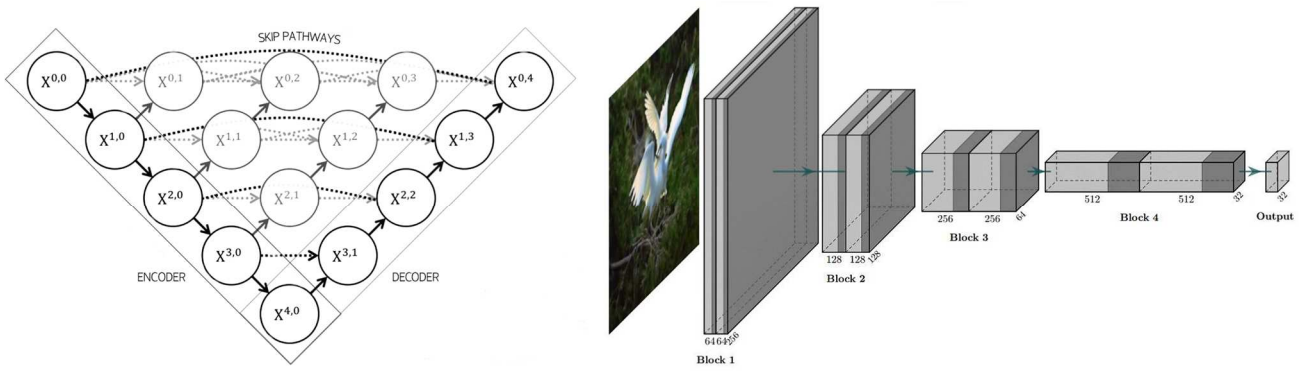
Going forward, we have the deep-learning approaches which are powered by CNNs [11] and GANs [12]. Most of these approaches involve supervised training and rely on paired data samples as discussed earlier. These low-lit images are often synthetically generated using methods like automatic light degradation, image retouching, color transformations etc. Hence, the presence of such unrealistic and factitious training data often results into poor generalization capability for a deep-learning model.

LL-Net [13] uses a stacked autoencoder for identifying the underlying signal features of a low-lit image. MSR-Net [14] learns a direct mapping between the bright and dark images using various Gaussian convolutional kernels. Lighten-Net [15] takes a low-light image as input, produces an illumination map and performs enhancement on the basis of the Retinex model. Retinex-Net [16] provides an end-to-end framework for decomposition of an image into their respective reflectance and illumination maps which are then used in the enhancement process of the image. KinD [17] is another Retinex based approach where a mapping is proposed which fits normal lit-images better than the previous gamma correction. MBLLEN [18] proposed a three-module CNN architecture which handles enhancement, extraction, and fusion for the suppression of image noise and artifacts while performing the low-light enhancement task.

In recent times, GANs have been proven successful in domains like image restoration, synthesis and translation. Low-light image enhancement involves image-to-image translation which can be based on both – supervised and unsupervised methods. Here, it is assumed that the input comes from one specific domain consisting of the dark environment while the enhanced images belong to another domain. Pix2PixGAN [19] uses paired data for training an image translator while CycleGAN [20] and DualGAN [21] examples of unsupervised learning based GANs which work on unpaired data.

LowLightGAN [23] proposed an enhanced GAN which uses spectral normalizations and multiple advanced loss functions to improve the results. EnlightenGAN [24] is an unsupervised GAN based method which uses unpaired samples of data for training. The Deep Photo Enhancer [26]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:32:44 UTC from IEEE Xplore. Restrictions apply.



(a) Generator (b) Discriminator

Fig. 3. *(a) U-Net++ architecture consisting of an encoder, a decoder, and nested skip pathways where each node $X^{i,j}$ represents a convolutional block. (b) Discriminator of the proposed network with output patches having size 32×32.*

has a two-way GAN similar to the CycleGAN. The enhancer's GAN is augmented and optimized to direct the understanding of 'good' characteristics in a well-lit picture, such as the scene-setting, lighting, etc. It then attempts to achieve these transformations in an input image, locally.

III. Proposed Method

We present our novel approach towards enhancing low-light images via unpaired image-to-image translations through a cycle-consistent adversarial network. We first establish the working of a basic cycle-consistent adversarial network and the nature of cycle consistency which makes image translations possible. Subsequently, we go in more detail about how we make the network generate enhanced well-lit images, without suffering from artifact generation or saturation abnormalities that the original CycleGAN suffers from. We then also provide details on the working of the network as well as the loss functions that ensure proper structural integrity and prevent aberrations and deformed contrast levels in the enhanced image.

A. Cycle-Consistent Adversarial Network

The Cycle-Consistent Adversarial Network [20], or CycleGAN for short, was proposed to perform image-to-image translations in the absence of any paired dataset. Before the CycleGAN, paired datasets were required to perform image-to-image translations [19]. However, this introduction made possible translations previously impossible due to the lack of paired data possible. These include translating to and from real-life photographs and Monet paintings; changing horses into zebras in pictures; translating between winter and summer images, or vice versa. Not only that, it can also be used for photo enhancement to generate narrow depth-of-field images with a bokeh effect.

The cyclic network works between translating two different image domains and has two separate generative adversarial networks. We may call the two domains P and Q and the two generative adversarial networks A and B. The GANs A and B are trained in such a manner so as to learn the mappings $A: P \rightarrow Q$ and $B: Q \rightarrow P$, respectively.

The eventual goal of the two generative networks is to ensure that the images generated by A from domain P, *viz* $A(P)$, are indistinguishable from the domain Q. Conversely, the generated images $B(Q)$ should be indistinguishable from the other domain, P. Both of these generators work against their own respective discriminator, that is D_Q and D_P. The discriminator D_Q identifies images from the domain Q and hence discriminates between Q and $A(P)$; while D_P identifies images from the domain P and discriminates P from $B(Q)$.

We use adversarial loss to ensure that images generated from a source domain are indistinguishable from images in the target domain.

However, just the adversarial loss above is not able to sufficiently constrain the network. To achieve the required constrain on the model, cycle consistency loss is applied. This cycle consistency loss influences the generators to operate as inverses of each other and serve as bijective mappings. Essentially, it ensures that images generated by one generator and then processed by the other generator should be as close to the original domain as possible.

$$B(A(P)) \approx P \quad (1)$$

$$A(B(Q)) \approx Q \quad (2)$$

The total loss of this unaugmented network now comes out to be the sum of the adversarial loss and cycle consistency loss, both weighed by an appropriate hyperparameter. These losses are discussed in more detail in *III (D)*.

B. Generator

The enhancement of low-light images requires that the well-lit images generated by the network are structurally sound and free from any visual artifacts. Essentially, this is made possible when the image segmentation process is able to identify structures in detail for subsequent enhancement.

For enhancing poorly-lit images, we utilize UNet++ [29] in both the generators of the cycle-consistent adversarial network. UNet++ is an enhancement over the original UNet architecture [25] and showcases an IoU gain of 3.9 points over the traditional architecture. Both these networks were originally intended for segmentation of biomedical images. However, they also serve well for us. This can be attributed to the original requirement of accurate biomedical image segmentation. UNet++ captures the rich texture information from the given input images and uses multiscale context information for synthesizing high-quality images. This makes it suitable for our purpose.

UNet++ is constituted by an encoder-decoder architecture with a nested skip-connection network and dense convolutional blocks. The presence of convolutional blocks on the skip-pathways allows semantic information from encoder's feature maps to be passed on to the decoder while also ensuring a gradient flow between the same.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:32:44 UTC from IEEE Xplore. Restrictions apply.

Our generator architecture consists of 5 down-sampling and up-sampling stages. Each down-sampling stage consists of a convolutional block made up of one 2×2 convolutional layer followed by batch-normalization and ReLu activation. The skip-pathways between each up-sampling and down-sampling block are also built using the same blocks with additional residual connections between each of the blocks concatenated at the input. The up-sampling stage involves a 2×2 deconvolutional layer having the concatenated residual connections as the input. The final output undergoes Tanh activation to obtain the enhanced image. The structure of the generator is shown in Fig. 3 (a).

C. Discriminator

The discriminator network determines whether an image sample belongs to the model distribution or the data distribution. For our purpose, the network categorizes whether the input normal-lit image sample was a real image or generated through the adversarial network.

We use two PatchGAN discriminators [19] coupled with Instance Normalization [33] and LeakyReLu [34] for our cycle consistent adversarial network, one in each block. These PatchGAN discriminator networks model the high-frequency features of an image sample while the low-frequency features are being implicitly captured by the different losses. Such networks have fewer number of parameters, train faster, and produce accurate outputs.

Here, the classification involves determining whether an image patch of size $N \times N$ is "real" or "fake". This allows the network to locally judge each patch. This is essential because our image may be constructed of a mix of light and dark portions as well as regions with varying structures and contrasts. If the discriminator just processed the image as a whole instead of patches, then visually different portions would be processed similarly and that would create visual artifacts and unnatural contrasts. This requirement of judging portions of image separately makes PatchGAN the ideal candidate for a discriminator, shown in Fig. 3 (b).

D. Losses

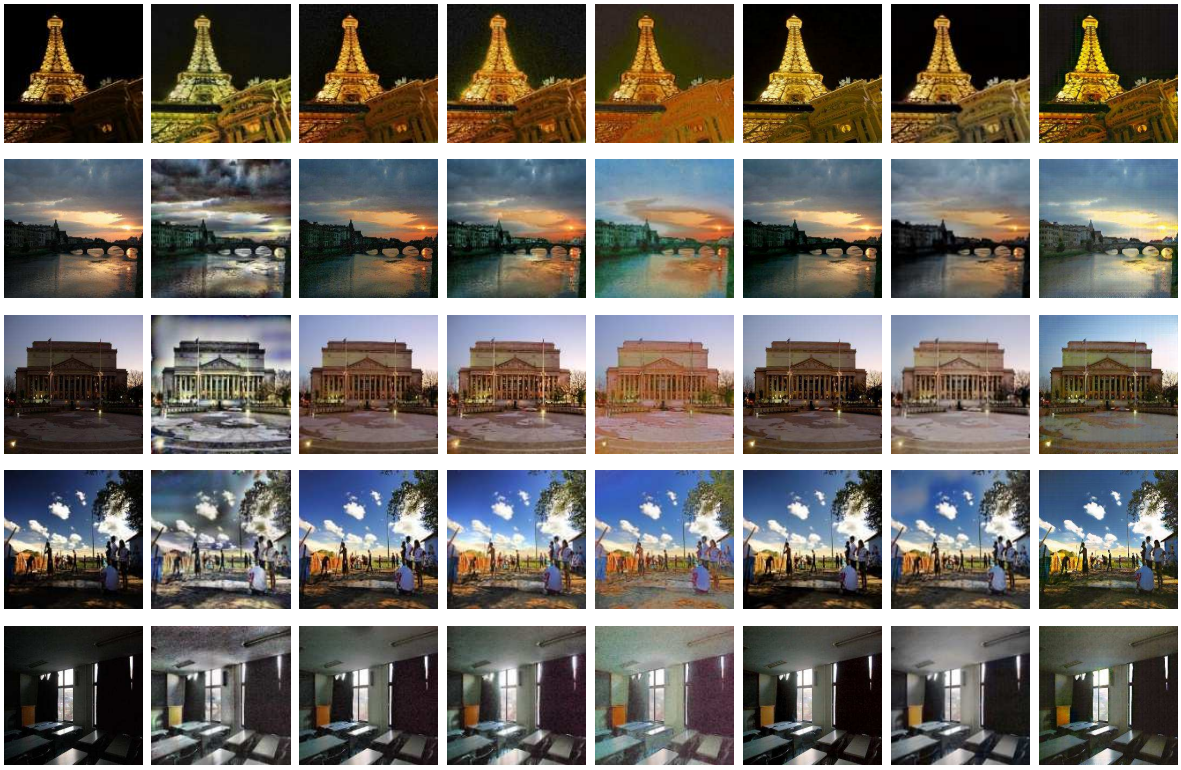
As discussed above in *III (A)*, a cycle consistent adversarial network has a few losses to ensure cycle consistency. This cyclic consistency is what enables the network to perform image-to-image translations. These implicit losses are:

1) Adversarial Loss: It is also known as GAN loss and was first introduced in [12]. The generator aims to minimize, while the discriminator opposes it and maximizes the log probability value for real images combined with the log of inverted probability of fake images. This results in the generator producing more plausible images, similar to the real data, thus making it hard for the discriminator to classify the generated image as fake. The adversarial loss is given by the following function:

$$\mathcal{L}_{gan}(A, D) = \mathbb{E}_{q \sim p_{data}(q)}[\log D(q)] + \mathbb{E}_{p \sim p_{data}(p)}[\log(1 - D(A(p)))] \quad (3)$$

where, $D(q)$ is the discriminator's probability estimate that q is from real dataset, $A(p)$ is the generated image, and $D(A(p))$ is the discriminator's probability estimate that the generated image is actually real.

2) Cycle Consistency Loss: Just the adversarial loss is not sufficient for ensuring image-to-image translations on unpaired datasets. Consider the following 2 mappings – $A: P \rightarrow Q$ *and* $B: Q \rightarrow P$. It enforces that both these mappings are inverse of each other, that is, $B(A(P)) \approx P$ and $A(B(Q)) \approx Q$. This results in reduction of possible mapping functions by enforcing forward and backward



(a) Input (b) CLAHE (c) SRIE (d) LIME (e) RetinexNet (f) MBLLEN (g) KinD (i) Ours

Fig. 4. *Visual Comparison with other Low-Light Image Enhancement methods*

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:32:44 UTC from IEEE Xplore. Restrictions apply.

consistency. The cycle consistency loss is given by the following function:

$$\mathcal{L}_{cyc}(A,B) = \mathbb{E}_{p\sim p_{data}(p)}[\|B(A(p)) - p\|_1] + \mathbb{E}_{q\sim p_{data}(q)}[\|A(B(q)) - q\|_1] \quad (4)$$

However, as discussed above these two only ensure proper image translations. For our purpose, we need to consider more details about the image structure to ensure good outputs. The process of low-light enhancement is bound to introduce structural abnormalities and artifacts. To make sure that such occurances do not occur, we utilize the following loss:

3) Total Variation Loss: In an uncontrolled network, the generation of enhanced well-lit images is bound to create visual artifacts or cause issues with the structures present in input image. To ensure that they are maintained and artifacts are suppresed, we use the total variation loss. The total variation loss is an indicator of the noise in an image and by minimizing it we subsequently minimize noise in our generated enhanced images. Using total variation loss as a structure preserving loss has a regularization effect to it. Traditional noise removal techniques may have an added disadvantage of accidental smoothing across edges. By contrast, total variation loss preserves edges and hence the structural integrity of the overall image. For our purpose, we compute the anisotropic total variation loss over all the pixels of the generated image. We multiply the calculated total variation loss value with a constant to normalize the value with respect to other losses. The total variation loss is given by the following function:

$$\mathcal{L}_{TV-aniso} = \sum_{k,l}\sqrt{|s_{k+1,l} - s_{k,l}|^2} + \sum_{k,l}\sqrt{|s_{k,l+1} - s_{k,l}|^2} = \sum_{k,l}\left(|s_{k+1,l} - s_{k,l}| + |s_{k,l+1} - s_{k,l}|\right) \quad (5)$$

where s is the generated 2D image, and $s_{k,l}$ denotes the l^{th} pixel from left in the k^{th} row from the top of the image. For our purposes, we use the following loss function in actual calculation:

$$\mathcal{L}_{TV} = \mu \times \mathcal{L}_{TV-aniso} \quad (6)$$

Here, μ is a constant factor used to normalize the effect of total variation loss. Minimizing the total-variation loss can have the unintentional result of reduction in image details. To counteract this effect and maintain details, a low value of μ is used. For our experiments, we set it to the order of 10^{-6}.

The total loss for the generative network now comes out to be the sum of these three values:

$$\mathcal{L}_G = \mathcal{L}_{gan} + \mathcal{L}_{cyc} + \mathcal{L}_{TV} \quad (7)$$

IV. Experiments

We compare and contrast how our cycle-consistent image enhancement network performs against various traditional and data driven techniques. We take a look at the metrics used for quantitative comparison and list the datasets used. Finally, we list the results obtained by different techniques on the defined metrics and showcase the visual outputs in Fig. 4.

TABLE I. Full-Reference Quantitative Assessment

Method	Metrics (PSNR ↑ / SSIM ↑)			
	DICM	LIME	MEF	NPE
HE	12.35/0.52	7.70/0.28	8.21/0.29	14.74/0.28
CLAHE	12.74/0.42	10.95/0.27	12.48/0.30	11.37/0.41
SRIE	17.09/0.65	16.19/0.53	16.26/0.52	16.77/0.71
LIME	13.63/0.57	12.79/0.42	12.53/0.40	13.30/0.62
RetinexNet	13.22/0.57	9.97/0.34	10.38/0.34	11.40/0.59
MBLLEN	18.25/0.70	14.39/0.54	16.21/0.59	19.59/0.77
KinD	15.04/0.67	12.90/0.42	13.52/0.46	14.51/0.66
UndarkGAN (Ours)	**19.41/0.75**	**21.38/0.75**	**19.39/0.67**	**19.81/0.80**

A. Datasets and Metrics

We propose an unpaired network which does not require any paired datasets. This lends us the freedom to utilize a variety of different datasets for the unpaired network. Our training data consists of 1016 normal-light images and 914 low-light images belonging to different natural scenes, indoor locations, and architecture buildings. Each PNG image during the training phase has the resolution of 600×400 and random patches of 256×256 are used.

For testing and benchmarking metrics, we used the following standard datasets: 1) DICM; 2) MEF; 3) LIME; and 4) NPE. Between them all, they contain a total of 180 images belonging to multifarious domains.

To calculate the quality of generated images, we have a variety of full- and no-reference metrics. Full-reference metrics compare our generated image with the original image. It includes 1) PSNR or Peak Signal to Noise Ratio and 2) SSIM [30] or Structural Similarity Index. No-reference metrics compare the image statistics of our generated image with a set of features from a database of pristine images. These metrics are 1) NIQE [31] or Naturalness Image Quality Evaluation and 2) BRISQUE [32] or Blind Reference-less Image Spatial Quality Evaluator.

TABLE II. No-Reference Quantitative Assessment

Method	Metrics (NIQE ↓ / BRISQUE ↓)			
	DICM	LIME	MEF	NPE
HE	6.01/26.54	8.41/28.51	5.63/27.96	5.46/22.91
CLAHE	5.60/27.11	8.82/32.26	5.77/27.32	5.64/28.31
SRIE	5.28/**25.73**	5.20/25.20	5.27/27.03	5.45/22.00
LIME	6.52/26.91	7.36/30.16	5.43/25.84	5.61/24.93
RetinexNet	7.82/30.57	9.02/35.57	7.30/24.67	6.39/27.56
MBLLEN	5.41/26.01	6.19/28.42	5.50/32.08	5.68/24.65
KinD	6.16/27.10	6.67/24.31	5.27/29.60	5.61/23.72
UndarkGAN (Ours)	**4.91**/25.73	**4.46**/**22.01**	**4.84**/**24.33**	**4.90**/**20.45**

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:32:44 UTC from IEEE Xplore. Restrictions apply.

(a) Input (b) UNet (c) UNet++ (d) UNet++ with TV

Fig. 5. *Visual Comparison from the ablation study of our network*

B. *Quantitative Assessment*

We analyze our generative network on the aforementioned metrics. For the analysis, we first employ the two full-reference quantitative metrics, PSNR and SSIM; and then the two no-reference metrics, BRISQUE and NIQE. All analysis is done on the four standard datasets; DICM, LIME, MEF, and NPE. All the images used for quantitative assessment have the uniform resolution of 256×256×3 and are in PNG format. First, we perform a full-reference analysis of the prevailing low-light enhancement methods. The values in Table I showcase our model's full-reference PSNR and SSIM metrics and compare them with results of earlier methods. Similarly, the no-reference quantitative assessment of the outputs with NIQE and BRISQUE are shown in Table II. The best values are highlighted in **bold** while the second best are underlined. From these values in Table I, we can showcase that our network is able to generate enhanced well-lit images which are accurate to the original structure. This indicates the lack of erroneous structures or any visual artifacts. Our model is able to achieve better full-reference values than others due to its focus on maintaining the structural integrity of our output. This is brought about by two key components of the network: (1) the UNet++ generator providing a detailed encoder-decoder network, and (2) the anisotropic total variation loss minifying the errors and noises in the generated image. The two-fold effect of these factors together contribute to the high scores we observe here. As we see again, our network provides better perceptual scores than the vast majority of other prevalent methods. We can attribute this to the cycle consistency enforced between the two networks. The network is able to translate poorly-lit images to well-lit conditions without any generation artifacts or poor contrast. The low NIQE and BRISQUE values are a quantitative testament to the fact that these translated images appear natural and don't have any visual oddities. It all indicates the high perceptual quality of our generated images.

TABLE III. ABLATION STUDY RESULTS

Method	Mean metrics			
	PSNR (↑)	SSIM (↑)	NIQE (↓)	BRISQUE (↓)
Resnet	17.98	0.563	6.52	30.73
U-Net	19.94	0.634	5.03	27.84
U-Net++	19.51	0.749	4.86	25.98
U-Net++ and TV	19.99	0.747	4.77	23.13

C. *Ablation Study*

We demonstrate the efficiency of various components in our cycle-consistent adversarial network via an ablation study. We study the effect of generator networks and the benefits introduced by bringing in the total variation loss function. We note the experimental values in Table III and show visual comparison in Fig. 5. **Effect of Generator:** The generator has a major contribution towards the images produced. We first use Resnet [35] encoder-decoder network as our generator. The images generated by it tend to be blurry, noisy, and lacking in detail. This is because of the lack of skip connections between the intermediate convolutional blocks, thus hindering the transfer of information in the network. To make up for this we use the UNet encoder-decoder network. An instant improvement in the quality of images is observed but they lack appropriate illumination enhancement. The original objective of the UNet network was accurate biomedical segmentation. This is apparent with the level of detail observed in the images generated by this network. The UNet++ network is an upgrade to the latter and has a nested structure of upsampling, downsampling and skip-pathway convolutional blocks facilitating the movement of semantic information across the network. **Contribution of Loss:** We compare the images created before and after introducing total

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:32:44 UTC from IEEE Xplore. Restrictions apply.

variation loss to the UNet++ generator. Although the difference may not be as apparent from the full-reference metrics, the no-reference perceptual scores and visual improvements clearly show that anisotropic total variation improves results.

V. Conclusion

Through this paper, we proposed a cycle-consistent adversarial network for enhancing low-light images. While most other low-light enhancement techniques are paired, the network we propose is unpaired. We describe the architecture of our network and then talk about how we perform image-to-image translations between dark and well-lit image domains. The proposed network enhances the darker regions of input images while keeping the well-lit portions as is. Moreover, our technique does not introduce any visual oddities like artifacts, oversaturated regions, or noise. Through experimentation, we define the ability of the network to outpace the vast majority of other common low-light enhancement techniques. We also performed an ablation study which analyzes the influence each key component has on the end result, both visually and quantitatively.

References

[1] Pizer, S., Amburn, E., Austin, J., Cromartie, R., Geselowitz, A., Greer, T., Romeny, B., Zimmerman, J., & Zuiderveld, K. (1987). Adaptive histogram equalization and its variations. *Computer Vision Graphics & Image Processing*, 39(3):355–368.

[2] Jafar, I., & Ying, H. (2007, May). Image contrast enhancement by constrained variational histogram equalization. In 2007 *IEEE International Conference on Electro/Information Technology* (pp. 120-125). IEEE.

[3] Land, E. H., & McCann, J. J. (1971). Lightness and retinex theory, *Josa*, vol. 61, no. 1, pp. 1–11.

[4] Parihar, A. S., & Singh, K. (2018). A study on Retinex based method for image enhancement. In 2018 *2nd International Conference on Inventive Systems and Control* (pp. 619-624).

[5] Singh, K., & Parihar, A. S. (2020). A comparative analysis of illumination estimation based Image Enhancement techniques. In 2020 *International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)* (pp. 1-5).

[6] Parihar, A. S. (2017). Fuzzy adaptive gamma correction for contrast enhancement. In *2017 International Conference on Intelligent Sustainable Systems* (pp. 625-629).

[7] Fu, X., Zeng, D., Huang, Y., Zhang, X. P., & Ding, X.. A weighted variational model for simultaneous reflectance and illumination estimation. In *CVPR*, pages 2782–2790, 2016.

[8] Fu, X., Zeng, D., Huang, Y., Liao, Y., Ding, X., & Paisley, J. (2016). A fusion-based enhancing method for weakly illuminated images. *Signal Processing*, *129*, 82-96.

[9] Wang, S., Zheng, J., Hu, H. M., & Li, B. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9):3538–3548, 2013.

[10] Guo, X., Li, Y., & Ling, H. (2016). LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, *26*(2), 982-993.

[11] Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017, August). Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)* (pp. 1-6). IEEE.

[12] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. In Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2 (pp. 2672–2680).

[13] Lore, K. G., Akintayo, A., & Sarkar, S. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017.

[14] Shen, L., Yue, Z., Feng, F., Chen, Q., Liu, S., & Ma, J. (2017). Msr-net: Low-light image enhancement using deep convolutional network. *arXiv preprint arXiv:1711.02488*.

[15] Li, C., Guo, J., Porikli, F., & Pang, Y. (2018). LightenNet: A convolutional neural network for weakly illuminated image enhancement. *Pattern Recognition Letters*, *104*, 15-22.

[16] Wei, C., Wang, W., Yang, W., & Liu, J. (2018). Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*.

[17] Zhang, Y., Zhang, J., & Guo, X. (2019, October). Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 1632-1640).

[18] Lv, F., Lu, F., Wu, J., & Lim, C. (2018, September). MBLLEN: Low-Light Image/Video Enhancement Using CNNs. In *BMVC* (p. 220).

[19] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).

[20] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223-2232).

[21] Yi, Z., Zhang, H., Tan, P., & Gong, M. (2017). Dualgan: Unsupervised dual learning for image-to-image translation. In *Proceedings of the IEEE international conference on computer vision* (pp. 2849-2857).

[22] A. S. Parihar, K. Singh, H. Rohilla, G. Asnani, & H. Kour (2020). A Comprehensive Analysis of Fusion-based Image Enhancement Techniques. In *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 823-828).

[23] Kim, G., Kwon, D., & Kwon, J. (2019, September). Low-lightgan: Low-light enhancement via advanced generative adversarial network with task-driven training. In *2019 IEEE International Conference on Image Processing (ICIP)* (pp. 2811-2815). IEEE.

[24] Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., ... & Wang, Z. (2019). Enlightengan: Deep light enhancement without paired supervision. *arXiv preprint arXiv:1906.06972*.

[25] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.

[26] Chen, Y. S., Wang, Y. C., Kao, M. H., & Chuang, Y. Y. (2018). Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 6306-6314).

[27] Kwon, D., Kim, G., & Kwon, J. (2020). DALE: Dark Region-Aware Low-light Image Enhancement. *arXiv preprint arXiv:2008.12493*.

[28] Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., & Cong, R. (2020). Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1780-1789).

[29] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support* (pp. 3-11). Springer, Cham.

[30] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, *13*(4), 600-612.

[31] Mittal, A., Soundararajan, R., & Bovik, A. C. (2012). Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, *20*(3), 209-212.

[32] Mittal, A., Moorthy, A. K., & Bovik, A. C. (2012). No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, *21*(12), 4695-4708.

[33] Ulyanov, D., Vedaldi, A., & Lempitsky, V. (2016). Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*.

[34] Xu, B., Wang, N., Chen, T., & Li, M. (2015). Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*.

[35] 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).

[36] Parihar, A. S., Singh, K., Rohilla, H., Asnani, G.. Fusion-based simultaneous estimation of reflectance and illumination for low-light image enhancement. *IET Image Process*. 2020; 1– 14.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2021 at 09:32:44 UTC from IEEE Xplore. Restrictions apply.