SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(111st Edition)

MARCH 2022

BY: CENTRAL LIBRARY

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

PREFACE

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

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

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

This issue contains the information published during **March, 2022**. 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 Novel Approach for Early Recognition of Cataract using VGG-16 and Custom User-based Region of Interest, ***8.Aaron Paul Jacob***, ***8.Aryan Bansal*** and ***3.Ruchika Malhotra***, Software Engineering, DTU

2. A novel protocol for efficient authentication in cloud-based IoT devices, ***6.Irfan Alam*** and ***3.Manoj Kumar***, CSE, DTU

3. A Survey and Feasibility Research Study on Computer Aided Process Planning (CAPP), ***3.Ashok Kumar Madan***, ***7.Kunal Krishna*** and ***7.Bipul Kumar***, Mechanical and Production, DTU

4. An intelligent railway surveillance framework based on recognition of object and railway track using deep learning, ***3.Rajiv Kapoor***, Rohini Goel and Avinash Sharma, Electronics, DTU

5. Analysis of Wind on Different Shape of High-Rise Structure, ***6.Rahul Kumar Meena***, ***3.Ritu Raj*** and ***3.S. Anbukumar***, Civil, DTU

6. Biophysical, Biochemical, and Molecular Docking Investigations of Anti-Glycating, Antioxidant, and Protein Structural Stability Potential of Garlic, Mohd W. A. Khan, Ahmed A. Otaibi, Abdulmohsen K. D. Alsukaibi, Eida M. Alshammari, Salma A. Al-Zahrani, Subuhi Sherwani, Wahid A. Khan, ***6.Ritika Saha***, ***3.Smita R. Verma*** and Nessar Ahmed, Biotechnology, DTU

7. Breast Cancer Mortality Trends and Predictions to 2030 and Its Attributable Risk Factors in East and South Asian Countries, Sumaira Mubarik, ***3.Rajesh Sharma***, Syeda Rija Hussain, Mujahid Iqbal, Nawsherwan, Xiaoxue Liu and Chuanhua Yu, USME, DTU

8. Can Unsupervised Knowledge Transfer from Social Discussions Help Argument Mining?, Subhabrata Dutta, ***8.Jeevesh Juneja***, Dipankar Das and Tanmoy Chakraborty, IT, DTU

9. Compatibility of a Truncated Fin-FinFET as a k-modulated Biosensor with Optimum parameters for Pre-emptive Diagnosis of Diseases, ***8.Mridul Prakash Kashyap***, ***8.Harshal Gudaghe*** and ***3.Rishu Chaujar***, Physics, DTU

10. Comprehensive analysis of design software application in solar distillation units, Om Prakash, Asim Ahmad, ***3.Anil Kumar***, S.M. Mozammil Hasnain and Gaurav Kumar, Mechanical, DTU

11. Crime Based Evaluation of GPS Network Using Machine Learning Techniques, ***3.Rahul***, Monika and ***8.Vartika Hari Durgapal***, Software Engineering, DTU

12. Decoding the BrainWaves using EEG signals for classifying Body Gestures by applying suitable ML & DL Techniques, ***3.Dr Ruchika Malhotra***, ***8.Dhananjay Deswal*** and ***8.Jai Chaudhry***, CSE, DTU

13. Design of 31 and 37 cores trench-assisted and air-hole assisted multi-core fibre for high-density space-division multiplexing, ***6.Garima Singh***, Amin Khorshidahmad, ***3.Gurjit Kaur*** and Ahmad Atieh, Electronics, DTU

14. Detection of Flooding Attacks UsingMultivariate Analysis, ***3.Priyanka Meel*** and Tanmay Singh, IT, DTU

15. DIAT-RadHARNet: A Lightweight DCNN for Radar based Classification of Human Suspicious Activities, Mainak Chakraborty, ***7.1.Harish C. Kumawat***, Sunita Vikrant Dhavale and Arockia Bazil Raj A, Electronics, DTU

16. Effect of Carrier gas on sensitivity of Surface Acoustic Wave Detector, Jitender Kumar, A.T.Nimal, ***7.1.Upendra Mittal***, Vinod Kumar and Vinay Kumar Singh, Electrical, DTU

17. Effect of Cr doping on Li2ZnTi3O8 as alternative anode material to enhance electrochemical properties of lithium-ion batteries, ***6.Anchali Jain***, ***3.Amrish K. Panwar*** and ***3.Pawan K. Tyagi***, Physics, DTU

18. Electrochemical Studies of Biofunctionalized MoS2 Matrix for Highly Stable Immobilization of antibodies and Detection of Lung Cancer Protein Biomarker, ***6.Ritika Khatri*** and ***3.Nitin K. Puri***, Applied Physics, DTU

19. ENERGY NETWORKS FOR STATE ESTIMATION WITH RANDOM SENSORS USING SPARSE LABELS, ***8.Yash Kumar*** and Souvik Chakraborty, Mechanical, DTU

20. EULER-RODRIGUES PARAMETERS: A QUANTUM CIRCUIT TO CALCULATE RIGID-BODY ROTATIONS, Emilio Peláez, Anuranan Das, ***6.Parmeet Singh Chani*** and Daniel Sierra-Sosa, Physics, DTU

21. Halophilic Alkaline Protease Enzyme on Biodegradation of Contaminants in Raw Textile Effluent, A. Yogeeswaran, R. Raja Jeya Sekar, S. Uma Maheswari, S.M. Vijila and ***3.T. Vijaya Kumar***, Environmental, DTU

22. IMPACT OF VARIATION IN SIZES OF BORON CARBIDE ON PROPERTIES OF NOVEL COMPOSITE OF ALUMINIUM ALLOY 6063-T6 AND BORON CARBIDE, ***Vibhu Singh***, ***6.Soni Kesarwani*** and ***3.M.S. Niranjan***, Mechanical, DTU

23. Implementation of GF-HOG Technique for Effective Commercial and Industrial Load Clustering and Classification for Better Demand Response, ***3.Aniruddha Bhattacharya*** and ***Madhusudan Singh***, Electrical, DTU

24. Investigations on Plasma Pretreatment of Catalyst Film and Catalyzed Growth of Carbon Nanotubes, ***6.Umang Sharma*** and ***3.Suresh C. Sharma***, Physics, DTU

25. Learning to Censor by Noisy Sampling, Ayush Chopra, ***8.Abhinav Java***, Abhishek Singh, Vivek Sharma, and Ramesh Raskar, CSE, DTU

26. Microstructural and mechanical behavior of micro-sized SiC particles reinforced friction stir processed/welded AA7075 and AA6061, ***8.1.Sumit Jain*** and ***3.R. S. Mishra***, Mechanical, DTU

27. Natural protein-based electrospun nanofibers for advanced healthcare applications: progress and challenges, Anushka Agarwal, Gyaneshwar K. Rao, Sudip Majumder, Manish Shandilya, Varun Rawat, ***3.Roli Purwar***, Monu Verma and Chandra Mohan Srivastava, Chemistry, DTU

28. Novel composite multilayer face masks for protection against airborne microorganisms, ***6.Priya Bansal***, ***7.Archana Gangwar***, ***8.Aman Verma*** and ***3.Roli Purwara***, Chemistry, DTU

29. Numerical modelling of Flow characteristics over sharp crested triangular and W-shaped weirs for establishing downstream scour behavior, ***7.Stuti Mishra***, ***3.Dr. Munendra Kumar*** and ***6.Vijay Kaushik***, Civil, DTU

30. Numerical study of microscale passive mixing in a 3-Dimensional spiral mixer design, ***Akash Sinha***, ***M. Zunaid***, ***Sulekh Tokas*** and Mubashshir Ahmad Ansari, Mechanical, DTU

31. Obscenity Detection in Videos through a Sequential ConvNet Pipeline Classifier, ***8.Neil Gautam*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

32. Oppositional Chaos Game Optimization Based Clustering with Trust Based Data Transmission Protocol for Intelligent IoT Edge Systems, M.Padmaa, T. Jayasankar, S. Venkatraman, Ashit Kumar Dutta, ***7.1.Deepak Gupta***, Shahab Shamshirband and Joel J. P. C. Rodrigues, DTU

33. Palmprint Biometric vs Encrypted Hash based Digital Signature for Securing DSP Cores Used in CE systems, Rahul Chaurasia, Aditya Anshul, Anirban Sengupta and ***6.Shikha Gupta***, CSE, DTU

34. Prioritizing Scheduling Parameters in the Automotive Industry Using Fuzzy TOPSIS-DEMATEL Model, ***6.Sumit Chawla*** & ***3.Ranganath M. Singari***, Mechanical, DTU

35. Projective Fisher Information for Natural Gradient Descent, ***8.1.Piyush Kaul*** and ***8.1.Brejesh Lall***, Electronics, DTU

36. Representation of Illness and Death in Bollywood: A Comparative Analysis of Devdas and Dil Bechara, ***6.Priya Meena*** and ***3.Rajiv Ranjan Dwivedi***, Humanities, DTU

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

38. Sentiment Analysis using Scikit Learn: A Review, ***8.Divij Gera*** and Amita Kapoor, CSE, DTU

39. Stator Inter-Turn Fault Diagnosis by Motor Current Signature Approach, ***6.Kuldeep Kumar Swarnkar***, ***3.Prof. J.N.Rai*** and Dr. Sulochana Wadhwani, Electrical, DTU

40. Techno-Economic and Environmental Analysis of Grid-Connected Electric Vehicle Charging Station Using AI-Based Algorithm, ***6.Mohd Bilal***, Ibrahim Alsaidan, Muhannad Alaraj, Fahad M. Almasoudi and ***3.Mohammad Rizwan***, Electrical, DTU

41. TRAFFIC VOLUME STUDY AND CONGESTION SOLUTIONS, ***8.Nishant Singh***, ***8.Nitin Chauhan***, ***8.Nishant Pande*** and ***3.Narad Muni Prasad***, Civil, DTU

42. Transductive Instance Transfer Learning for Cross-Language Defect Prediction, ***8.Rahul Chopra***, ***8.Shreoshi Roy*** and ***3.Ruchika Malhotra***, Software Engineering, DTU

43. UAVs assisted Network Partition Detection and Connectivity Restoration in Wireless Sensor and Actor Networks, Aditi Zear and ***3.Virender Ranga***, IT, DTU

44. Wind Excited Action around Tall Building Having Different Corner Configurations, ***6.Rahul Kumar Meena***, ***3.Ritu Raj*** and ***3.S. Anbukumar***, Civil, DTU

45. Word Sense Disambiguation using Cooperative Game Theory and Fuzzy HindiWordNet based on ConceptNet, ***3.GOONJAN JAIN*** and D. K. LOBIYAL, Mathematics, DTU

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

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

A Novel Approach for Early Recognition of Cataract using VGG-16 and Custom User-based Region of Interest

Aaron Paul Jacob
Department of Software Engineering, Delhi Technological University
aaronpauljacob_2k18se003@dtu.ac.in

Aryan Bansal
Department of Software Engineering, Delhi Technological University
aryanbansal_2k18se038@dtu.ac.in

Ruchika Malhotra
Head of Department, Department of Software Engineering, Delhi Technological University
ruchikamalhotra@dtu.ac.in

ABSTRACT

Cataract is a severe eye disease and is one of the main causes of vision impairment across the world. There have been many researches about cataract prediction using machine and deep learning state-of-the-art methods but all of these methods rely on automatic image classification without using the knowledge of the clinicians. This study proposes a novel approach that incorporates artificial intelligence with the human intelligence of experienced medical professionals. Our model allows manual selection of the region in the fundus image that is most likely to have cataract and gives it more importance as compared to the rest of the region in the image. After that, a VGG-16 model was applied to classify the input images. Accuracy was chosen as the performance metric and the maximum accuracy observed for the model was 98.83%.

CCS CONCEPTS

• **Computing methodologies** → Artificial intelligence; Computer vision; Computer vision problems; Object recognition; Machine learning; Learning paradigms; Supervised learning; Supervised learning by classification.

KEYWORDS

Cataract Recognition, Region of Interest, Deep learning, Machine Learning, VGG-16

ACM Reference Format:
Aaron Paul Jacob, Aryan Bansal, and Ruchika Malhotra. 2022. A Novel Approach for Early Recognition of Cataract using VGG-16 and Custom User-based Region of Interest. In *2022 4th Asia Pacific Information Technology Conference (APIT 2022), January 14–16, 2022, Virtual Event, Thailand*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3512353.3512356

1 INTRODUCTION

According to the World Health Organisation (WHO), around 2.2 billion people in the world have vision impairment and blindness. One of the major and leading cause of this are cataracts. It is to be emphasized that there are around 1 billion people out of those 2.2 billion people where cataract could have been treated if proper attention and early treatment were provided to them. It is clear from these statistics that early recognition of cataract could be so useful for the masses.

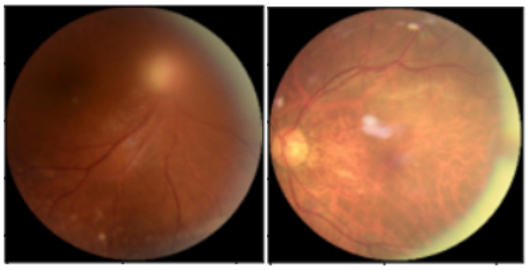


Figure 1: Fundus Images of a patient from ODIR database

A fundus image is a projection of the fundus captured by a monocular camera on a 2D plane. Unlike other eye scans, such as OCT images and angiographs, fundus images can be acquired in a non-invasive and cost-effective manner, making them more suitable for large-scale screening [1]. An example of a fundus image is provided in Figure 1

In this study, we have proposed a novel approach to classify fundus images as cataract or normal. Our approach combines the deep learning approach with the human intelligence associated with medical professionals to predict the outcome class for the input fundus image. The proposed model will enable the clinicians to select the area of interest in the fundus image that could possibly be having cataract and will give more importance to that region of interest. We ardently believe that by incorporating this idea the accuracy of a cataract detection system can be significantly increased.

The primary aim of the research is to develop a Manual-Selection Convolutional Neural Network (MS-CNN) architecture that enables the user to select the region of interest and assign a higher priority to that region as compared to the rest of the region of the fundus image. The model should then be able to classify the input image based on whether the fundus image contains cataract or is a normal eye image. The aim should not be confused with classification based on automatic cataract detection models that are previously employed to classify the fundus images.

In recent years, CNNs have emerged as the de facto standard for solving image classification problems, achieving error rates lower than 4% in the ImageNet challenge [5]. CNN is capable of extracting and learning the relevant features of the image and hence for our research problem, CNN is the best choice.

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.
APIT 2022, January 14–16, 2022, Virtual Event, Thailand
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9557-1/22/01…$15.00
https://doi.org/10.1145/3512353.3512356

2 RELATED WORK

There has been much research in this field until now. The majority of the work focuses on automatic classification of retinal diseases like glaucoma [8], keratoconus [9], macular degeneration [7] and diabetic retinopathy [4]. Different deep learning and machine learning models have been applied to classify the fundus images.

Gour et al. [15] multi-class multi-label classification of eye disorders in fundus images. The dataset used for the research study was Ocular Disease Intelligent Recognition (ODIR) database. "Four different pre-trained CNN architectures with two different optimizers are used and it is observed that VGG16 pre-trained architecture with SGD optimizer performs better for multi-class multi-label fundus images classification on ODIR database." The validation accuracy achieved was 87.16% with an AUC score of 84.93% using the VGG-16 architecture and SGD optimizer.

Junayed et al. [16] proposed a CataractNet model for the classification of fundus images as healthy or cataract-prone. Adam optimizer was used in the proposed methodology and the activation function and loss were tuned so that the model is trained with small kernels. Average accuracy of 99.13% was achieved on the 4746 augmented images that were used in the study. For the Kaggle dataset, the accuracy was observed to be 98.62%.

Junjun et al. [3] used a Dense Correlation Network (DCNet) to prepare a multi-label ocular disease classification model. The DCNet consisted of three different components: ResNets of different depths were used as the backbone CNN, a spatial correlation module and a classifier. A maximum AUC score of 0.930 was observed using a ResNet-101 model along with the SCM module.

3 PROPOSED METHODOLOGY

The entire proposed methodology is depicted as a flow diagram in Figure 2.

4 EXPERIMENTATION

We have developed a Manual-Selection Convolutional Neural Network (MS-CNN) architecture that enables the user to select the region of interest and assign a higher priority to that region as compared to the rest of the region of the fundus image. The model is then able to classify the input image based on whether the fundus image contains cataract or is a normal eye image.

4.1 Dataset

The proposed model was tested on images taken from the Ocular Disease Intelligent Recognition (ODIR) dataset which is available on Kaggle. The dataset consists of about 7,000 coloured fundus images of 5,000 patients that were taken from different hospitals and medical centers of China. The dataset has classified the fundus images in eight different classes including the normal and cataract class.

For our research, we used the cataract labelled images of both the left and right fundus of the eyes along with the images of a perfectly normal fundus.

There were a total of 588 cataract fundus images present in the dataset. In order to create a balanced dataset, 688 normal eye fundus images (i.e. not having cataract) were used. Hence there were a total of 1276 images used for the research. The data was then split into training and testing sets with a validation split of 20%. Hence, there were 1020 images in the training set and 256 images in the testing set.

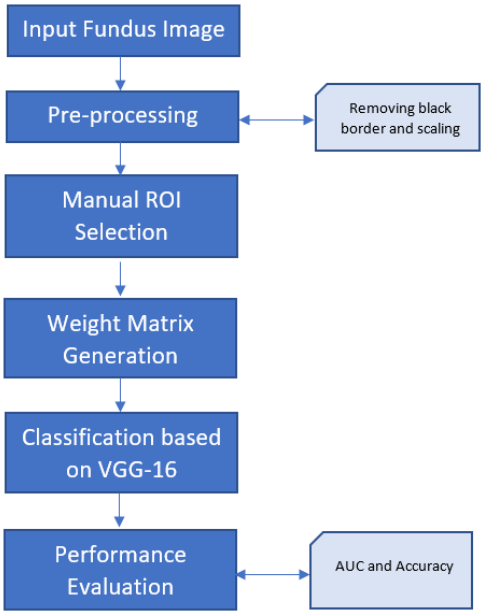


Figure 2: Flow Diagram of the Proposed Methodology

4.2 Data Pre-processing

The data preprocessing consisted of majorly two steps. Firstly, the dataset images were preprocessed by removing black border paddings. Secondly, the processed images were scaled to 224x224 pixels, containing 3 channels for RGB colors.

4.3 Extraction of Region of Interest

For the purpose of giving higher priority to the custom-defined region of interest, a weight matrix generator program was used. A 2D weight matrix was created that contained higher values for the pixels present in the manually selected region of interest. All the values were between [0.5,1.0]. The weight matrix when multiplied with an image matrix, highlighted and enhanced the values of the corresponding pixels as shown in Figure 3

All the images of normal and cataract fundus in the dataset were swapped in the background and the regions of high-visual importance were highlighted to increase the corresponding weights for that region in the weight matrix of that image and a new set of images were generated with the effect of the custom-selected region by the user as shown in Figure 4. Each image in the dataset was multiplied by the weighted filter to provide enhanced pixel values to the region of interest. Some random image examples after this step are provided in Figure 5

4.4 Model Training

The model was trained with 224*224 sized images that were generated in the previous step. Adam optimizer was used and the final

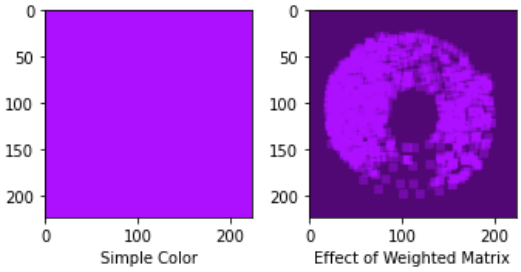


Figure 3: Image a) without weighted matrix and b) with weighted matrix of the selected region

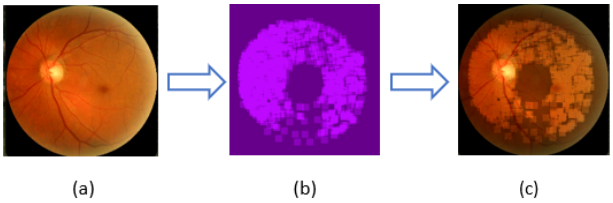


Figure 4: (a) Original Image (b) Weight Matrix of the original image (c) Weighted Image - capturing the selected region of interest

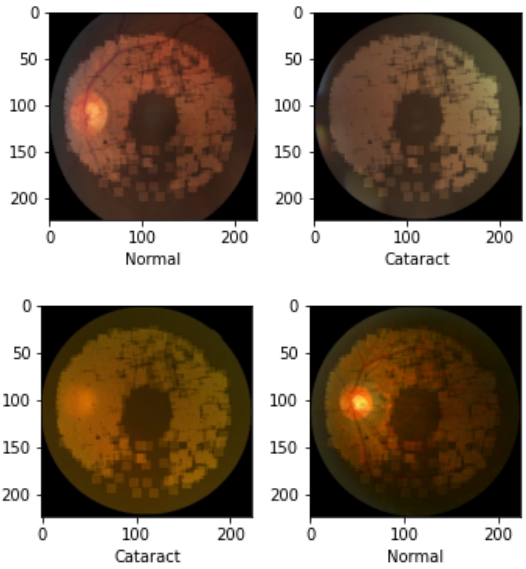


Figure 5: Fundus Images after being multiplied by the weighted filter

layer used sigmoid activation for predicting the output class. The model was trained with VGG-16 architecture with encoder weights set equal to ImageNet weights.

The loss function used is binary cross-entropy. This loss function compares the predicted probabilities of the binary classes in the problem to actual class output. Then the score is calculated based on the distance of the predicted probabilities from the expected value and penalizes them.

We have used early stopping and reduced learning rate on plateau callbacks to train the model and get the results quickly.

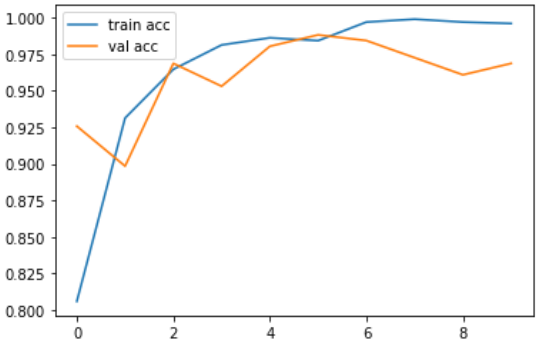


Figure 6: Accuracy vs Epoch Plot

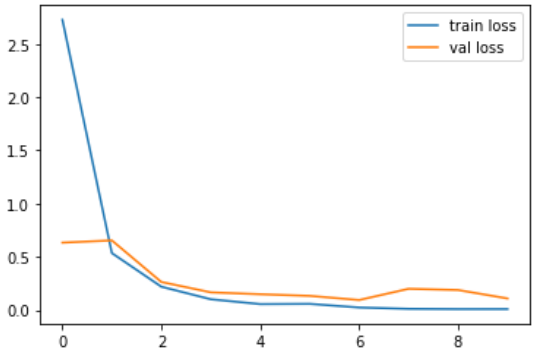


Figure 7: Loss vs Epoch Plot

4.5 Results and Analysis

Our proposed model showed some remarkable and record-breaking results. The performance metric chosen for the research was accuracy as we had a balanced dataset in hand. The maximum accuracy achieved using our novel weighted matrix approach was 98.83%. It evidently proved that the proposed approach performed better than the other state-of-the-art methods applied to solve the given problem. The accuracy and loss plots are presented in Figure 6 and Figure 7

While observing the Accuracy vs Epoch plot, in the initial stages of the training when the epoch is about 1, a better validation accuracy is observed as compared to the training set accuracy. This could be possible due to a number of reasons such as the model is underfit when epoch is one, or the validation set had less variance as compared to the training set variance from the best fit model. The same applies to the loss vs epoch plot since the validation set has lesser loss as compared to the training set loss.

5 CONCLUSION AND FUTURE WORK

The proposed model performed exceptionally well in classifying the fundus images. The accuracy achieved during the experimentation phase was significantly greater than the other state-of-the-art methods deployed for the classification problem. It is evident from the results that the proposed MN-CNN model can be deployed to solve the problem of early detection of cataract on a large scale.

There are some limitations to the study which can be covered in the future work. The research is carried out using the VGG-16 model only. Other deep learning models could also be implemented to compare the performance of the proposed approach with different models. The model will be tested more with various other architectures like ResNet, AlexNet, MobileNet, etc. Furthermore, we will compare the performance of the proposed MS-CNN model with the existing object detection models like Mask R-CNN, Fast R-CNN, etc. as these models automatically divide the image in various region proposals as compared to the manual selection done in our approach.

ACKNOWLEDGMENTS

We are highly grateful to the Kaggle platform for extending their valuable resources to us. The dataset used for the study was taken from the Kaggle platform. Early phase implementation of the research was carried out on the Kaggle platform. The specifications provided by the platform are Nvidia P100 16 GB GPU, 16 GB ram and 20 GB disk storage.

We would also like to extend our gratitude to Dr. Ruchika Malhotra mam for her active guidance throughout the completion of this project. The final research work implementation and manuscript writing was carried out under her constant support and guidance.

REFERENCES

[1] Edupuganti, V.G., Chawla, A., Kale, A., 2018. Automatic optic disk and cup segmentation of fundus images using deep learning, in: 2018 IEEE International Conference on Image Processing, ICIP 2018, Athens, Greece, October 7-10, 2018, IEEE. pp. 2227–2231.URL: https://doi.org/10.1109/ICIP.2018.8451753, doi: 10.1109/ICIP.2018.8451753.

[2] Linglin Zhang *et al.*, "Automatic cataract detection and grading using Deep Convolutional Neural Network," 2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC), 2017, pp. 60-65, doi: 10.1109/ICNSC.2017.8000068.

[3] Junjun He, Cheng Li, Jin Ye, Yu Qiao, Lixu Gu,Multi-label ocular disease classification with a dense correlation deep neural network,Biomedical Signal Processing and Control,Volume 63,2021,102167,ISSN 1746-8094,https://doi.org/10.1016/j.bspc.2020.102167.(https://www.sciencedirect.com/science/article/pii/S1746809420303062)

[4] Chorage, S., Khot, S.S.: Detection of diabetic retinopathy and cataract by vessel extraction from fundus images. In: 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), vol. 1, IEEE, pp. 638–641 (2017)

[5] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, 2015, doi: 10.1007/s11263-015-0816-y.

[6] Fan,W., Shen, R., Zhang, Q.,Yang, J.J., Li, J.: Principal component analysis based cataract grading and classification. In: 2015 17th International Conference on E-health Networking, Application and Services (HealthCom), IEEE, pp. 459–462 (2015)

[7] Güven, A.: Automatic detection of age-related macular degeneration pathologies in retinal fundus images. Comput. Methods Biomech. Biomed. Eng. 16(4), 425–434 (2013)

[8] Nayak, J.,Acharya,R., Bhat, P.S., Shetty,N., Lim, T.C.:Automated diagnosis of glaucoma using digital fundus images. J. Med. Syst.33(5), 337 (2009)

[9] Souza, M.B.,Medeiros, F.W., Souza, D.B.,Garcia, R.,Alves,M.R.:Evaluation of machine learning classifiers in keratoconus detection from orbscan ii examinations. Clinics 65(12), 1223–1228 (2010)

[10] Wang W, Han X, Zhang J, *et al.* Br J Ophthalmol Epub ahead of print: [please include Day Month Year]. doi:10.1136/ bjophthalmol-2020-318609

[11] A. Soni and A. Rai, "A Novel Approach for the Early Recognition of Diabetic Retinopathy using Machine Learning," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-5, doi: 10.1109/ICCCI50826.2021.9402566.

[12] Imran, A., Li, J., Pei, Y. *et al.* Fundus image-based cataract classification using a hybrid convolutional and recurrent neural network. Vis Comput 37, 2407–2417 (2021). https://doi.org/10.1007/s00371-020-01994-3

[13] V. Nair, S. Suranglikar, S. Deshmukh and Y. Gavhane, "Multi-labelled Ocular Disease Diagnosis Enforcing Transfer Learning," 2021 55th Annual Conference on Information Sciences and Systems (CISS), 2021, pp. 1-6, doi: 10.1109/CISS50987.2021.9400227.

[14] Ji-Jiang Yang, Jianqiang Li, Ruifang Shen, Yang Zeng, Jian He, Jing Bi, Yong Li, Qinyan Zhang, Lihui Peng, Qing Wang, Exploiting ensemble learning for automatic cataract detection and grading, Computer Methods and Programs in Biomedicine, Volume 124, 2016, Pages 45-57, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2015.10.007.

[15] Neha Gour, Pritee Khanna, Multi-class multi-label ophthalmological disease detection using transfer learning based convolutional neural network, Biomedical Signal Processing and Control, Volume 66, 2021, 102329, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2020.102329.(https://www.sciencedirect.com/science/article/pii/S1746809420304432)

[16] M. S. Junayed, M. B. Islam, A. Sadeghzadeh and S. Rahman, "CataractNet: An Automated Cataract Detection System Using Deep Learning for Fundus Images," in IEEE Access, vol. 9, pp. 128799-128808, 2021, doi: 10.1109/ACCESS.2021.3112938.

[17] Behera, Manoj & Chakravarty, Sujata & Gourav, Apurwa & Dash, Satyabrata. (2020). Detection of Nuclear Cataract in Retinal Fundus Image using RadialBasis FunctionbasedSVM. 278-281. 10.1109/PDGC50313.2020.9315834.

[18] SungMin Suh, Yongeun Park, KyoungMin Ko, SeongMin Yang, Jaehyeong Ahn, Jae-Ki Shin, SungHwan Kim, "Weighted Mask R-CNN for Improving Adjacent Boundary Segmentation", Journal of Sensors, vol. 2021, Article ID 8872947, 8 pages, 2021. https://doi.org/10.1155/2021/8872947

[19] Tao Li, Wang Bo, Chunyu Hu, Hong Kang, Hanruo Liu, Kai Wang, Huazhu Fu, Applications of deep learning in fundus images: A review, Medical Image Analysis, Volume 69, 2021,101971, ISSN 1361-8415,https://doi.org/10.1016/j.media.2021.101971.(https://www.sciencedirect.com/science/article/pii/S1361841521000177)

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-022-11927-y

A novel protocol for efficient authentication in cloud-based IoT devices

Irfan Alam [1] · Manoj Kumar [1]

Received: 14 July 2021 / Revised: 8 December 2021 / Accepted: 3 January 2022
Published online: 24 February 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
The Internet of Things (IoT) has emerged as one of the most revolutionary technological innovations with the proliferation of applications within almost all fields of the human race. A cloud environment is the main component of IoT infrastructure to make IoT devices efficient, safe, reliable, usable, and autonomous. Reduction in infrastructure cost and demand accessibility of shared resources are essential parts of cloud-based IoT (CIoT) infrastructure. Information leakage in cloud-assisted IoT devices may invite dangerous activities and phenomena. Various cloud-based systems store IoT sensor data and later on access it accordingly. Some of them are public, and some of them are private. Private cloud services must be secured from external as well as internal adversaries. Hence, there must be a robust mechanism to prevent unauthorized access to devices. This paper proposes a novel and efficient protocol based on the Elliptic Curve property known as Elliptic Curve Discrete Logarithm Problem (ECDLP) with hash and XOR functions for the authentication in cloud-based IoT devices. In comparison to the existing protocols, the proposed protocol is resistant to attacks and other security vulnerabilities. The one-way hash function and XOR function effectively ensure a reduction in computation cost. AVISPA and BAN logic have been used for formal analysis of the proposed protocol. As per the performance analysis results, it is clear that the proposed protocol is efficiently suitable for cloud-assisted IoT devices.

Keywords Authentication · Protocol · ECC · Attacks · XOR · AVISPA

1 Introduction

The Internet of Things (IoT) is a setup of different entities like objects, people, and animals capable of transferring data over networks without head-to-head or head-to-computer

✉ Irfan Alam
irfanrmamu@gmail.com

[1] Department of Computer Science and Engineering, Delhi Technological University, New Delhi 110042, India

connection. IoT is an acronym for Any time, Anything, Any place connections. In recent years, the Internet of Things (IoT) has risen to prominence as a critical technological component for overcoming interoperability, heterogeneity, and Internet-aware resistances.

The deployment of IoT devices has increased exponentially, resulting in a large amount of data handling and analysis [8]. There is a need for a standard platform to manage heterogeneous natured gadgets and data. Cloud-based infrastructures are well suited to this need. Ray, in his survey paper, describes applications for the IoT cloud platforms [39]. In general, cloud services are of two types: public cloud services and private cloud services. A person or organization owns private cloud services that make cloud security essential [53]. Analysis of the cloud-based IoT (CIoT) device mechanism [2, 7, 14, 36] shows that most applications carry essential information, and leakage of such information may lead to a significant loss and several attacks. Also, the nature of attacks changes from time to time. Computing resources are employed on-demand via a network in the Cloud services. It consumes a significant amount of energy, which does not measure easily with computational resources. Pete et al. [37] proposed an experimental energy-efficient model for Virtual Machines (VMs) scheduling over the Cloud; furthermore, Kumar et al. [27] illustrate an economically efficient model for virtualization over the Cloud using a docker container.

Authentication is a primary security mechanism for all network-based services that prevent unauthorized users or adversaries from gaining access [48]. It is compulsory to ensure that only authenticated users should access the resources. For information exchange through an unsecured network such as the Internet, IoT authentication can ensure the trust of IoT devices. A robust IoT authentication model is needed to protect unauthorized users and devices. Traditional public-key cryptosystems are commonly used in authentication procedures. Traditional public-key cryptosystems, such as RSA, have large key sizes and use a lot of computing resources [46]. As a result, most standard authentication systems are incompatible with the limited computational power of IoT devices. Recently, several research papers addressed IoT authentication and the possibility of attacks [4, 22, 30, 34, 51, 52]. However, many attacks are still unresolved, such as packet analysis, gateway forgery, and Sybil attacks. Choosing an appropriate protocol to address recent attacks is tough task, due to unavoidable resource constraints, such as low power support, small storage, low computational support, and low latency support. The proposed protocol is a solution for above said challenges. The significant contributions of this paper have been listed as follows:

- A novel protocol has been proposed for the IoT infrastructure based on a cloud environment.
- The proposed novel authentication protocol prevents adversaries from common attacks such as Man in the middle, Masquerade, Denial of Services (DoS), Forging, Guessing, and physical attacks.
- It can provide robust and secure authentication using the Elliptic Curve Discrete Logarithm Problem (ECDLP) property of ECC with hash and XOR functions. NP-hard complexity of ECDLP makes the protocol resistant to attacks such as Forging, guessing, and Man in the middle, and the inclusion of XOR and hash keeps it lightweight.
- The proposed protocol has been compared with the existing protocols to explore its ability to resist attacks, reduce computation power and storage requirements, etc. Its computational cost is very low (0.2 s) and resists more than fifteen modern attacks. It also shows its usefulness more than other existing protocols.

- The modular structure of protocols varies from application to application. We have tried to modal the proposed protocol in such a way so that it suits 90% of IoT infrastructures.
- Simulation of the proposed protocol using AVISPA [5, 44] and BAN logic [43, 50] proves its power against active adversaries.
- Description of BAN logic and AVISPA in the present article would allow readers to understand and apply these tools for other protocols easily.

The rest of the paper deals with the following way - Section 2 discusses the related work, followed by section 3, which describes the Preliminaries. Section 4 presents the details of the proposed protocol. Section 5 discusses the performance and security analysis. Finally, the paper presents the conclusion and future work in section 6.

2 Related works

The selection of proper methodology for securing the flow of information plays a vital role in secured and safe communication. This selection varies for different applications, such as polynomial congruence is used for multimedia encryption over the Cloud [26]. Similarly, the secret-sharing concept has been used to encrypt video [31, 40]. The proposed protocol meets the requirements of real-time applications by minimizing the required time and enhancing the efficiency of the encryption process. Similarly, Gadicha et al. [17] present a multimode approach of data encryption in images through Quantum Steganography. This work opens a new chapter in modern quantum cryptography dealing with images in multi-modal aspects.

In general, Confidentiality, Integrity, and Availability (CIA) are the main features to be taken care of in the case of traditional security. In contrast, in the case of IoT, security measures vary from application to application. In 2020, Iqbal et al. [21] assessed IoT security requirements, challenges, and remedies in-depth. From there, it is clear that authentication is mandatory for almost all applications related to cloud-based IoT infrastructure.

In their survey paper, Nandy et al. [35] reviewed all aspects of IoT authentication. It also explains the taxonomy of all possible attacks on IoT authentication, factors involved in choosing the authentication technique, and the tools for simulation of authentication protocols for IoT devices. Authentication protocols are usually based on three parameters: biometric features of users, credentials, and smart cards. In 1981, Lamport [29] proposed a password-based authentication protocol for untrusted networks. However, a stolen-verifier attack was found in password table-based protocols at the server end. After that, numerous protocols have been proposed using a single or combination of the techniques such as Smart card, Password, and Biometrics. In 1985, Victor Saul Miller [33] and Neal I. Koblitz [19] came up with the elliptic curve in cryptography. Elliptic curve cryptography (ECC) algorithms have been used since 2004 widely due to their small size and low computational requirements with the same security level as RSA [21, 48]. In [10, 23, 28, 38], authors have worked with the Computational hardness of ECC for IoT devices. In [24], Kim et al. have done cryptanalysis of the hash function. Protocols like [22, 42, 52], and others have used hashed-based algorithms, which are helpful to ensure the integrity of the messages and verify the sender as well.

In 2015, Li et al. [30] proposed a mutually authenticated protocol based on a smart card in cloud computing. Their protocol does not persist against physical attack, forging attack, and masquerade attack [50]. Furthermore, Sun et al. [43] proposed a novel remote user authentication and key agreement protocol for the mobile client-server Environment. Still, it does not provide a pleasant process of password update and biometric change [50].

After that, in 2016, an efficient authentication protocol was proposed for IoT-enabled devices in a distributed cloud environment by Amin et al. [4]. In this work, different attacks has been discussed, such as user anonymity, insider attack. In 2018, Challa et al. [26] analyzed that Amin et al. protocol [4] does not resist masquerade attacks such as privileged-insider attacks and impersonation attacks. An adversary can send a valid login message on behalf of the user. In 2018, Wu et al. [51] explained that Irshad et al. [22] and Amin et al. [4] are vulnerable to Privileged Insider (PI) attacks and Offline Password Guessing (OPG) attacks, respectively. They both do not guarantee User Anonymity (UA).

Recently, Wu et al. [52] analyzed Xu et al. protocol [51]. They concluded that Xu et al. protocol could not resist Privileged Insider (PI) and stolen Smart Card (SSC) attacks and did not provide pre-verification and Perfect Forward Secrecy (PFS). If we look deep into the hash functions, we can get an exciting result. If the attacker replaces the contents of two messages, such as msg1 with msg2, the receiver will not detect this change since both messages generate the message same hash. This defect may allow adversaries to play with the networks that lead to various attacks such as user and sensor impersonation attacks and guessing attacks. In this scenario, the protocols discussed above may not be able to resist modern attacks.

Considering the shortcomings of the above-discussed protocols, we have proposed this protocol. It uses ECC along with the hash and XOR operators. ECDLP makes the proposed protocol resist the above-discussed attacks and other recent attacks due to its NP-hard complexity. The use of hash and XOR functions keeps the proposed protocol computationally lightweight.

3 Preliminaries

3.1 Elliptic curve discrete logarithm problem (ECDLP)

For two points, P and Q, of an elliptic curve $E_P(a,\ b)$ the **ECDLP** is to find an integer $k \in [1,\ n\ -\ 1]$ such $Q\ =\ k.\ P$.

ECDLP is more complicated than most discrete logarithm and factorization problems in cryptography. There are several attempts to break this problem; however, it is still unbreakable.

3.2 One-way cryptographic hash function

Here we describe the brief properties of the hash function, which are as follows

- This function is deterministic h :$\{0,\ 1\}^* \rightarrow \{0,\ 1\}^n$
- h is the hash function, where output produces y = h(x) (hash digest)$y \in \{0,\ 1\}$.
- Any tiny modification to the input string potentially leads to an entirely different hash value (message digest).

3.3 System model

The modular architecture of IoT-based systems may vary from application to application. We have tried to consider the most common approach suited for 90% of IoT infrastructure. As shown in Fig. 1, the main components of the proposed system are following:

- User (U),
- Cloud Server (S_m) directly associated with IoT devices, and
- Trusted authority (TA) sometimes called a control server

Both user (U) and cloud server (S_m) must register with TA before U can access (S_m) data via a secure channel. Necessary Credentials for U and (S_m) are generated by a Trusted Authority (TA) and stored in their memory. A session key is established (After mutual authentication between U and (S_m) for future independent communication.

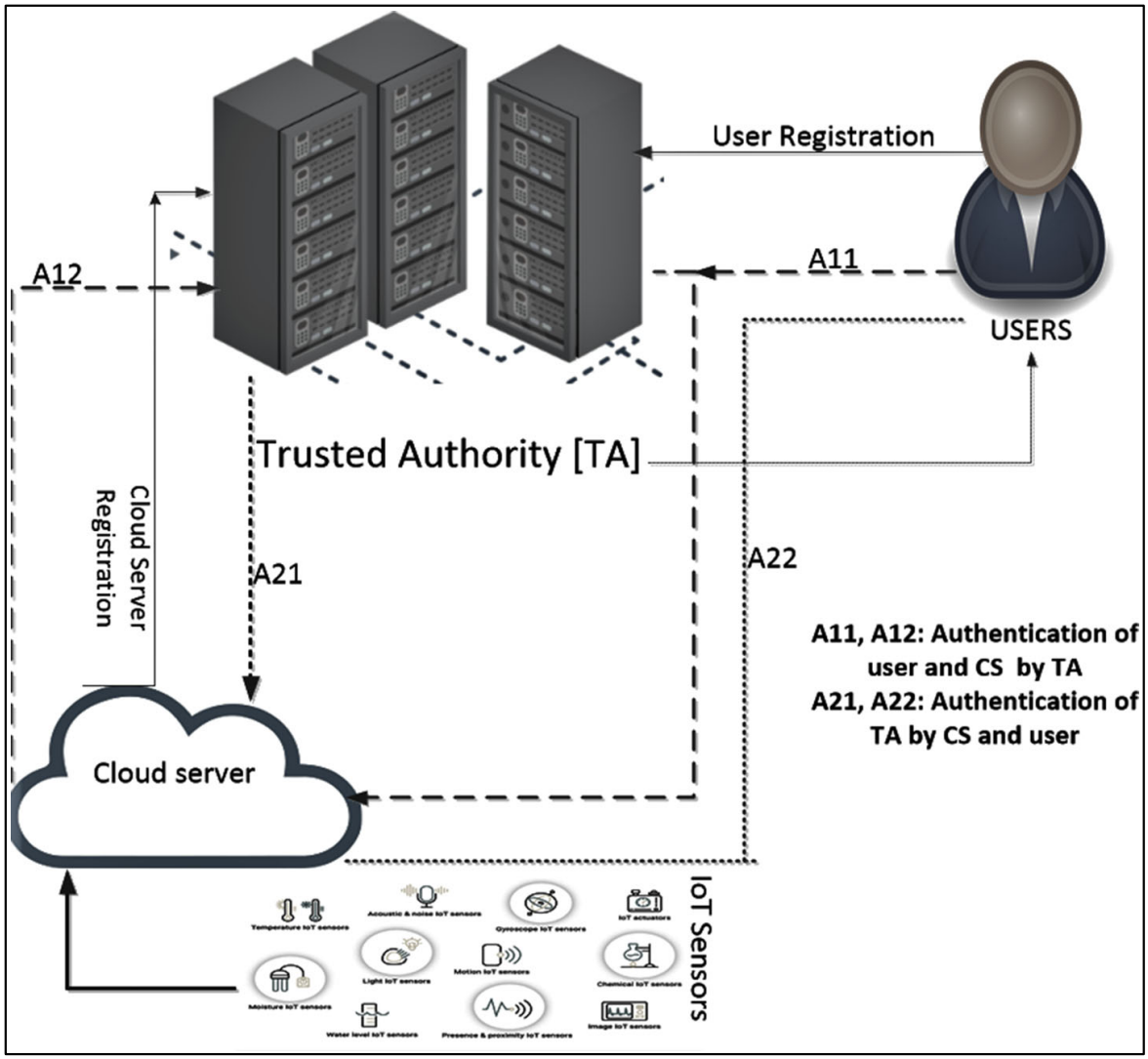


Fig. 1 Cloud-based IoT environment authentication model

3.4 Adversary model

If there is a protocol for some security purpose, there must be some adversary model against which the protocol is safe. A fundamental and most common adversary model was proposed by Dolev and Yao [16] in 1983. It is a broadly accepted model. According to this model, an adversary can read, alter and perform decryption (if got right keys) on messages. Adversaries are unable to carry out any statistical or cryptanalytic assaults. Due to rapid transformation in technologies, adversaries are now more advanced and have extraordinary capacities. Along with the Dolev and Yao model, we have considered that adversaries would know critical information from smart cards (if lost/stolen) using power analysis of smart cards [25, 32]. An adversary would also extract data from network flow using a network analyzer and recent AI techniques [15, 54].

4 Proposed protocol

This section presents an ECC-based protocol designed mainly to ensure that the devices are accessible in a secure manner on public channels. Cloud involvement in all entities and clock synchronization is our assumption for the proposed protocol. In Table 1, we listed necessary and frequent notations with their explanation for the proposed protocol. The proposed protocol comprises different phases, viz., pre-deployment, registration, login, and authentication. In the pre-deployment phase, the setup of the elliptic curve is discussed with the declaration of public and private keys. Registration of cloud server and user have been described in sections 4.2 and 4.3, respectively. After successful registration, verification of the user is done in the login phase via smart card by a trusted authority. In section 4.5, firstly, user authentication proceeded by cloud server authentication is done by a trusted authority. Later, trusted authority is authenticated by both user and cloud server consecutively. At the same time, a session key is established for further communications between the user and the cloud server.

Table 1 Notation used in the proposed protocol

Explanation	Notation
An elliptic curve	$E_P(b_1, b_2)$
Cloud server	S_m
Identity of the i^{th} User	U_i
Trusted authority	TA
Password of User	p_u
Card Reader	CR
Z_p= {0, 1, · ·, p−1}, a prime finite field	Z_p
Transmission delay	ΔT
Timestamps of CS, User and TA, respectively	TS_m, TS_u, T_{TA}
Pair of public key, private key of TA	$\{Q_{TA}, d_{TA}\}$
Concatenation operation	$\|\|$
Hash functions	H (.)
Session key	SK
XOR function	$\oplus$

4.1 Pre-deployment phase

An elliptic curve $E_P(b_1, b_2)$of non-singular nature over a finite field Z_p is selected by TA, where P is a large prime and $\mathbf{4}b_1^3 + \mathbf{27}b_2^2 \neq \mathbf{0}(mod\ p)$. At the same time, a base point P of order n over $E_P(b_1, b_2)$ where TA also chooses n.P = O, and a private key$d_{TA} \in Z_p$. Then TA computes the public key $Q_{TA} = d_{TA}$.P corresponding tod_{TA}. TA chooses the one-way cryptographic hash function. In the same way, cloud servers and the user also compute private keys and public keys using ECC properties.

4.2 Registration of cloud server

1. Cloud server S_m chooses identity (sid_m),$d_{cs} \in Z_{cs}$and $Q_{cs} = d_{cs}.P$ after that sends {sid_m, d_{cs} and Q_{cs}} securely to TA as shown in Fig. 2.
2. TA computes$psid_m$= h ($sid_m||d_{cs}$) and$B_{sm} = h(psid_m||\ Q_{TA})$.
3. TA sends B_{sm}toS_m, which saves B_{sm} and d_{cs} in the server memory.

4.3 User registration phase

1. User chooses identity (U_i), password (P_u), $d_u \in Z_P$ and $Q_u = d_u.P$ as shown in Fig. 3.
2. User calculate $A_u = h(p_u||d_u)$, $PID_u = h(U_i||Q_u)$ *and* $bb_u = Q_u \oplus A_u$
3. Then user send A_u and PID_u to a Trusted Authority (TA)
4. After receiving A_u and PID_u, TA calculates $h(A_u||PID_u)$, $D_u = h(PIDu||d_{TA})$ *and* $E_u = D_u \oplus A_u$
5. After step 4, TA send [C_u, E_u] to the user
6. User again calculate $DP_u = h(U_i||P_u) \oplus d_u$
7. User saves [C_u, E_u, DP_u]in the smart card.

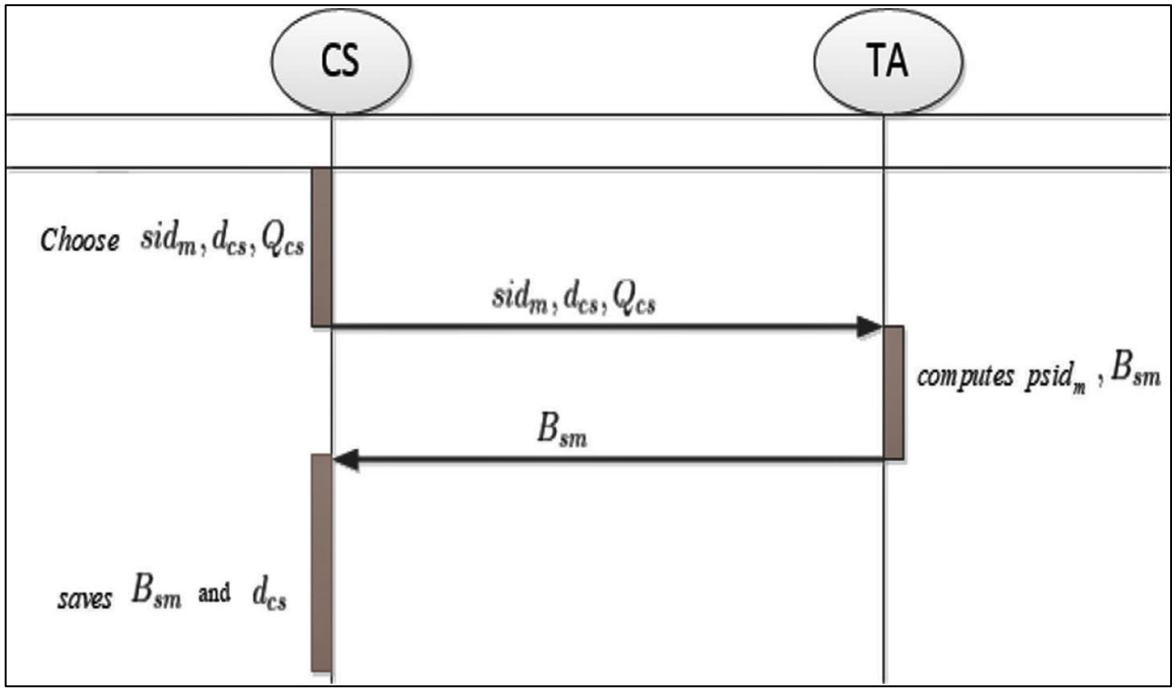


Fig. 2 Cloud Server Registration

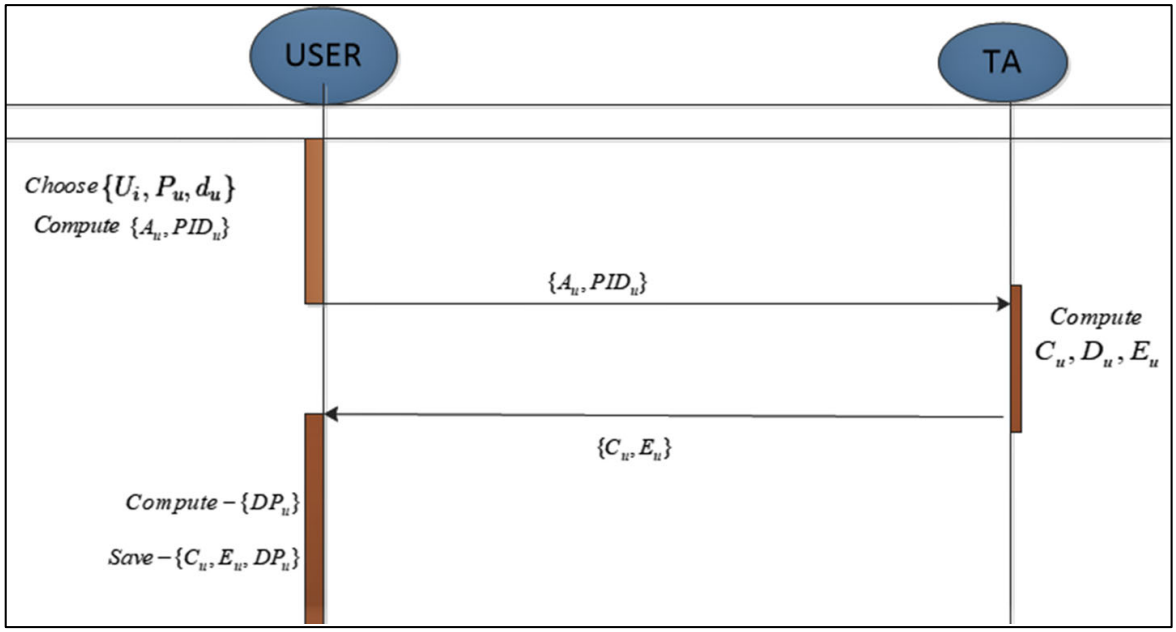


Fig. 3 User Registration

8. Finally, a smart card holds [C_u, E_u, DP_u, $h(.)$, *ECC parameters*]

4.4 Login phase

1. Firstly, a smart card is punched into the card reader by the legal user for accessing server resources, and then the user provides Id and password (U_i*, P_u*) in the card reader terminal, as shown in Fig. 4.

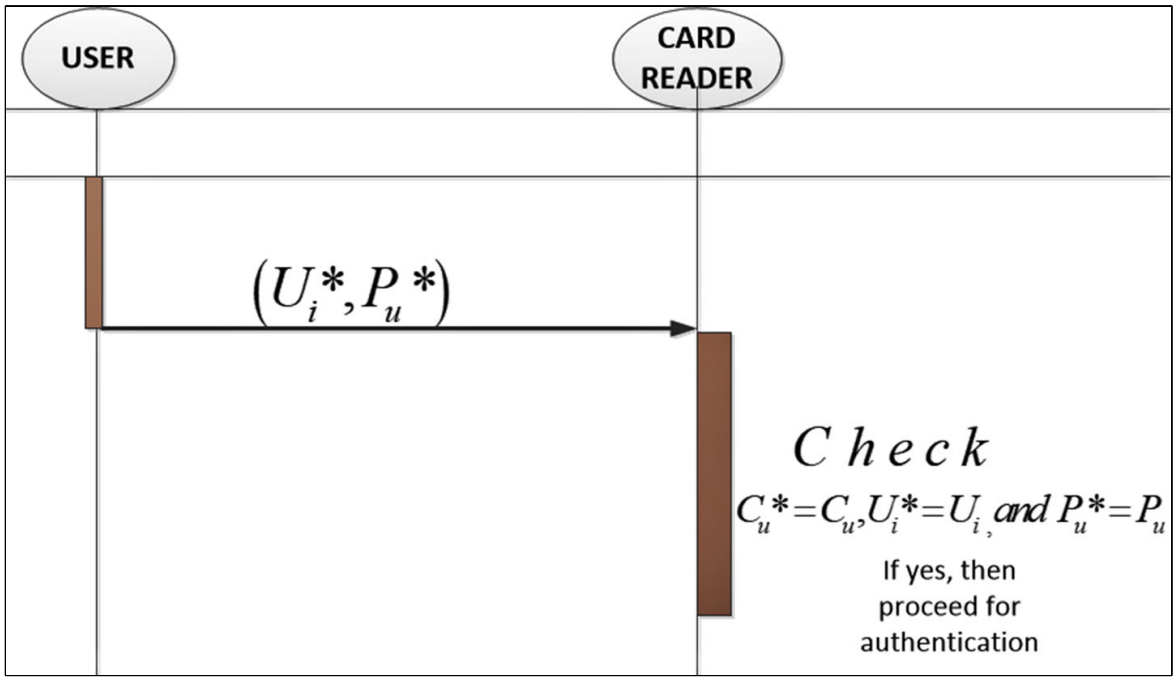


Fig. 4 pre-verification from card reader

2. After first step card reader calculates the following

(i)

$$d_u* = DP_u \oplus h\left(U_i*\left(P_u\right)\right) \quad and \quad A_u* = h\left(p_u*\left(d_u*\right)\right)$$

(ii)

$$Q_u* = bb_u* \oplus A_u*$$

(iii)

$$PID_u* = h\left(U_i*\left(Q_u*\right)\right)$$

(iv)

$$C_u* = h\left(A_u*\left(PID_u*\right)\right)$$

Card reader checks the condition $C_u * = C_u$, $U_i * = U_i$,*and* $P_u * = P_u$ if satisfies, then proceed for an authentication phase.

4.5 Authentication phase

After the login phase, the smart card produces a nonce N_u, Calculate $D_u = E_u \oplus A_u$, $F_u = D_u \oplus N_u$, *and* $Z_u = SID_m \oplus h(D_u \| N_u)$ and send $[F_u, Z_u, PID_u, TS_u]$,to the cloud-serverS_m, here (SID_m) is the cloud server identity chosen by the user and TS_u is the user's timestamp.

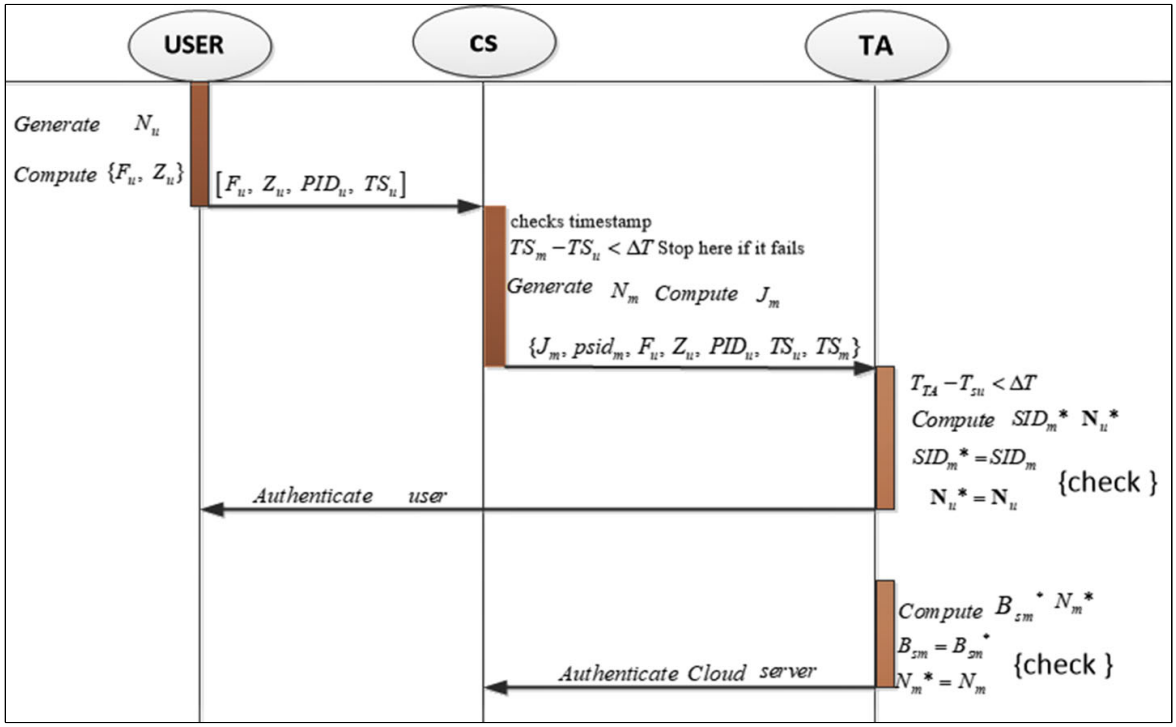


Fig. 5 Authentication of User and Cloud Server

The following steps are done on the Cloud server-side, as shown in Fig. 5.

1.1 At the cloud server, it checks timestamp $TS_m - TS_u < \Delta T$ (Stop here if it fails, where TS_m is a current timestamp and ΔT is the time interval to be expected during message transmission, respectively)
1.2 Cloud server produces 128 bits nonce N_m .

1. 3 Calculates $J_m = B_{sm} \oplus N_m$ and sends $[J_m, psid_m, F_u, Z_u, PID_u, TS_u, TS_m]$ to TA.
2. After receiving the above keys, TA confirms the validity of timestamp $(T_{TA} - T_{su} < \Delta T)$ and calculates$N_u * = F_u \oplus D_u$, $SID_m * = Z_u \oplus D_u$ and check $SID_m * = SID_m$ and $N_u^* = N_u$. If this is ok then TA authenticates the user as the legal user.
3. After authenticated User, TA computes $B_{sm} = h(psid_m \| Q_{TA})$ and $N_m^* = B_{sm}^* \oplus J_m$ and checks $B_{sm}^* = B_{sm}$ and $N_m * = N_m$ if it is ok, TA authenticates the Cloud server.
4. After Authenticated User and Cloud server, TA chooses a 128-bit nonce $N_{TA}$and computes the following:

(v)

$$P_{TA} = N_m \oplus N_{TA} \oplus h\left(N_u\left(D_u\right)\right.$$

(vi)

$$SK_{TA} = h(N_u \oplus N_m \oplus N_{TA})$$

(vii)

$$R_{TA} = N_u \oplus N_{TA} \oplus h\left(B_{sm}^* \| N_m^*\right)$$

(viii)

$$Z_{TA} = h\left((N_m \oplus N_{TA})\left(\quad SK_{TA}\right)\right.$$

(ix)

$$V_{TA} = h\left(\left(N_u\left(\quad N_{TA}\right)\left(\quad SK_{TA}\right)\right)\right.$$

Finally SK_{TA} is the secret session key.

The TA sends $[P_{TA}, R_{TA}, Z_{TA}, V_{TA}]$to the S_m through public communication.

5. After receiving from TA the S_m computes following:

(x)

$$W_m = h\left(\mathbf{B_{sm}}\left(\quad N_m\right)\right., N_u \oplus N_{TA} = R_{TA} \oplus W_m, SK_m = h(N_u \oplus N_{TA} \oplus N_m)$$

(xi)

$$and \quad V_{TA}* = h\Big((N_u \oplus N_{TA})\Big(\quad SK_m\Big)$$

Then, S_m check the condition ($V_{TA}* = V_{TA}$). If yes, then proceed and send [P_{TA}, Z_{TA}] it to the user publicly.

6. On obtaining [P_{TA}, Z_{TA}] fromS_m, the user calculates the following:

$L_u = h(N_u \| D_u)$, *and* $SK_u = h(N_m \oplus N_{TA} \oplus N_u) Z_{TA}* = h((N_m \oplus N_{TA}) \| SK_u)$

A user checks the condition($Z_{TA}* = Z_{TA}$) *and* $N_m \oplus N_{TA} = P_{TA} \oplus L_u$, and an affirmative answer proves that S_m and TA are authentic.

Finally, we see that (as shown in Fig. 6)

- Mutual authentication is achieved among users S_m and TA.
- Session key SK_m=SK_u are established for future communication.

5 Security analysis

To demonstrate and validate the security and correctness of our proposed protocol, we use BAN logic [3, 9, 18, 41, 43, 50], followed by Automated Validation of Internet Security Protocols and Applications (AVISPA) [5, 6, 13, 45]. Furthermore, an informal discussion has also been done in the latter part of this section.

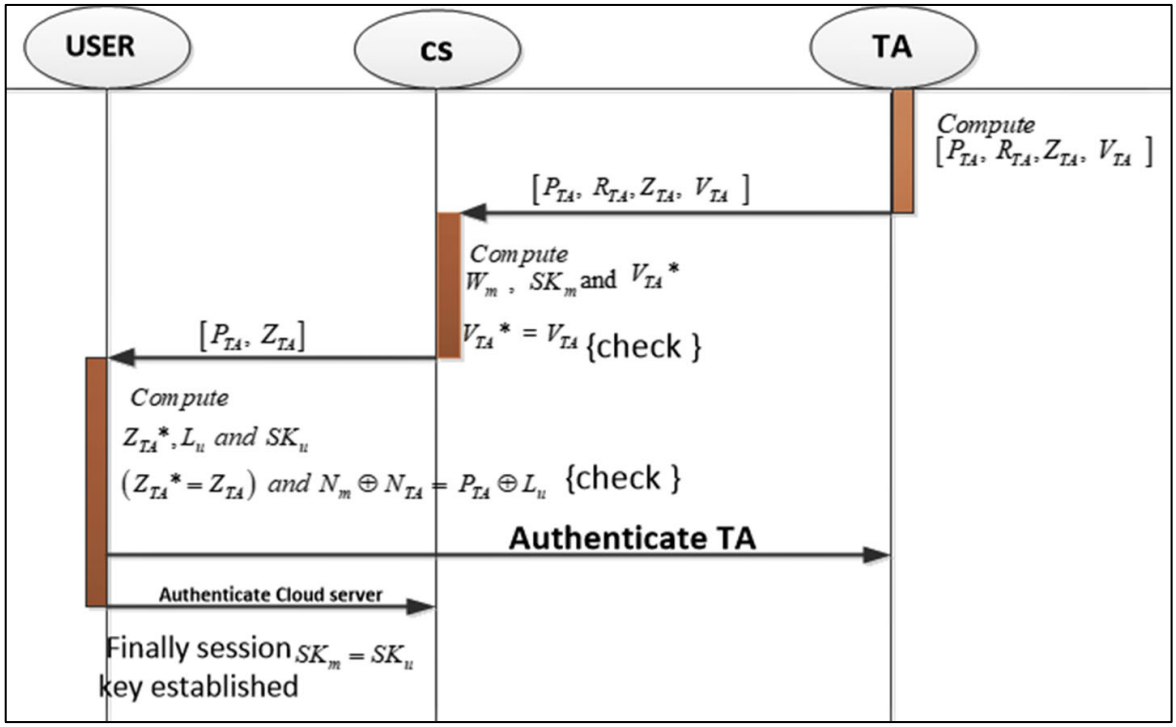


Fig. 6 TA authentication

5.1 BAN logic (a formal approach of security analysis)

It is very much essential to ensure the correctness of the authentication protocol. In Feb. 1990, MICHAEL BURROWS, MARTIN ABADI, and ROGER NEEDHAM came logically to verify the correctness of an authentication protocol, efficiency, and applicability, termed BAN Logic. Although detailed descriptions are illustrated by M. Burrows et al. [9] and more such as [3, 18, 41, 43, 50], we have tried to summarize BAN logic to understand the main concepts of BAN logic.

By using BAN logic, we can get the following important information about an authentication protocol:

1. The purpose of each protocol (Goal).
2. The cryptosystem that was used.
3. Whether secrets are used or not (other than the key).
4. Is there a guarantee that messages will arrive on time?
5. Whether protocol establishes each party's presence.
6. To remove redundancy.

The main goal of our protocol is that a user (U), Cloud server (S_m), and Trusted authority (TA) must authenticate one another. A session key (SK) is established for further communication between the server and the user.

Following are the steps used in the proposed protocol. Sentences with bold letters are common in all authentication protocols.

In First step, goals are written

G1: U believes $U \Leftrightarrow^{SK} S_m$

G2:UbelievesS_m believes $U \Leftrightarrow^{SK} S_m$

G3: S_m believes $U \Leftrightarrow^{SK} S_m$.

G4:$S_m$believes$U$believes $U \Leftrightarrow^{SK} S_m$.

G5:$S_m$believes$S_m \Leftrightarrow^{SK} TA$

G6:$S_m$believes TA believes $S_m \Leftrightarrow^{SK} TA$

G7: TA believes$S_m \Leftrightarrow^{SK} TA$

G8: TA believesS_m believes $S_m \Leftrightarrow^{SK} TA$

In the second step, the proposed protocols are idealized and written in the form of language of formal logic.

M1 (Message 1): $U \rightarrow S_m$: pid_u, TS_u, A_u, $E_u : \langle A \rangle_{(D_u)}$, $F_u : \langle N_u \rangle_{(D_u)}$, $Z_u : \langle SID_m \rangle_{\left(D_u\left(N_u\right)\right)}$.

M2 (Message 2):$S_m \rightarrow$TA: M1, $psid_m$, $J_u : \langle N_m \rangle_{(B_{sm})}$.

M3(Message3): TA$\rightarrow S_m$: $P_{TA} : \langle N_m \oplus N_{TA} \rangle_{\left(h\left(N_u\left(D_u\right)\right)\right)}$, $V_{TA} : \langle N_u \oplus N_{TA} \rangle_{(SK_{TA})}$.

M4(Message4):$S_m \rightarrow U$: $P_{TA} : \langle N_m \oplus N_{TA} \rangle_{\left(h\left(N_u\left(D_u\right)\right)\right)}$

In the third step, we identify the assumptions which show the initial state of the proposed protocol.

A1 (First assumption): $U \mid \equiv \#(N_u)[U$ believes fresh$(N_u)]$.

A2: $S_m \mid \equiv \#(N_u)$, A3: TA $\mid \equiv \#(N_u)$, A4: $S_m \mid \equiv \#(N_m)$,

A5: $U| \equiv \#(N_m)$, A6: TA $|\equiv \#(N_m)$.

A7: TA$|\equiv \#(N_{TA})$, A8: $U| \equiv \#(N_m \oplus N_{TA})$, A9: $U |\equiv \#(N_u \oplus N_{TA})$.

A10:$U|\equiv U \Leftrightarrow^{D_u} S_m$, A11:$S_m|\equiv(U \Leftrightarrow^{SK} S_m)$,

A12:$S_m|\equiv(S_m \Leftrightarrow^{B_{sm}} TA)$, A13: TA$|\equiv(U \Leftrightarrow^{SK} S_m)$. Many other assumptions can be added in the initial assumption, such as $S_m|\equiv U$ controls (N_u), TA $|\equiv S_m$ controlsN_m, but they are already evident, so we do not expand that.

In the fourth step (proof), the idealized forms (M1, M2… Mn) of the proposed protocol are analyzed. The basis of the analysis is based on the assumptions and BAN logic rules such as message meaning (MM), Freshness- Conjunction (FC), Belief (BL), Nonce -Verification (NV), Jurisdiction (JR), Session keys (SK). Based on the analysis in the fourth step, we reach the goals.

Using Message 1: $U \rightarrow S_m$: pid_u, TS_u, A_u, $E_u : \langle A \rangle_{(D_u)}$, $F_u : \langle N_u \rangle_{(D_u)}$, $Z_u : \langle SID_m \rangle_{\left(D_u\left(N_u\right)\right)}$ and seeing rule we get.

S1: $S_m \triangleleft pid_u$, TS_u, A_u, $E_u : \langle A \rangle_{(D_u)}$, $F_u : \langle N_u \rangle_{(D_u)}$, $Z_u : \langle SID_m \rangle_{\left(D_u\left(N_u\right)\right)}$.

Applying MM rule with S1, A11 results in S2: $S_m \mid \equiv U \mid \equiv N_u$.

Applying FC and NV rules A2, S2, results in S3: $S_m \mid \equiv U \mid \equiv N_u$ (hereN_u is a required parameter of the proposed protocol's session key.)

Applying JR rule with A14, S3 results in S4: $S_m \mid \equiv (N_u)$.

Applying SK rule with A2, S3 results in S5: $S_m|\equiv U \Leftrightarrow^{SK} S_m$**(G3).**

Applying the NV rule with A2, S3 results in S6: $S_m|\equiv U|\equiv U \Leftrightarrow^{SK} S_m$**(G4).**

Applying seeing rule on M2: $S_m \rightarrow$TA: M1, $psid_m$, $J_u : \langle N_m \rangle_{(B_{sm})}$and results S7: $TA \triangleleft M1$, $psid_m$, $J_u : \langle N_m \rangle_{(B_{sm})}$.

Applying MM rule with A13, S7, and results in S8: $TA \mid \equiv S_m \mid \equiv N_m$.

Applying A6, S7, MM, and NV rule results S9: $TA \mid \equiv S_m \mid \equiv N_m$. ($N_m$is a required parameter of the proposed protocol's session key.)).

Applying JR rule with A15, S9 results in S10: $TA \mid \equiv N_m$.

Applying SK rule with A6, S10 results in S11: $TA|\equiv S_m \Leftrightarrow^{SK} TA$**(G7).**

Applying NV rule with A6, S11 results in S12: $TA|\equiv S_m|\equiv S_m \Leftrightarrow^{SK} TA$**(G8).**

Applying seeing rule with M3 results S13: $S_m \triangleleft P_{TA}$, V_{TA}. Results S14: $S_m \mid \equiv TA \mid \equiv (N_u \oplus N_{TA})$.

Applying FC and NV rule with A12, S14 results in S15: $S_m \mid \equiv TA \mid \equiv (N_u \oplus N_{TA})$.

Applying SK rule with A9, S15 results in S16: $S_m|\equiv S_m \Leftrightarrow^{SK} TA$**(G5).**

Applying NV rule with A9, S16 results in S17: $S_m|\equiv TA|\equiv S_m \Leftrightarrow^{SK} TA$**(G6).**

Applying seeing rule with M4: $S_m \rightarrow U$: $P_{TA} : \langle N_m \oplus N_{TA} \rangle_{\left(h\left(N_u\left(D_u\right)\right)\right)}$ results in S18: $U \triangleleft P_{TA}$.

Applying MM rule with S18, A8 results in S19: $U \mid \equiv S_m \mid \equiv (N_m \oplus N_{TA})$.

Applying FC and NV rule with S19, A10 results in S20: $U \mid \equiv S_m \mid \equiv (N_m \oplus N_{TA})$.

Applying SK rule with A10, S20 results in S21: $U|\equiv U \Leftrightarrow^{SK} S_m$ **(G1).**

Applying NV rule with A8, S21 results in S22: $U|\equiv S_m|\equiv U \Leftrightarrow^{SK} S_m$**(G2).**

Thus, the above steps are used in BAN logic to verify the authentication protocols.

5.2 AVISPA: A simulator for authentication protocols

It is vital to have tools that support the rigorous analysis of security protocols to speed up the next generation of security protocols and improve their security. For that, it detects weaknesses and establishes their correctness. In 2005, Armando et al. [6] came with a push-button tool for the

Table 2 Security features comparison

Protocols	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
Xu et al. [51]	√	×	×	×	×	×	×	√	×	×	×	×	×	√	×
Wu et al. [52]	√	×	√	×	×	√	√	√	×	×	×	√	√	√	√
Chang et al. [11]	√	×	×	×	×	×	√	√	√	×	×	×	×	√	√
Amin et al. [4]	×	×	×	×	×	×	×	√	×	×	√	×	√	√	√
Sun et al. [42]	√	×	×	×	×	×	√	√	√	×	×	×	×	√	×
Challa et al. [10]	√	×	×	×	√	√	√	√	√	√	×	√	√	√	√
Proposed protocol	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√

S1: User anonymity and un-traceability, S2: OGA, S3: PFS, S4: SA, S5: KSTIA, S6: SSCA, S7: PIA, S8: U-IA, S9: C-IA, S10: TA-IA, S11: Pre-verification in smart card, S12: K-KA, S13: Formal verification using AVISPA, S14: Mutual Authentication, S15: RA, ×: Insecure against a particular attack or does not support specific features, √: Secure against a specific attack or defend the specific features.

automated validation of Internet security-sensitive protocols and applications named AVISPA. The tool is used by a protocol designer who describes a security problem in the High-Level Protocol Specification Language HLPSL [20]. Moreover, detailed studies of AVISPA and its execution are available in [5, 6, 13]. In addition to this, some essential tricks to write the protocols in HLPSL language are reported in our work, followed by screenshots of our protocol result.

- First and foremost, the Role of each participant is written, which contains the Role name, declaration of local as well as constant variables, and transition.
- When the Roles of participants are defined, Roles are combined in a session.
- Define the Environment in which protocol is analyzed that contains prior knowledge about the intruder, the scenario to be executed, and the session instances to be run in parallel.
- Finally, we declare security properties of protocol to be executed.

We have passed our protocol with all four utilities provided by AVISPA tools that are following:

1. **On-the-fly Model-Checker (OFMC**): Performs protocol falsification and bounded verification.
2. **Constraint-Logic-based Attack Searcher (CL-AtSe):** It can identify type flaws and manage message concatenation associativity.
3. **SAT-based Model-Checker (SATMC**): constructs a propositional formula encoding a bounded unrolling of the IF's specified transition relation, the initial state, and the set of conditions denoting a breach of the security properties.
4. **TA4SP (Tree Automata based on Automatic Approximations for the Analysis of Security Protocols**): Regular tree languages and rewriting are used to approximate attacker knowledge. Our protocol passed all four tests.

5.3 Informal analysis of attacks

In these subsections that follow, we do a preliminary study to show the secrecy of the proposed protocol against a wide range of threats.

5.3.1 User anonymity and un-traceability

The identity of user must not be revealed during the exchange of keys among the user, cloud server, and trusted authority. User, cloud server, and trusted authority are generated for each fresh session nonce, respectively. At the same time, different timestamps (T_{TA},T_{su}andT_{sm}) validate the session. User identity is concatenated with Q_u and hashed with PID_u: ($PID_u = h(U_i \| Q_u)$). It is passed to the trusted authority. Hence, it is obvious that extracting information about the user's identity (U_i) is impossible by a third party. So, the proposed protocol preserves user anonymity and the un-traceability of the user's identity.

5.3.2 Offline guessing attacks (OGA)

The adversary can know various information saved in smart cards using power analysis or differential power analysis attacks [32, 54]. In the proposed protocol, the smart card holds $\{C_u,E_u,DP_u,\}$. Concatenation and XOR operation on user identity and password preceded by hashing results above parameters. One-way collision resistance property of hash function and uniqueness of ECC makes too hard to get knowledge about user identity and password. Wu.et. al [51] and Tsu et al. [52] are also using almost the same parameters as Amin's protocol [4]. Possibilities of the same attacks are there in these two protocols. In the proposed protocol, parameters such as user id (Ui), password, and server identity are concatenated with the private key generated on ECC. Then it is hashed with one-way hash function functions. So these consecutive steps eliminate the chance of power analysis attacks.

5.3.3 Perfect forward secrecy (PFS)

Some keys and parameters are readily available to adversaries. PFS ensures that session keys among users, cloud servers, and trusted authorities are unknown to adversaries at any cost. Here in the proposed protocol, assume an adversary wants to compute $SK_u = SK_m = h(N_m \oplus N_{TA} \oplus N_u)$.These three parameters $N_m N_{TA}$ and N_u are not available to the adversary because these parameters are always used in hashed conditions and concatenated with elliptic curve points. It is very tough to get a polynomial-time solution to this problem. Hence, the proposed protocol can provide PFS.

5.3.4 Sybil attack (SA)

In a Sybil attack, multiple accounts/nodes are created by an adversary to take over the network. In the proposed protocol, TA's elliptic curve Q_{TA} is involved in the registration and first phase of authentication. Thus ECDLP property of ECC eliminates the chance of private computing key from known parameters. So the uniqueness of private keys disaffects the harm of multiple accounts.

5.3.5 Known session-specific temporary information (KSTIA)

In KSTIA attack, session ephemeral secrets are exposed to the adversary, based on this an adversary would be able to get the session keys. In the proposed protocol, almost all information is hashed and concatenated with the private key. It is very tough to extract information from temporary information. Some researcher explains this attack with the name of Ephemeral Secret Leakage (ESL) Attack.

5.3.6 Stolen smart card attack (SSCA)

At the worst, there may be the situation that information $\{C_u,E_u,DP_{u,}\}$ are known to intruder A due to the loss or steal of a smart card. But due to h (·) protection secret entities of U, viz.,C_u,E_u and $DP_{u,}$, are not known to A. Therefore, our protocol is secure against this attack.

5.3.7 Privileged insider attack (PIA)

There is always the possibility of two types of adversaries in a system, outside adversaries and inside adversaries. An inside adversary may have privileged access to TA. In this case an adversary can get information such as A_u and PID_uwhile user is registering to TA, but $A_u = h(P_u\|d_u)$, $PID_u = h(U_i\|Q_u)$ and $\{d_u,Q_u\}$ are the points on elliptic curve, so to reveal P_u and U_i from A_u and PID_u is almost impossible. Furthermore, if the intruder fetches the information $\{C_u, E_u$ and $DP_{u,}\}$ from the smart card directly by smart card attack [43, 50], in this case also getting P_u and U_i is very difficult as discussed in stolen smart card attack.

5.3.8 User impersonation attack (U-IA)

An adversary may put P_u and U_i in authentication phase, to impersonate a user U, but to complete authentication by TA, there is need of SID_m and N_u. Getting SID_m and N_u is not possible because $N_u = F_u \oplus D_u$ and $D_u = E_u \oplus A_u$, forA_u, P_u is concatenated with d_u (point of elliptic curve) and transferred to TA in hashed form. Therefore getting N_u is too tough, similar is the condition with SID_m.

5.3.9 Cloud server impersonation attack (C-IA)

An adversary may get the identity of a cloud server (sid_m) to impersonate the cloud server, but to be authenticated from trusted authority (TA), $B_{sm} = h(psid_m\| Q_{TA})$ should be known where $psid_m$= h ($sid_m\|d_{cs}$). Here, in this case Q_{TA} and d_{cs} are the points of two different elliptic curves. Hence, by using the ECDLP property of ECC, we conclude that our protocol is safe from Cloud server Impersonation attacks.

5.3.10 Trusted authority (TA) impersonation attack (TA-IA)

To impersonate TA, an adversary may attempt to get the key $\{P_{TA}, R_{TA}, Z_{TA}, V_{TA}\}$ but at cloud server end $V_{TA} = h((N_u\| N_{TA})\| SK_m)$and at user end $Z_{TA} = h((N_m \oplus N_{TA})\| SK_u)$ are used for authentication, getting nonces N_u, N_{TA} and N_m in correct form is tough, also collision resistance property of hash function makes almost impossible to get correct keys. Thus, this protocol is safe from TA impersonation attack.

5.3.11 Pre-verification in the smart card

In the login phase several protocols such as [12, 28, 47], does not support smart card to verify the identity and password of user. It puts an extra burden on the server. While in the proposed protocol the smart card checks $C_u* = C_u, U_i * = U_{i,} and\ P_u * = P_u$ in the login phase. If it is found valid, the proposed protocol proceeds for the authentication phase. Otherwise, the session will be deferred until the right password and identity have been submitted. This means

Table 3 Comparison of the computational overhead of our protocols with related protocols

Protocols	User (U)	Trusted authority (TA)	Cloud server	Total overhead (sec)
Proposed protocol	$T_{ecm}+10Th$	$T_{ecm}+9Th$	$T_{ecm}+3Th$	$3T_{ecm}+22Th\approx0.2002$
Wu et al. [52]	13Th	19Th	8Th	$40Th\approx0.02$
Amin et al. [4]	9Th	10Th	19Th	$23Th\approx0.0115$
Sun et al. [42]	$4T_{ecm}+3Th$	–	$6T_{ecm}+4T_{eca}+4Th$	$10T_{ecm}+7Th+4T_{eca}\approx0.67775$
Li et al. [30]	$2T_{ecm}+6Th$	–	$6T_{ecm}+4Th+4T_{eca}$	$8T_{ecm}+10Th+4T_{eca}\approx0.35667$
Chang-Le [11]	$4T_{ecm}+12Th$	–	9Th	$4T_{ecm}+21Th\approx0.2628$
Challa et al. [10]	$2T_{ecm}+17Th+T_{fe}$	5Th	$T_{ecm}+5Th$	$3T_{ecm}+27Th+T_{fe}\approx0.333$

that the proposed protocol saves computational and communication expenses whether there is inaccurate input or an unlawful user. As a result, the proposed protocol successfully provides pre-verification..

5.3.12 Known-key attack (K-KA)

If one session key is compromised, it does not necessarily mean that all other session keys would be as well. The accepted session key in the proposed protocol is based on three random ephemeral secrets N_u, N_{TA} and N_m, which are different for each session. Due to the difficulty of the ECDLP problem, the adversary may not be able to derive all of these at the same time. As a result, revealing one session key does not allow the adversary to learn about additional session keys.

Sometimes researchers call it No key control property.

5.3.13 Replay attack (RA)

In the authentication phase, we have used timestamps $\{TS_u, TS_m, TS_{TA}\}$, and accepted delay (ΔT) is sufficiently tiny, which makes replaying of old messages useless. Hence, our protocol is secure against replay attacks. In Table 2, the summary of the security features of recently published protocols and proposed protocols is given.

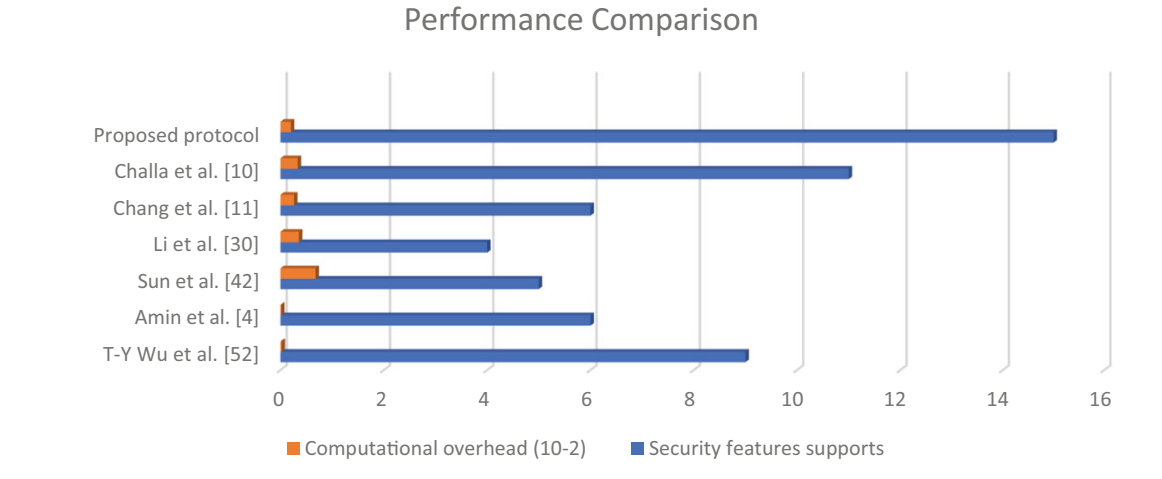


Fig. 7 Performance Comparison

5.4 Performance comparison

The computational cost of the proposed protocol is compared to that of other protocols in Table III. We have taken experimented values as reported in [1, 3, 10, 49] are following:

Time to compute hash function (T_h): 0.0005 s, time to compute ECC multiplication (T_{ecm}):0.06307 s, time to compute fuzzy extractor (T_{fe}):0.063075. Time to compute ECC point addition (T_{eca}): 0.010875 s. we have ignored time consumed in XOR and concatenation operations. At last, we have plotted a graph for better judgment between security features supported and time consumed by the proposed protocol and other protocols. The graph of summary of security analysis is presented for the ease of readers.

5.5 The efficiency of the proposed protocol

Table 2 compares the proposed protocol to other relevant existing protocols in terms of security and functional elements that are desired or required. Our protocol includes all essential security features. Existing protocols are missing several key features (as discussed in section 5.3, informal analysis of proposed protocols), such as protection against offline password guessing attacks, known key attacks, stolen card attacks, and impersonation attacks. In addition, the proposed protocol provides rigorous security analysis and formal security verification using the widely-accepted AVISPA tool. Table 3 shows that the total overhead of the proposed protocol is 0.2 s, which is much less than most of the existing protocols. Results from Tables 2 and 3 proves the efficiency of the proposed protocol. Furthermore, Fig. 7 makes our claim more claim in terms of performance comparison.

6 Conclusion

This paper first discusses crucial role of IoT devices in our daily lives, the advantages of using cloud computing in IoT devices, and authentication requirements in such an environment. Efficient authentication is the need for the hours for IoT devices. The proposed protocol is computationally very light and successfully resists attacks that are not covered by the currently existing protocol. Elliptic curve discrete logarithm problem (ECDLP) is used with one way hash function and XOR operator. ECDLP property makes the proposed protocols hard to break. The use of one way hash function and XOR operator maintains the efficiency of authentication along with secrecy. The proposed protocol is verified using BAN logic and simulated using HLPSL language for the AVISPA tool. We have added essential tricks to play with BAN logic and the AVISPA tool for any protocols related to the secure use of IoT devices. The generalized approach of the system model makes the proposed protocol implementable for the different scenarios of IoT devices such as Medical IoT, Industrial IoT, and cyber-physical systems. The proposed protocol can be extended for healthcare IoT devices and Industrial IoT by incorporating novel biometric, homomorphic encryption, and iterative learning techniques.

Acknowledgments We would like to thank the anonymous reviewers for their valuable comments, which helped us to improve the organization, content, and quality of this Manuscript.

Declarations

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

References

1. Almuhaideb AM (2021) Re-AuTh: lightweight re-authentication with practical key Management for Wireless Body Area Networks. Arab J Sci Eng 46:8189–8202. https://doi.org/10.1007/s13369-021-05442-9
2. Alzahrani BA (2021) Secure and efficient cloud-based IoT authenticated key agreement scheme for e-health wireless sensor networks. Arab J Sci Eng 46:3017–3032. https://doi.org/10.1007/s13369-020-04905-9
3. Amin R, Biswas GP (2016) A secure light weight scheme for user authentication and key agreement in multi-gateway based wireless sensor networks. Ad Hoc Netw 36:58–80. https://doi.org/10.1016/j.adhoc.2015.05.020
4. Amin R, Kumar N, Biswas GP, Iqbal R, Chang V (2018) A light weight authentication protocol for IoT-enabled devices in distributed cloud computing environment. Futur Gener Comput Syst 78:1005–1019. https://doi.org/10.1016/j.future.2016.12.028
5. Alessandro Armando, David Basin, Jorge Cuellar MR and LV (2001) The high level protocol specification language deliverable details. Avispa
6. Armando A, Basin D, Boichut Y, et al (2005) The AVISPA tool for the automated validation of Internet security Protocols and applications. Pp 281–285
7. Bae W, Kwak J (2020) Smart card-based secure authentication protocol in multi-server IoT environment. Multimed Tools Appl 79:15793–15811. https://doi.org/10.1007/s11042-017-5548-2
8. Banerjee S, Odelu V, Das AK, Srinivas J, Kumar N, Chattopadhyay S, Choo KKR (2019) A provably secure and lightweight anonymous user authenticated session key exchange scheme for Internet of things deployment. IEEE Internet Things J 6:8739–8752. https://doi.org/10.1109/JIOT.2019.2923373
9. Burrows M, Abadi M, Needham R (1990) A logic of authentication. ACM Trans Comput Syst 8:18–36. https://doi.org/10.1145/77648.77649
10. Challa S, Das AK, Gope P, Kumar N, Wu F, Vasilakos AV (2020) Design and analysis of authenticated key agreement scheme in cloud-assisted cyber–physical systems. Futur Gener Comput Syst 108:1267–1286. https://doi.org/10.1016/j.future.2018.04.019
11. Chang CC, Le HD (2016) A provably secure, efficient, and flexible authentication scheme for ad hoc wireless sensor networks. IEEE Trans Wirel Commun 15:357–366. https://doi.org/10.1109/TWC.2015.2473165
12. Chen F, Tang Y, Cheng X, Xie D, Wang T, Zhao C (2021) Blockchain-based efficient device authentication protocol for medical cyber-physical systems. Secur Commun Networks 2021:1–13. https://doi.org/10.1155/2021/5580939
13. Chevalier Y, Compagna L, Cuellar J, et al (2006) A High Level Protocol Specification Language for Industrial Security-Sensitive Protocols To cite this version: HAL Id: inria-00099882 A High-Level Protocol Specification Language for Industrial Security-Sensitive Protocols ∗. https://hal.inria.fr/inria-00100219
14. Chintala RR, Kallepalli H, Kotapati J, et al (2021) Implementing security framework for cloud based IOT network implementing security framework for cloud based IOT network
15. Do Xuan C, Dao MH, Nguyen HD (2020) APT attack detection based on flow network analysis techniques using deep learning. J Intell Fuzzy Syst 39:4785–4801. https://doi.org/10.3233/JIFS-200694
16. Dolev D, Yao AC (1983) On the security of public key Protocols. IEEE Trans Inf Theory 29:198–208. https://doi.org/10.1109/TIT.1983.1056650
17. Gadicha AB, Gupta VBB, Gadicha VB, et al (2021) Multimode approach of data encryption in images through quantum steganography. Pp 99–124
18. Haack C (2008) What is BAN logic ? Verification of security Protocols what are questions that BAN logic aims to answer ? What are limitations of BAN logic ? BAN logic ' s model of time formulas: syntax domains formulas: basic formulas formulas: idealized messages
19. Hankerson D, Menezes A (2011) Elliptic curve cryptography. In: Encyclopedia of cryptography and security. Springer US, Boston, MA, pp. 397–397
20. Internet A, Protocols S (2006) HLPSL Tutorial. In: Society

21. Iqbal W, Abbas H, Daneshmand M, Rauf B, Bangash YA (2020) An in-depth analysis of IoT security requirements, challenges, and their countermeasures via software-defined security. IEEE Internet Things J 7:10250–10276. https://doi.org/10.1109/JIOT.2020.2997651
22. Irshad A, Ahmad HF, Ramzan MS (2016) An efficient and anonymous Chaotic Map based authenticated key agreement for multi-server architecture KSII Trans Internet Inf Syst 10. https://doi.org/10.3837/tiis.2016.12.023
23. Islam SKH, Biswas GP (2012) An improved pairing-free identity-based authenticated key agreement protocol based on ECC. Procedia Eng 30:499–507. https://doi.org/10.1016/j.proeng.2012.01.890
24. Kim H, Kim D, Yi O, Kim J (2019) Cryptanalysis of hash functions based on blockciphers suitable for IoT service platform security. Multimed Tools Appl 78:3107–3130. https://doi.org/10.1007/s11042-018-5630-4
25. Kocher P, Jaffe J, Jun B (1999) Differential power. Analysis pp:388–397
26. Koppanati RK, Kumar K (2021) P-MEC: polynomial congruence-based multimedia encryption technique over cloud. IEEE Consum Electron Mag 10:41–46. https://doi.org/10.1109/MCE.2020.3003127
27. Kumar K, Kurhekar M (2016) Economically efficient virtualization over cloud using Docker containers. In: 2016 IEEE international conference on cloud computing in emerging markets (CCEM). IEEE, pp 95–100
28. Kumari S, Karuppiah M, Das AK, Li X, Wu F, Gupta V (2018) Design of a secure anonymity-preserving authentication scheme for session initiation protocol using elliptic curve cryptography. J Ambient Intell Humaniz Comput 9:643–653. https://doi.org/10.1007/s12652-017-0460-1
29. Lamport L (1981) Password authentication with insecure communication. Commun ACM 24:770–772. https://doi.org/10.1145/358790.358797
30. Li H, Li F, Song C, Yan Y (2015) Towards smart card based mutual authentication schemes in cloud computing. KSII Trans Internet Inf Syst 9:2719–2735. https://doi.org/10.3837/tiis.2015.07.022
31. Manupriya P, Sinha S, Kumar K (2017) V⊕SEE: Video secret sharing encryption technique. In: 2017 Conference on information and communication technology (CICT). IEEE, pp 1–6
32. Messerges TS, Dabbish EA, Sloan RH (2002) Examining smart-card security under the threat of power analysis attacks. IEEE Trans Comput 51:541–552. https://doi.org/10.1109/TC.2002.1004593
33. Miller VS (1986) Use of elliptic curves in cryptography. In: Advances in cryptology — CRYPTO '85 proceedings. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 417–426
34. Muhammad G, Alhussein M (2021) Security, trust, and privacy for the Internet of vehicles: a deep learning approach. IEEE Consum Electron Mag 2248:1–1. https://doi.org/10.1109/MCE.2021.3089880
35. Nandy T, Yamani M, Bin I et al (2019) Review on security of Internet of things authentication mechanism. IEEE Access 7:151054–151089. https://doi.org/10.1109/ACCESS.2019.2947723
36. Naseer O, Ullah S, Anjum L (2021) Blockchain-based decentralized lightweight control access scheme for smart grids. Arab J Sci Eng 46:8233–8243. https://doi.org/10.1007/s13369-021-05446-5
37. Pete P, Patange K, Wankhade M et al (2018) 3E-VMC: an experimental energy efficient model for VMs scheduling over cloud. In: 2018 first international conference on secure cyber computing and communication (ICSCCC). IEEE:322–327
38. Rangwani D, Om H (2021) A secure user authentication protocol based on ECC for cloud computing environment. Arab J Sci Eng 46:3865–3888. https://doi.org/10.1007/s13369-020-05276-x
39. Ray PP (2017) A survey of IoT cloud platforms. Futur Comput Informatics J 1:35–46. https://doi.org/10.1016/j.fcij.2017.02.001
40. Sharma S, Kumar K (2018) GUESS: genetic uses in video encryption with secret sharing. In: Advances in Intelligent Systems and Computing. Springer Verlag, pp. 51–62
41. Sierra JM, Hernández JC, Alcaide A, Torres J (2004) Validating the use of BAN LOGIC. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 3043:851–858. https://doi.org/10.1007/978-3-540-24707-4_98
42. Sun H, Wen Q, Zhang H, Jin Z (2013) A novel remote user authentication and key agreement scheme for mobile client-server environment. Appl math. Inf Sci 7:1365–1374. https://doi.org/10.12785/amis/070414
43. Syverson P, Cervesato I (2001) The logic of authentication Protocols. Pp 63–137
44. Team A (2006) AVISPA v1. 1 user manual. Inf Soc Technol 1:1–88
45. Team A (2006) HLPSL Tutorial, A Beginner's Guide to Modelling and Analysing Internet Security Protocols Inf Technol Solut 1–52
46. Tsai JL, Lo NW (2015) A privacy-aware authentication scheme for distributed Mobile cloud computing services. IEEE Syst J 9:805–815. https://doi.org/10.1109/JSYST.2014.2322973
47. Wazid M, Das AK, Kumari S, Li X, Wu F (2016) Provably secure biometric-based user authentication and key agreement scheme in cloud computing. Secur Commun Networks 5:422–437. https://doi.org/10.1002/sec
48. Wazid M, Das AK, Hussain R, Succi G, Rodrigues JJPC (2019) Authentication in cloud-driven IoT-based big data environment: survey and outlook. J Syst Archit 97:185–196. https://doi.org/10.1016/j.sysarc.2018.12.005

49. Wazid M, Das AK, Bhat KV, Vasilakos AV (2020) LAM-CIoT: lightweight authentication mechanism in cloud-based IoT environment. J Netw Comput Appl 150:102496. https://doi.org/10.1016/j.jnca.2019.102496
50. Wessels J (2001) Applications of Ban-Logic
51. Wu F, Li X, Xu L, Sangaiah AK, Rodrigues JJPC (2018) Authentication protocol for distributed cloud computing: an explanation of the security situations for Internet-of-things-enabled devices. IEEE Consum Electron Mag 7:38–44. https://doi.org/10.1109/MCE.2018.2851744
52. Wu T-Y, Lee Z, Obaidat MS, Kumari S, Kumar S, Chen CM (2020) An authenticated key exchange protocol for multi-server architecture in 5G networks. IEEE Access 8:28096–28108. https://doi.org/10.1109/ACCESS.2020.2969986
53. Xue K, Hong P, Ma C (2014) Journal of computer and system sciences a lightweight dynamic pseudonym identity based authentication and key agreement protocol without verification tables for multi-server architecture. J Comput Syst Sci 80:195–206. https://doi.org/10.1016/j.jcss.2013.07.004
54. Yang S, Greenberg A, Endsley M (2011) Social computing, Behavioral-Cultural Modeling and Prediction. Springer Berlin Heidelberg, Berlin, Heidelberg

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

iJRASET
International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** II **Month of publication:** February 2022

DOI: https://doi.org/10.22214/ijraset.2022.38581

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue II Feb 2022- Available at www.ijraset.com

A Survey and Feasibility Research Study on Computer Aided Process Planning (CAPP)

Ashok Kumar Madan[1], Kunal Krishna[2], Bipul Kumar[3]
[1]*Professor, Department of Mechanical Engineering, Delhi Technological University, New Delhi-110042, India*
[2, 3]*PG Student, Production Engineering, Delhi Technological University, New Delhi-110042, India*

Abstract: ***Computer Aided Process Planning (CAPP) is the use of computer system to help and guide process planners in planning and executing different functions. It has emerged as one of the crucial links between integration of Computer Aided Design (CAD) and Computer Aided Manufacturing (CAM). In this paper, an attempt is made to provide a complete survey of CAPP based on different methodology namely, features and solid model, neural networks, optimization, fuzzy logic, knowledge based, web based, hybrid manufacturing, artificial intelligence, virtual, (STEP)-compliant method. The aim of this paper is to provide an update, in and after year 2000, in field of CAPP systems using survey and representing graphically for easy understanding***

Keywords: CAPP, CAD, CAM, Methodology, Variant, Generative.

I. MOTIVATION

The significance of CAPP in a manufacturing facility cannot be undervalued. One of the logics for this support is that it provides a critical link between design and manufacturing and at the same time it reduces the time and cost and improves the quality as well. So, a comprehensive survey is done to present its role.

II. INTRODUCTION

In the past two decades, more than hundred different CAPP systems with variant, generative and hybrid approaches are developed by human beings to assist them in planning. It has emerged as one of the important links between integration of CAD and CAM. It is one of the principal components in computer integrated manufacturing. A lot of research has been carried out in this area, one of the main reasons being reduction in throughout time and improvement in quality. CAPP is the use of computer to help process planner in planning functions. There are broadly two approaches in CAPP: Variant and Generative.

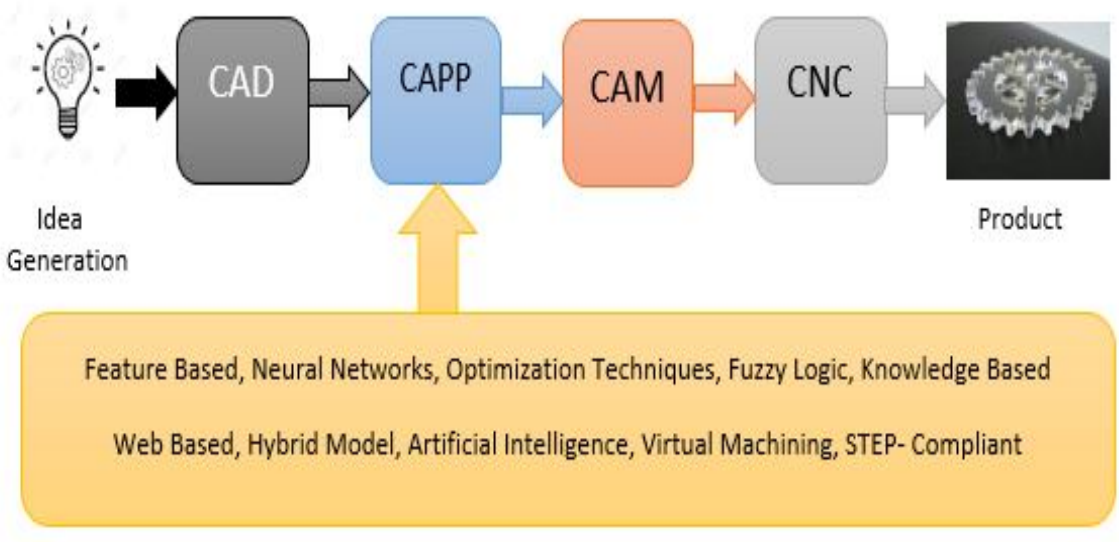


Fig.1 CAPP in manufacturing cycle.

© IJRASET: All Rights are Reserved |

A. Variant

Also called as retrieval method. In this approach, process plan for a brand-new part is created by recollecting, recognizing and retrieving an existing plan for an analogous part and making necessary modifications in the existing part. It basically selects generic process plan from established master process plan, developed for each part family and then modify it to suit as per the requirement of new part using a group technology code (GT codes). The benefit of this method includes the simplicity of maintenance, but the lack of an on-time calculation of the manufacturing process and the quality of the process plans still rely on the knowledge of a process planner, and it still requires human inputs for the formation of mass data into manufacturing processes.



Fig.2 Variant approach used in CAPP

B. Generative

This involves creating a process plan from scratch for each component without the need of humans. These systems are built to automatically synthesise process data in order to generate a part's process plan. These systems have the logic to build a process plan for a specific component by using a manufacturing database and relevant part description schemes. Generative approach obliterates the drawback of the variant approach in CAPP and bridges the gap between the computer-aided design (CAD) and computer-aided manufacturing (CAM).

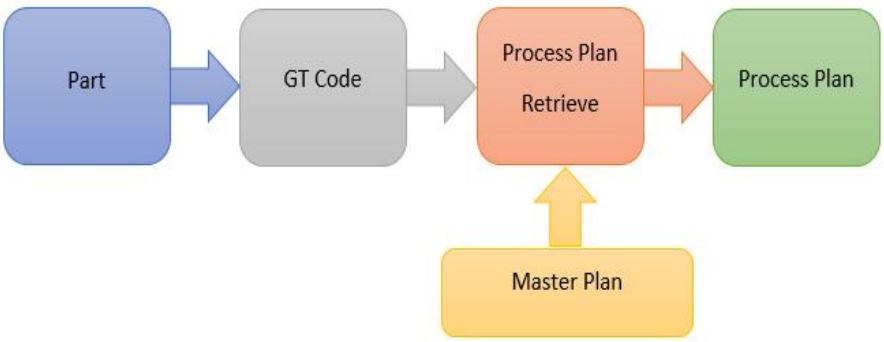


Fig.3 Generative approach used in CAPP

1) Advantages of using CAPP

CAPP has numerous advantages, some of them are listed here

a) Reduction in process planning and production lead time.
b) CAPP ensures process plan consistency.
c) It can give more complete and detailed process plans.
d) It can give faster response to rapid changes in engineering.

e) It caters an improved ability to introduce latest manufacturing technology and rapidly update present process plans so as to utilize the improved technology.
f) Accessing up-to-date information in a central database system can be made available.
g) It results in improved cost-estimation procedures and eliminates calculation errors.
h) It facilitates with production scheduling and capacity utilization of machinery.
i) Makes efficient use of resources leading to reduction of cost.
j) Maximisation of JIT performance.

When compared with manual process planning the CAPP systems have fewer disadvantages:

- When compared with manual process planning initial cost of establishing a CAPP system are high.
- Also, CAPP systems have inability to show special manufacturing techniques.

III. OBJECTIVE

To present a Comprehensive Survey for easy understanding of the past, present and future of CAPP with brief introduction of CAPP approaches, methods, conclusion.

IV. CAPP SURVEY

A. *Brief review of past survey*

1) *1980's: I*n the year 1984, Harold presented the first review on CAPP mainly discussing about developing a generative type automated planning system.[1] In 1988, Ham and Lu presented an assessment of CAPP stating that future research lies on integration of design, manufacturing and artificial intelligence.[2] Next in 1989, 128 systems of CAPP were surveyed by Gouda and Taraman discussing about the four types of CAPP systems namely variant, generative, semi-generative and an expert process planning system.[3]
2) *1990's:* In 1993, Hoda ElMararghy conducted a survey where issues relating to quality and evolving standards were discussed. It also discussed about the major development in CAPP, evolving trends, integration of design, challenges at that time as well as production planning.[4] In 1995, author Kamrani overviewed the techniques used and role of process planning in manufacturing. It also listed the issues and characteristics associated with selection and evaluation of CAPP systems.[5] Same year Leung published a comprehensive review on CAPP where he stated that solid modelling in CAPP system is not sufficient and the future process planning system to be built on AI techniques.[6] A review done in 1996 and presented two years later by Marri and others discussed about the advantages and disadvantages of production planning system along with generative approach for CAPP.[7] Later in 1997, past 8-year survey was done by Cay and Chassapis that provided an overview of manufacturing and feature recognition techniques with CAPP.[8]
3) *2000's:* In 2000, Kulkarni and others presented a review of process planning techniques in layered manufacturing to analyse the situation and suggest future approach of research in this area.[9] In the same year, Tan and Khoshnevis presented a review in integration of process planning and scheduling to discuss the extent of applicability of various approaches and suggest directions for future research.[10] Ahmad and others in 2001 presented current trend in computer aided process planning to classify the presented research work according to their focus.[11] In 2007, Zhang and Xie provided a review on agent technology for collaborative process planning.[12] Professor Hoda ElMaraghy presented evolution and future perspectives of CAPP in 2007-08 to assess the evolution of various CAPP approaches, techniques and their effectiveness in industrial application, future perspective and research. [13]
4) *2010's:* In the year 2010, Li and others presented a review on integrated process planning and scheduling systems which classified the various research work presented and suggested some future trends.[14] In very next year 2011 a critical review of recent developments and future trends in computer aided process planning was presented by Xun Xu and Wang.[15] Three years later in 2014, Yusof and Latif came up with survey on computer aided process planning to classify the focused issues of research work in different areas.[16] Later in same year Isnaini and Shirase presented a review of computer aided process planning systems for machining operation in which discussion of effective parameters relating to development of CAPP systems for process plan for metal removal were highlighted.[17] Mohsen Soori presented application of CAD/CAM system in developing the machining operations using virtual machining in following years 2013, 2014, 2016, 2017 and 2020. Mazin and Atanas in 2018 presented a survey on smart automated computer-aided process planning (ACAPP) techniques which deals with the missing aspects in smart ACAPP generation, the limitations of current systems in recognizing new features and justifying the process selection in CAPP.[18]

B. Present survey

The present survey relating to CAPP methods (past and present) is divided under ten categories to which the related technologies belong and contribution of each method is represented using a figure at last.

1) *Feature Based and Solid model-based Process Planning:* Almost all CAPP systems function on the basis of features and/or require features as an input data. The two approaches to obtaining features are feature recognition and design by features. The feature recognition probes the topology and geometry of the part and determines the existence and definitions of features. While the design by features method involves both adding and subtracting features to develop desired part. In 2017, Vinod V Rampur and other investigated computer aided process planning using STEP neutral file for producing automotive parts from STEP AP-214 of 3D model extracting the basic features of automotive part from STEP AP-214.[19] In the paper by S P Leo Kumar and others in 2014, a new methodology for automatic part feature extraction and CNC code generation for micro features in tool-based micromachining processes was presented.[20] Vamsi Krishna and others in 2011, presented a new idea to generate process plan from feature-based modeling, based on an integrated geometric modeling that proposes both feature-based modeling and its information storage.[21] Chandra R. Devireddy analysed a method of integration of the CAD & CAPP aspects of manufacturing using the idea of Feature-Based Modelling (FBM) and Artificial Neural Networks (ANNs) in year 2010.[22] Patil and Pandey developed an intelligent feature-based process planning system for prismatic parts in 2002.[23] A feature-based inspection process planning system for co-ordinate measuring machine presented by Zhang in 2000.[24] Same year Case and Wan integrated process planning and assembly system to generate new CAD/CAM systems that became capable of assisting in the optimization of product design.[25]

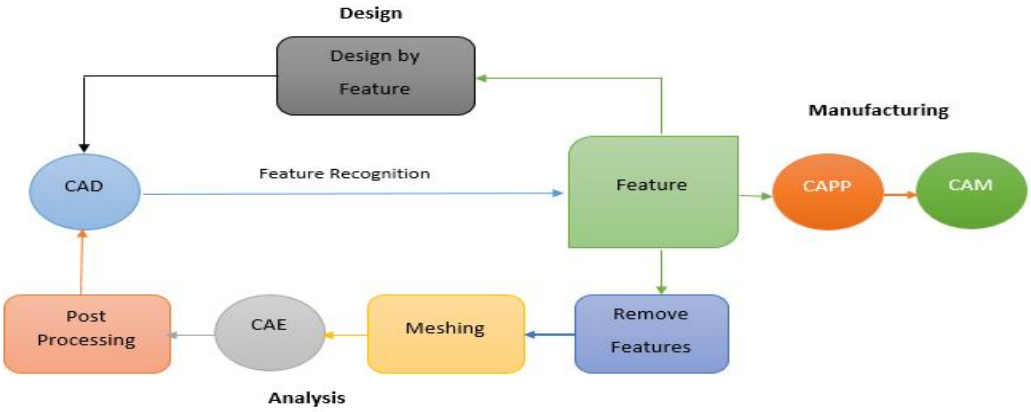


Fig.4 Feature recognition methods in CAPP.

2) *Neural network in CAPP:* It is a network of neuron which are structured in the same way as our brain, it consists of three layers namely an input layer, a hidden layer (which basically of one or more than one layer) and an output layer. The neurons are often called as node. Each node is connected to one another in such a way it resembles a biological neuron and has a corresponding weight and threshold. The chance of obtaining a good result depends upon how dense a network is, i.e., a denser network will provide a high and accurate result. In manufacturing this type of system is used for the selection of machine tools to optimize its performance. The input node in neural network is used to represent its machining part characteristics as well as its operation while on the other hand output node is used to represent specific machine tools which are being used to perform certain operations. Niechen Chena in 2021 came up with research on artificial neural network (ANN) based approach for machining process planning, mostly about the toolpath planning for milling operations.[26] In year 2019, Guang Hu along with others optimized the hardness of SLA printed objects by using the neural network and genetic algorithm.[27] Natarajan and other in 2018 focused on the application and role of artificial neural network techniques in CAPP.[28] In 2017, Izabela Rojek made use of neural network models in CAPP and developed it to aid CAPP for complex real systems.[29] Earlier in 2012, Jun Wang introduced the artificial neural network for manufacturing knowledge modeling and it is represented as neural network weight value matrix, and then forming ANN database to support intelligent CAPP.[30] In 2010, Romdhane and others came up with a model for integration of neural network approach in CAD/CAM for the task of generating a process plan for machining features.[31] While in 2009, Sunil and Pande came up with ANN-based feature-recognition (FR) system, used to recognize complex machining features by analyzing feature topology and geometry.[32]

3) *Optimization techniques using CAPP:* To increase efficiency in process of part production, optimization techniques are generally applied to the CAPP process. Thereby decreasing time and cost of production using the optimized parameters of part manufacturing. In a manufacturing system optimization leads to enhance the efficiency of system by maximizing profit, productivity and quality of product and minimizing the production cost and non-productive time. In 2017 Phung and others presented an effective method of operation sequence optimization in CAPP based on modified clustering algorithm that highlighted the sequence and compared their traveling costs to choose the optimal operation which resulted in the minimum traveling cost in the sequence.[33] Earlier in 2016, Mitin and other investigated on mathematical modelling in the computer aided process planning.[34] An Optimization of operation sequencing in CAPP using simulated annealing technique (SAT) presented by Nallakumarasamy in year 2011 described techniques to obtain optimal sequences in a single run with lesser computational time.[35] In 2009, Shao came up with a new integration model with a modified GA-based approach to facilitate the integration and optimization of various CAPP systems.[36] Same year Salehi and others optimized machine tools, cutting tools and tool access direction by using the genetic algorithm.[37] A year earlier in 2008, Siva Shankar optimized process parameters for rotational components that lead to reduce machining time and cost as compared to handbook readings and traditional practices.[38]

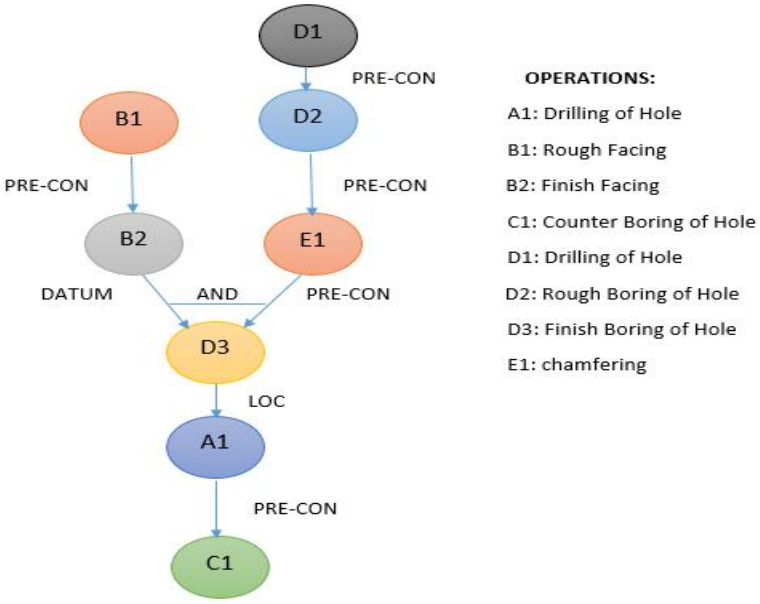


Fig.5 Optimization of operation sequencing in CAPP (Super hybrid genetic algorithms

4) *Fuzzy logic in Process Planning System:* The various decision-making in the real world such as process planning takes place in a domain in which goals and constraints are fuzzy (not known accurately). So, approximation techniques are used to obtain a logical model in a real environment. This theory deals with similar type of problems by transforming human knowledge to mathematical formulae and puts it into the system together with other information like mathematical models and sensory measurements in which goals and constraints can be transformed to model by fuzzy sets. In 2014, Pendokhare developed a fuzzy logic model for drilling operations to select drilling speeds for different types of materials. The fuzzy relationship between various cutting materials, various apertures and feed rates is used to analyse and evaluate various problems between the hardness of a given material and the drilling speed, and these results were given.[39] In 2013, Amaitik developed a logic model with fuzzy characteristics that can help select machining parameters such as cutting speed and feed rate in the automatic CAP system.[40] In 2010, Mula and others studied the effectiveness of fuzzy mathematical programming for any production plan where the demand in the supply chain is essentially fuzzy.[41] Three years ago, in 2007, Leung and others introduced a powerful optimization model for production planning problems in multiple locations in any unsafe environment.[42] In the same year, Aliev and others studied the fuzzy genetic method of aggregate production planning and distribution in supply chain management.[43]

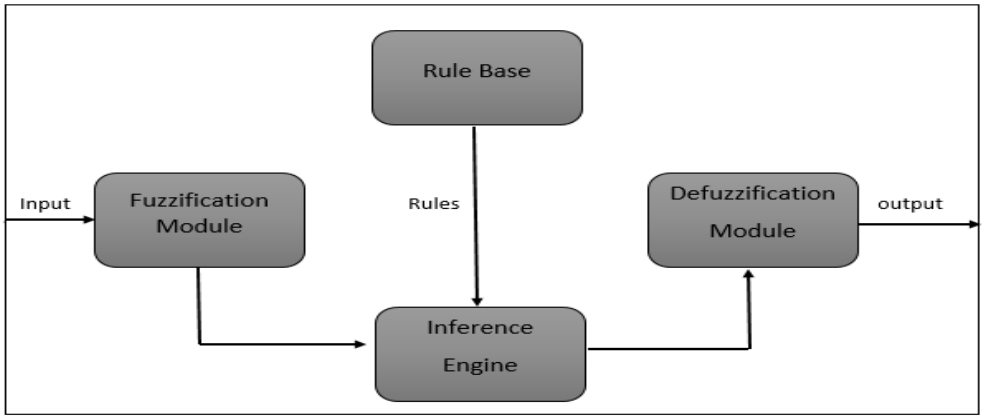


Fig.6 Fuzzy logic architecture

5) *Knowledge based CAPP System:* In this system the process planning to a large extent depends upon the past experience of a process planner or expert, therefore this system is also known as knowledge-based system. In traditional approach the decision of process planning surrounds entirely around a process planner. The planner uses his skills based on past experience to generate processing steps required in manufacturing. There are numerous disadvantages of this system which includes; inconsistency, dependency on skilled experts, time consuming etc. To overcome this bottleneck, it got automated with the help of computers so that better decision-making function can be incorporated in process planning. S P Leo Kumar in 2018, made an effort to review knowledge based expert system applications in manufacturing planning.[44] In 2015, Ravi and others presented knowledge-based computer aided process plan and computer numeric control code generation for the development of a generative process planning methodology used in automated generation of CNC codes for 3 axis milling centres with the CAD geometry as the only primary input.[45] The hybrid method of knowledge representation in a computer aided process planning knowledge-based system brought up by Grabowik and others in 2012.[46] An overview of the CAPP field based on an integrated approach that combined technological and business considerations was studied and presented by Denkena under title knowledge management in process planning in year 2007.[47] Previously in 2001, Law in his presentation investigated object-oriented knowledge-based computer-aided process planning system for circuit boards manufacturing.[48]

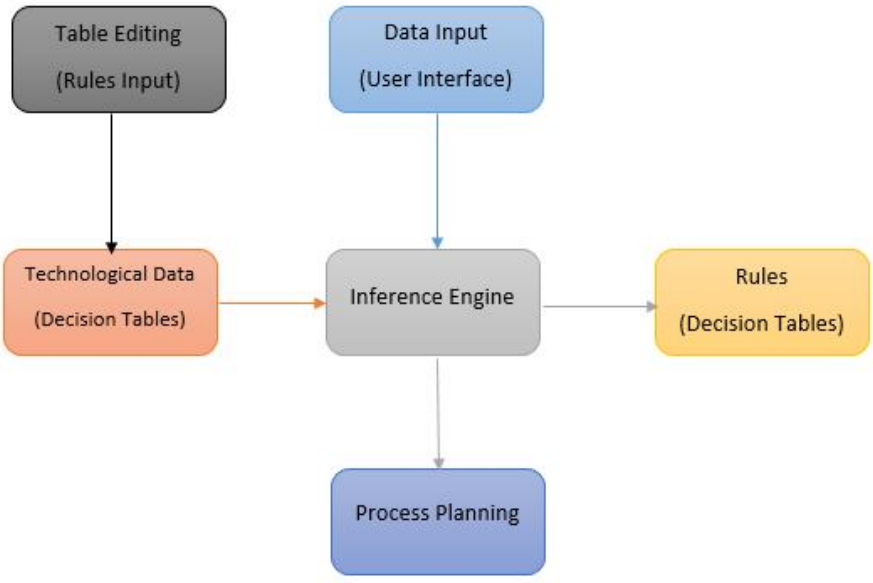


Fig.7 Knowledge based CAPP system for drawing dies.

6) *Web Based Technologies in Process Planning:* The Manufacturing application of web-based technology leads to interconnection and automation of various infrastructure mainly machine control unit in the shop floor area as well as ERP in production process. The integration of manufacturing and production can be carried out with the help of web-based technology but for that it requires updated infrastructure or some form of investments for creating adaptable infrastructure. In 2019, Murena and others used web-based process planning system to effectively bend varieties of sheet metal product using various bending press machine.[49] Mijodrag and others in 2017, presented e-CAPP in which authors discussed about a system for distributed collaborative system which assisted manufacturing enterprises and experts in discussing, suggesting, evaluating and selecting best process plans for manufacturing parts of same family.[50] Shan Wan and others in 2015 investigated a web-based process planning for machine tool maintenance and services and presented a collaborative maintenance planning system resulting to manage information and knowledge to support decision making.[51] A year earlier in 2014, Nagaraj and others implemented the web-based manufacturing process planning workflow in pressure die casting industry.[52] In year 2013, an Internet-based CAD/CAM/CAPP prototype system for the remote manufacturing of different mechanical components was evolved.[53] Earlier in year 2009, Agrawal and others addressed the DPP problem in the e-manufacturing by presenting a multiagent system consisting of three autonomous agents namely global manager, design, and optimization.[54A] In year 2008, Hu and others presented their work and tried to solve the operation sequencing problem for web-based process planning.[55]

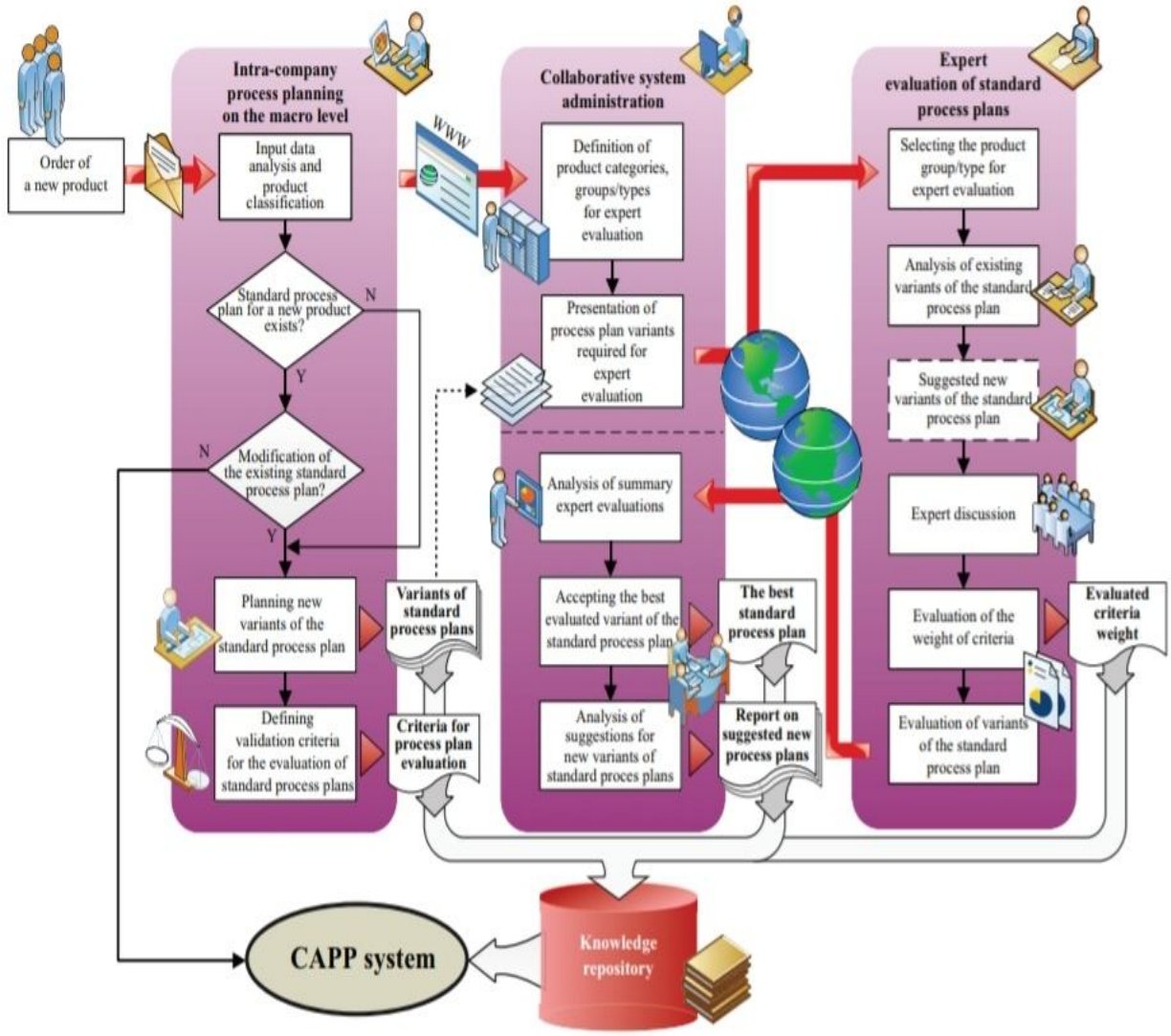


Fig.8 Web based technology in process planning. [a]
Source: Mijodrag et al. (2016)

7) *Hybrid Manufacturing Process Planning System:* In hybrid type two or more manufacturing process are integrated together e.g., Additive manufacturing and subtractive manufacturing has gain considerable recognition in recent years. The process extracts the benefit of individual processes while overcoming their limitation at the same time. With the integration of CAPP and Hybrid manufacturing, process planning has been optimized to a great extent. Evolution of computer-aided process planning for hybrid additive/subtractive process presented by Osama and others in year 2020 which discussed mathematical models to optimize part orientation as well as minimize additive and subtractive times. The feasibility and capability of the proposed approach and the optimized CAPP for the hybrid process are explained in a case study.[56] Ann-Kathrin in 2019, investigated a hybrid-manufacturing concept where standardized base parts were produced in large volume production first and then assembled to finalize by additive and subtractive manufacturing.[57] In year 2018, Basinger and others came up with an analysis in the form of a modular computer-aided process planning (CAPP) system for additive-subtractive hybrid manufacturing used for pockets, holes, and flat surfaces. [58] Newman in 2015 developed process planning for additive and subtractive manufacturing technologies.[59] Earlier in 2012, Zhu presented a novel process planning approach for hybrid manufacturing consisting of additive, subtractive and inspection processes.[60]

8) *Artificial Intelligence (AI) in CAPP:* In its simplest form, artificial intelligence is a field, which combines the application of computer and large datasets, to solve a given problem. In this methodology machines learn from experience, and then adjust to new inputs and perform the task more accurately and efficiently. With the help of this methodology, a large set of data is processed and patterns in the data are recognized to accomplish certain tasks. In 2020, Raffaele and others came up with a research to analyze, the scientific literature relating to the application of artificial intelligence and machine learning (ML) in various industries.[61] Zhang Bo in 2016 presented his work titled research into technology decision methods of CAPP Artificial Intelligence that discussed the method of automatic processing in CAPP and established an effective intelligent process decision-making system.[62] In year 2012, Prakash and others developed a mathematical model to analyse the randomized CIM environment using advanced CAPP systems.[63] Stryczek in 2007, reviewed computational intelligence in computer aided process planning mainly discussing about elements of a typical CAPP system such as FBM, part feature extraction, integration, process planning system development resulting to develop the CAPP systems.[64] Earlier in 2000, Chang and Chang generated multiple process plans using an integrated artificial intelligent, fuzzy logics.[65]

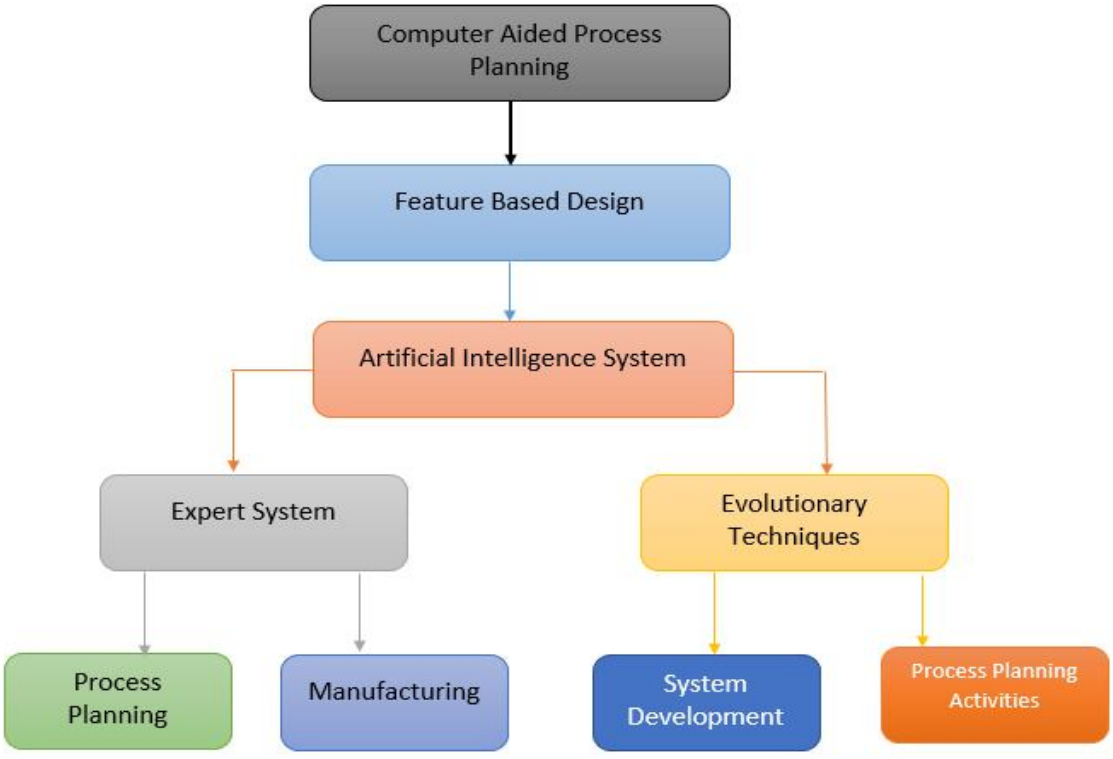


Fig.9 AI systems application in process planning and manufacturing.

9) *Virtual machining system for CAPP:* Virtual machining is defined as the application of computers to simulate and model the use of various machine tools for different manufacturing processes. This method replicates the behaviour and show the errors of a real environment in virtual reality systems. This leads to manufacture of products without its physical testing on the shop floor. As a result of which both time and cost of products can be reduced to a significant level. With the help of this it becomes easier to evaluate geometrical and dimensional flaw as well as cutting forces in machine tools. Parag and others in year 2017 presented their research titled computer numerical control machine tool information reusability within virtual machining systems in which authors reviewed research trends in the domain of virtual machining and later discussed how much of research has been taken on board by software vendors in order to facilitate machine tool information reusability.[66] Same year, wang and others investigated practical models of the cloud-based system and distributed process planning services were developed accordingly.[67] In 2013, Kim and Woo presented their work and developed a virtual corner detection system in their study to modify the process planning of parts using virtual machining system.[68] Earlier in year 2010, Ahmad and others presented applications of genetic algorithms in process planning which they came up with a systematic method for automatically selecting an optimal tool sequence for 2.5-axis pocket machining in their study.[69] Rajeev and others in 2008, generated the algorithm for optimal or nearly optimal identity of sequence of operation of prismatic part of minimum production.[54B] In 2000, Narita investigated a virtual machining simulation to analyse and optimize the cutting tool paths for milling operation in virtual environments.[70]

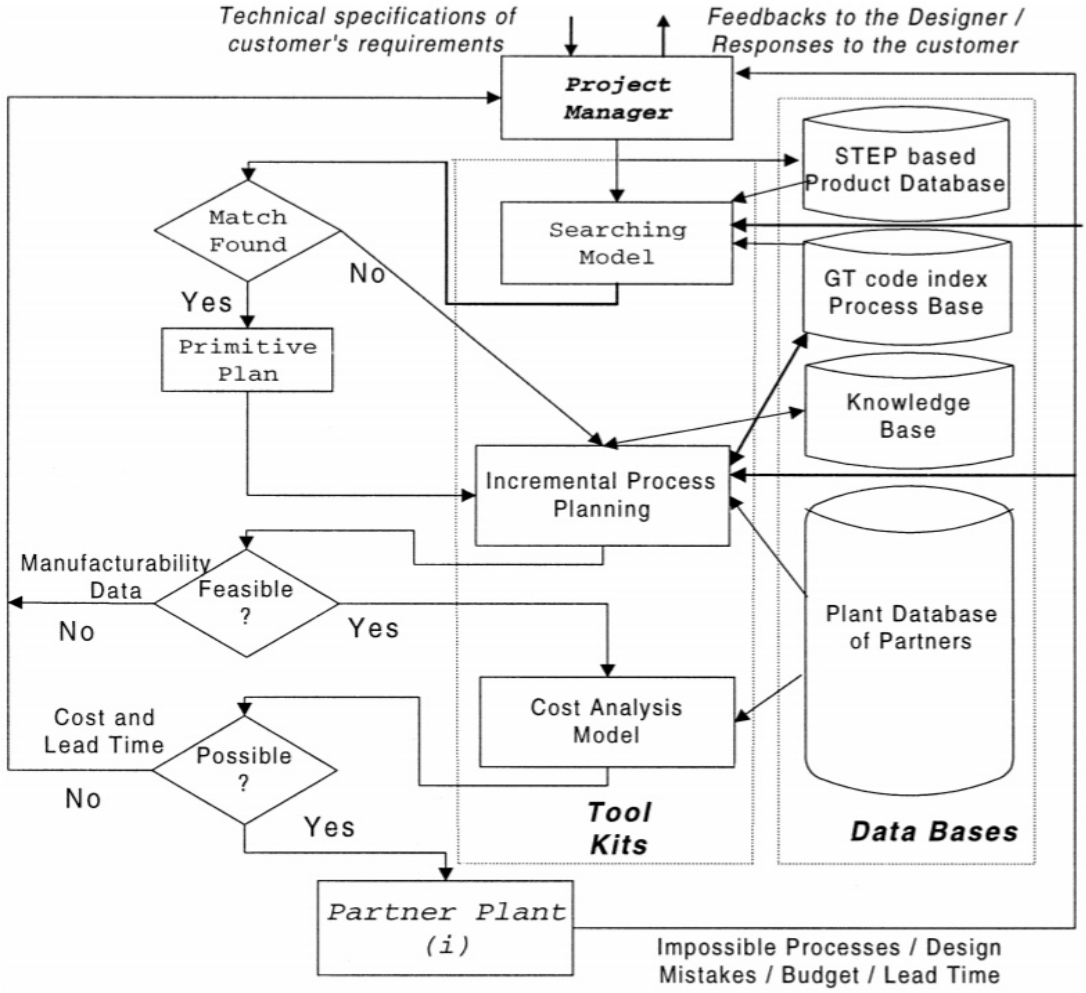


Fig.10 A reference architecture of CAPP system in a virtual company. [b]
Source: Yiliu Tu et al. (1998)

10) *STEP-Compliant CAPP:* Although with automation, a higher degree of accuracy in Design and manufacturing areas is achieved sharing of information between different system and software including CAPP were still lagging behind, to overcome this limitation the International Organisation for Standardisation (ISO) developed a methodology by which it is now possible to share the information between numerous other applications. Yu Zhang in 2021, presented an Intelligent feature recognition for STEP-NC-compliant manufacturing which was based on artificial bee colony algorithm and back propagation neural network.[71] In 2019, Vinod V Rampur and other came up with a computer aided process planning using STEP neutral file that proved useful for automotive parts.[19B] In 2012, Han and others came up with his work titled a framework of STEP-NC manufacturing system integrating CAD, CAPP, CAM and CNC in which a STEP-NC manufacturing system using networked manufacturing systems was presented.[72] In 2005, Amaitik proposed a STEP-based feature modeller for computer-aided process planning to be used for prismatic part.[73]

V. CONCLUSIONS

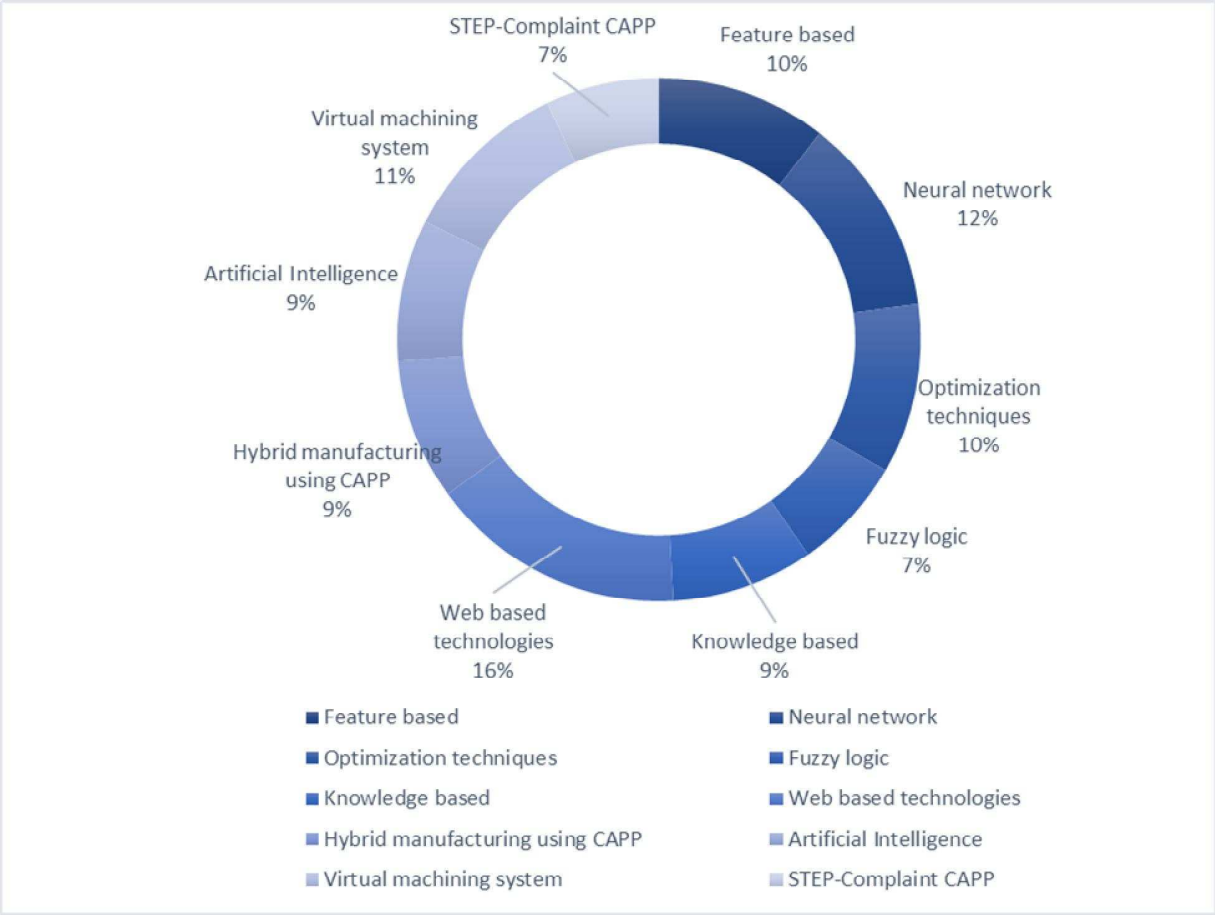


Fig.11 Contribution of various methods in the presented survey.

CAPP is need of the hour, as with the rapid growth and technological advancement in the field of manufacturing where to meet high demand requires high production rate. This demand can be fulfilled by integration of computer aided design (CAD) and computer aided manufacturing (CAM). So, at present there is a need of CAPP in all those manufacturing system which can justify the cost of production. Also, new demands of CAPP systems are created day by day due to increasing part complexities and rapid manufacturing developments. Nowadays, CAPP systems are connected by using web to share data between different steps of process planning.

In this survey, it has been found that CAPP tries to resolves most of the issues pertaining to planning, operation, setup, scheduling, machine selection, tool selection, job sequencing, reorganization, tool path generation, optimization and various others. Various research papers from different journals were studied to present the contribution of feature based and solid model-based process planning, neural network in CAPP, optimization techniques using CAPP, fuzzy logic in process planning, knowledge-based CAPP system, web-based technologies in process planning, hybrid manufacturing process planning, artificial intelligence in CAPP, virtual machining system for CAPP and STEP-compliant CAPP. Taking in account the number of prominent journals in various different years in and after year 2000, the contribution of each ten methods is displayed in figure 11.

VI. SUGGESTIONS

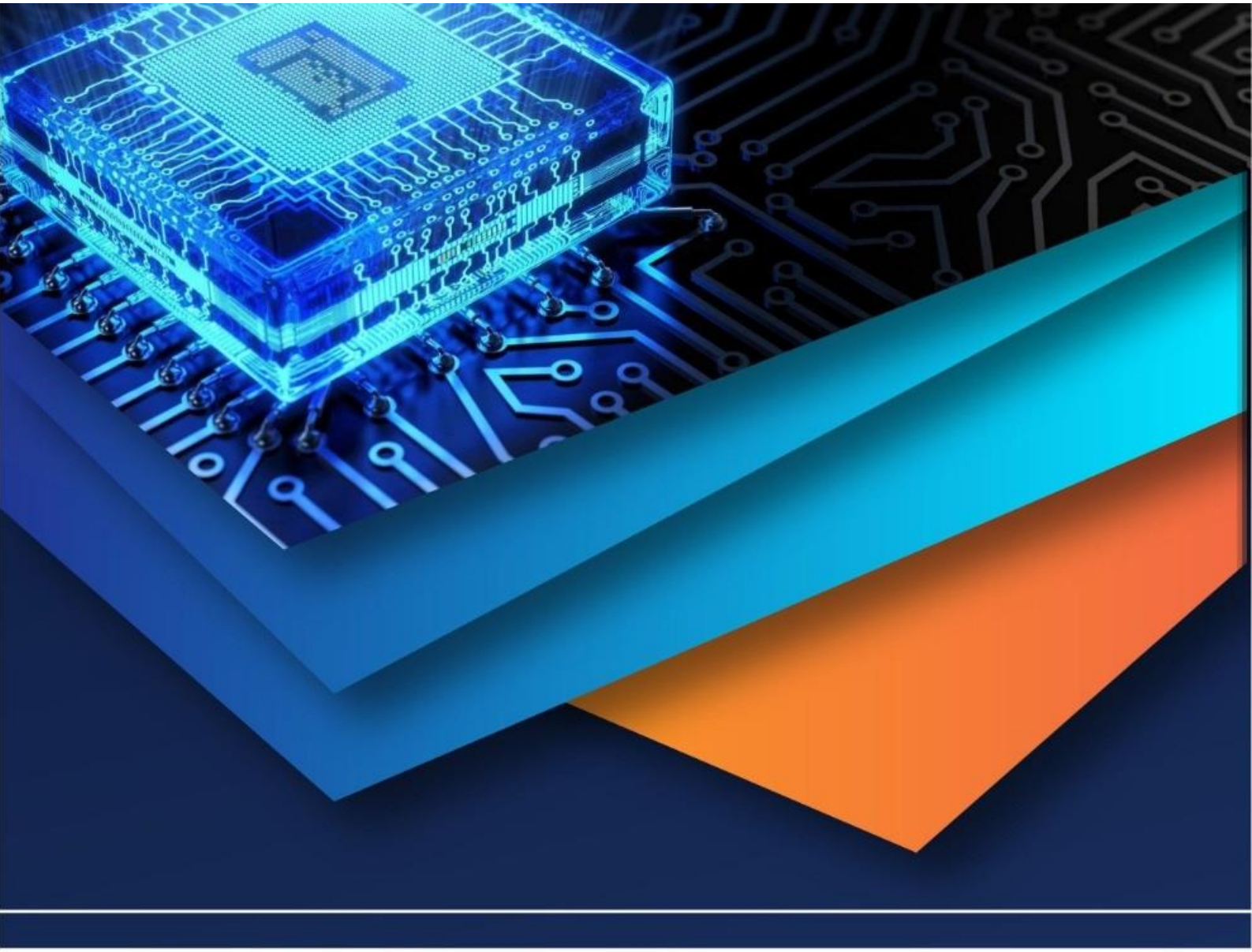
Manufacturing world day by day is moving towards more advance, intelligent and flexible technology. As a result, there is demand for advance and intelligent CAPP. While fulfilling these demand, global environmental issues pertaining to environmental changes should be taken into account. In future, the key area of research for CAPP should be such that it contributes towards green manufacturing environment, less energy consumption, non-traditional manufacturing processes, rapid prototyping process, micro manufacturing and nano manufacturing.

REFERENCES

[1] Harold J (1984) Computer-aided process planning: past, present and future. Int J Prod Res 22(2):253–266
[2] Ham I, Lu S (1988) Computer-aided process planning: the present and the future. CIRP Ann Manuf Technol 37(2):591– 601
[3] Gouda S, Taraman K (1989) CAPP: AAST, present and future. Society of Manufacturing Engineers, Dearborn.
[4] ElMaraghy H (1993) Evolution and future perspectives of CAPP. CIRP Ann Manuf Technol 42(2):739–751
[5] Kamrani A, Sferro P, Handelman J (1995) Critical issues in design and evaluation of computer aided process planing systems. Comput Ind Eng 29(1):619–623
[6] Leung H (1996) Annotated bibliography on computer-aided process planning. Int J Adv Manuf Technol 12(5):309–329
[7] Marri H, Gunasekaran A, Grieve R (1998) Computer-aided process planning: a state of art. Int J Adv Manuf Technol 14(4):261– 268
[8] Cay F, Chassapis C (1997) An IT view on perspectives of computer aided process planning research. Comput Ind 34(3):307– 337
[9] Kulkarni, P., Marsan, A., & Dutta, D. (2000). A review of process planning techniques in layered manufacturing. Rapid Prototyping Journal, 6(1), 18-35.
[10] Tan, W., & Khoshnevis, B. (2000). Integration of process planning and scheduling—a review. Journal of Intelligent Manufacturing, 11(1), 51-63.
[11] Ahmad, N., Haque, A., & Hasin, A. (2001). Current trend in computer aided process planning. Paper presented at the Proceedings of the 7th Annual Paper Meet and 2nd International Conference
[12] Zhang, W., & Xie, S. (2007). Agent technology for collaborative process planning: a review. The International Journal of Advanced Manufacturing Technology, 32(3-4), 315-25.
[13] Zab, A., Perusi, G., ElMaraghy, H. A., & Urbanic, J. (2007). Semi-generative macroprocess planning for reconfigurable manufacturing. Digital Enterprise Technology, 251-8. Springer.
[14] Li, X., Gao, L., Zhang, C., & Shao, X. (2010). A review on integrated process planning and scheduling. International Journal of Manufacturing Research, 5(2), 161-80.
[15] Xun Xu, Lihui Wang (2011) Computer-aided process planning-A critical review of recent developments and future trends. International Journal of Computer Integrated Manufacturing, doi 10.1080/2010.518632
[16] Yusof, Y., & Latif, K. (2014). Survey on computer-aided process planning. The International Journal of Advanced Manufacturing Technology, 75(1-4), 77-89
[17] Isnaini, M. M. R., & Shirase, K. (2014). Review of computer-aided process planning systems for machining operation–future development of a computer-aided process planning system–. International Journal of Automation Technology, 8(3), 317-32.
[18] Mazin Al wswasi, Atanas Ivanov, Harris Makatsoris (2018) A survey on smart automated computer-aided process planning (ACAPP) techniques. The International Journal of Advance Manufacturing Technology (2018) 97:809–832.
[19] A], [19B] Vinod V Rampur, Dr. Sachhidanand Reur (2017). Computer Aided Process Planning using STEP Neutral File for Automotive Parts. International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181.
[20] S.P. Leo Kumar, J.Jerald & S.Kumanan (2014). Automatic Feature Extraction & CNC Code Generation in a CAPP system for Micromachining, International Conference on Advances in Manufacturing and Materials Engineering, ICAMME 2014, 2014.07.531
[21] Vamsi Krishna, P., Shankar, N.V.S. & Surendra Babu (2011). Feature Based Modeling and automated process plan generation for turning components. Advances in Production Engineering & Management 6 (2011) 3, 153-162, ISSN 1854-6250
[22] Chandra R. Devireddy (2010). Feature-based modelling and neural networks-based CAPP for integrated manufacturing. International Journal of Computer Integrated Manufacturing ISSN: 0951-192x (Print) 1362-3052.
[23] Patil, L., & Pande, S. (2002). An intelligent feature-based process planning system for prismatic parts. International journal of production research, 40(17), 4431-47.
[24] Zhang, S., Ajmal, A., Wootton, J., & Chisholm, A. (2000). A feature-based inspection process planning system for co-ordinate measuring machine (CMM). Journal of materials processing technology, 107(1-3), 111-8.
[25] Case, K., & Wan Harun, W. (2000). Feature-based representation for manufacturing planning. International journal of production research, 38(17), 4285-300.
[26] Niechen Chena (2021). An evolutionary neural network approach to machining process planning: A proof of concept. 49th SME North American Manufacturing Research Conference (NAMRC 49,2021), Procedia Manufacturing, Volume 53, 2021, Pages 690-696.
[27] Guang Hu, Zhi Cao, Michael Hopkins, Conor Hayes, Mark Daly, Haiying Zhou, Declan M Devine, (2019). Optimizing the hardness of SLA printed objects by using the neural network and genetic algorithm. 29th International Conference on Flexible Automation and Intelligent Manufacturing, Volume 38, Pages 1-1848.

[28] K.K. Natarajan, J. Gokulachandran, (2018). Application of artificial neural network techniques in computer aided process planning - a review. International Journal of Process Management and Benchmarking , Volume 11, Issue 1 , ISSN: 1460-6739.
[29] Rojek I. (2017). Technological process planning by the use of neural networks. AI EDAM, 31(1), 1-15.
[30] Jun Wang, Haili Zhang and Zhangyue Su (2012). Manufacturing Knowledge Modeling Based On Artificial Neural Network for Intelligent CAPP, Applied Mechanics and Materials (Volume 127), AMM.127.310, 310-315.
[31] Romdhane BenKhalifa, Noureddine BenYahia and Ali Zghal (2010). Integrated neural networks approach in CAD/CAM environment for automated machine tools selection, Journal of Mechanical Engineering Research Vol. 2(2), pp. 25-38
[32] Sunil V, Pande S (2008) Automatic recognition of machining features using artificial neural networks. Int J Adv Manuf Technol 41(9):932–947.
[33] Lan Xuan PHUNG,Dich Van TRAN, Sinh Vinh HOANG,Son Hoanh TRUONG (2017). Effective method of operation sequence optimization in CAPP based on modified clustering algorithm. Journal of Advanced Mechanical Design, systems and manufacturing, Volume 11 Issue 1, Pages JAMDSM0001.
[34] S Mitin and P Bochkarev (2016). Mathematical modelling in the computer-aided process planning. IOP Conf. Ser.: Mater. Sci. Eng. 124 012077.
[35] Nallakumarasamy, G., Srinivasan, P., Raja, K. V., & Malayalamurthi, R. (2011). Optimization of operation sequencing in CAPP using simulated annealing technique (SAT). The International Journal of Advanced Manufacturing Technology, 54(5-8), 721-8
[36] Shao, X., Li, X., Gao, L., & Zhang, C. (2009). Integration of process planning and scheduling—a modified genetic algorithm-based approach. Computers & Operations Research, 36(6), 2082-96.
[37] Salehi, M., & Tavakkoli-Moghaddam, R. (2009). Application of genetic algorithm to computer-aided process planning in preliminary and detailed planning. Engineering Applications of Artificial Intelligence, 22(8), 1179-87.
[38] Siva Sankar, R., Asokan, P., Prabhaharan, G., & Phani, A. (2008). A CAPP framework with optimized process parameters for rotational components. International journal of production research, 46(20), 5561-87.
[39] Devendra G. Pendokhar, Taqui Z. Quazi (2014) Fuzzy logic-based drilling control process. International Journal of Scientific & Engineering Research, Volume 5, Issue 12, ISSN 2229-5518
[40] Saleh AMAITIK (2013). Fuzzy Logic Models for Selection of Machining Parameters in CAPP Systems, International Journal of Computer and Information Technology, ISSN: 2279 – 0764.
[41] Peidro, D., Mula, J., Jiménez, M., & del Mar Botella, M. (2010). A fuzzy linear programming-based approach for tactical supply chain planning in an uncertainty environment. European Journal of Operational Research, 205(1), 65-80.
[42] Leung, S. C., Tsang, S. O., Ng, W.-L., & Wu, Y. (2007). A robust optimization model for multi-site production planning problem in an uncertain environment. European Journal of Operational Research, 181(1), 224-38.
[43] Aliev, R. A., Fazlollahi, B., Guirimov, B., & Aliev, R. R. (2007). Fuzzy-genetic approach to aggregate production–distribution planning in supply chain management. Information Sciences, 177(20), 4241-55.
[44] S P Leokumar (2018). Knowledge based expert system in manufacturing planning: state of the art review, International Journal of Production Research, Vol 57, Issue 15-16, Pages 4766-4790.
[45] Ravi Yerigeri, Abhimanyu Posangiri, Dr.S.S.Hebbal (2015). A knowledge-based Computer Aided Process Plan and CNC code generation, International Journal of Scientific & Engineering Research, Volume 6, Issue 12, ISSN 2229-5518.
[46] Grabowik, C., Krenczyk, D., & Kalinowski, K. (2012). The hybrid method of knowledge representation in a CAPP knowledge-based system. Paper presented at the International Conference on Hybrid Artificial Intelligence Systems.
[47] Denkena, B., Shpitalni, M., Kowalski, P., Molcho, G., & Zipori, Y. (2007). Knowledge management in process planning. CIRP annals, 56(1), 175-80.
[48] Law, H.-W., Tam, H.-Y., Chan, A. H., & Hui, I. (2001). Object-oriented knowledge-based computer-aided process planning system for bare circuit boards manufacturing. Computers in industry, 45(2), 137-53.
[49] Eriyeti Murena, Khumbulani Mpofu, Alfred T Ncube, Olasumbo Makinde, John A Trimble & Xi Vincent Wang (2019). Development and performance evaluation of a web-based feature extraction and recognition system for sheet metal bending process planning operations, International Journal of Computer Integrated Manufacturing. Volume 34, Issue 6, pages 598-620.
[50] [a] MijodragMilosevic, DejanLukic, AcoAntic, BojanLalic, MirkoFicko, GoranSimunovic (2017). e-CAPP: A distributed collaborative system for internet-based process planning, Journal of Manufacturing Systems Volume 42, January 2017, pages 210-223.
[51] ShanWan, JamesGao, DongboLi, YifeiTong & FeiHe (2015). Web-based Process Planning for Machine Tool Maintenance and Services, Procedia CIRP, Volume 38, 2015, Pages 165-170.
[52] M. Nagaraj, Dr G Madhan Mohan (2014). Implementation of Web Based Manufacturing Process Planning Workflow in a Pressure Die Casting Industry, IOSR Journal of Computer Engineering 16(2):126-134.
[53] Julio Cesar T. Benavente, Joao Carlos E. Ferreira, Cassio Marcos Goulart, Victor Gomes de Oliveira (2013) A STEP-NC compliant system for the remote design and manufacture of mechanical components through the Internet, International Journal of Computer Integrated Manufacturing 26(5):412-4.
[54] [54A], [54B] Rajeev Agrawal, S. K. Shukla, S. Kumar, M. K. Tiwari (2008). Multi-agent system for distributed computer-aided processnplanning problem in e-manufacturing environment, Int J Adv Manuf Technol 44:579–594.
[55] Hu C, Li Z, Zheng L, Li N, Wen P (2008) An XML-based implementation of manufacturing route sheet documents for context-sensitive and web-based process planning. Int J Comput Integr Manuf 21(6):647–656.
[56] Osama Abdulhameed, Abdulrahman Al-Ahmari, Syed Hammad Mian, Abdulmajeed Dabwan, Hisham Alkhalefah (2020). Evolution of Computer-Aided Process Planning for Hybrid Additive/Subtractive Process, Advances in Material Science and Engineering, Vol. 2020, ID 7458239, pages 21.
[57] Ann-Kathrin Reichlera, Julian Redekerb, Felix Gabriela, Fabio Kai Falkea, Thomas Vietorb, Klaus Drodera (2019). Combined Design and Process Planning for Incremental Manufacturing. Procedia CIRP, 93, 2020, Pages 927-932
[58] Basinger, K. L., Keough, C. B., Webster, C. E., Wysk, R. A., Martin, T. M., & Harrysson, O. L. (2018). Development of a modular computer-aided process planning (CAPP) system for additive-subtractive hybrid manufacturing of pockets, holes, and flat surfaces. The International Journal of Advanced Manufacturing Technology, 96(5-8), 2407-20.
[59] Newman, S. T., Zhu, Z., Dhokia, V., & Shokrani, A. (2015). Process planning for additive and subtractive manufacturing technologies. CIRP annals, 64(1), 467-70.

[60] Zhu, Z., Dhokia, V., & Newman, S. (2012). A novel process planning approach for hybrid manufacturing consisting of additive, subtractive and inspection processes. Paper presented at the 2012 IEEE International Conference on Industrial Engineering and Engineering Management.
[61] Raffaele Cioffi, Marta Travaglioni, Giuseppina Piscitelli, Antonella Petrillo, Fabio De Felice (2020). Artificial Intelligence and Machine Learning Applications in Smart Production: Progress, Trends, and Directions, Sustainability 2020, 12(2), 492
[62] Zhang Bo (2016). Research into Technology Decision Methods of CAPP Artificial Intelligence, International Conference on Education, Management, Computer and Society (EMCS 2016).
[63] Prakash, A., Chan, F., & Deshmukh, S. (2012). Application of knowledge-based artificial immune system (KBAIS). for computer aided process planning in CIM context. International journal of production research, 50(18), 4937-54.
[64] Stryczek, R. (2007). Computational intelligence in computer aided process planning-a review. Advances in manufacturing science and technology, 31(4), 77-92.
[65] Chang, P.-T., & Chang, C.-H. (2000). An integrated artificial intelligent computer-aided process planning system. International Journal of Computer Integrated Manufacturing, 13(6), 483-97.
[66] Parag Vichare, Xianzhi Zhang, Vimal DhokiaVima (2017). Computer numerical control machine tool information reusability within virtual machining systems, Proceedings of the institution of Mechanical Engineers, Part B- Journal of Engineering Manufacture, Volume: 232, issue 4, pages: 593-604.
[67] Xi Vincent Wang, Mohammad Givehchi, Lihui Wang (2017) Manufacturing system on the cloud: a case study on cloud-based process planning, 50th CIRP Conference on Manufacturing Systems, Procedia CIRP 63 (2017) 39 – 45
[68] Kim, S.-H., &Woo, Y. (2013). Determination of maximum cutter sizes for planar milling by virtual corner detection. International Journal of Precision Engineering and Manufacturing, 14(9), 1565-70.
[69] Ahmad, Z., Rahmani, K., & D'Souza, R. M. (2010). Applications of genetic algorithms in process planning: tool sequence selection for 2.5-axis pocket machining. Journal of Intelligent Manufacturing, 21(4), 461-70.
[70] Narita, H., Shirase, K., Wakamatsu, H., & Arai, E. (2000). Pre-process evaluation of end milling operation using virtual machining simulator. JSME International Journal Series C Mechanical Systems, Machine Elements and Manufacturing, 43(2), 492-7.
[71] Yu Zhang, Yongsheng Zhang, Kaiwen He, Dongsheng Li, Xun Xu, Yadong Gong (2021). Intelligent feature recognition for STEP-NC-compliant manufacturing based on artificial bee colony algorithm and back propagation neural network, Journal of Manufacturing Systems, doi-j.jmsy.2021.01.018.
[72] Han, Z. Y., Hu, P., Han, D. D., & Fu, H. Y. (2012). A Framework of STEP-NC Manufacturing System Integrating CAD, CAPP, CAM and CNC. Paper presented at the Applied Mechanics and Materials.
[73] Amaitik SM, Kilic¸ SE (2007). An intelligent process planning system for prismatic parts using STEP features. Int J Adv Manuf Technol 31(9-10):978–993
[74] Yiliu Tu, Xulin Chu (2000). Computer-aided process planning in virtual one-of-a-kind production, Computers in Industry 41(1):99-110.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-022-12059-z

An intelligent railway surveillance framework based on recognition of object and railway track using deep learning

Rajiv Kapoor [1] · Rohini Goel [2] · Avinash Sharma [2]

Received: 19 February 2021 / Revised: 14 July 2021 / Accepted: 4 January 2022
Published online: 14 March 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

In high speed railways, the intelligent railway safety system is necessary to avoid the accidents due to collision between trains and obstacles on the railway track. The unceasing research work is being performed to reinforce the railway safety and to diminish the accident rates. The rapid development in the field of deep learning has prompted new research opportunities in this area. In this paper, a novel and efficient approach is proposed to recognize the objects (obstacles) on the railway track ahead the train using deep classifier network. The 2-D Singular Spectrum Analysis (SSA) is utilized as decomposition tool that decomposes the image in useful components. That component is further applied to the deep classifier network. The obstacle recognition performance is enhanced by the combination of 2D-SSA and deep network. This method also presents a novel measure to identify the railway tracks. In addition, the performance of this approach is analyzed under different illumination conditions using OSU thermal pedestrian benchmark database. This system can be a tremendous support to curtail rail accidental rate and monetary loads. The results of proposed approach present good accuracy as well as can effectively recognize the objects (obstacles) on the railway track which helps to the railway safety. It also achieves a better performance with 85.2% accuracy, 84.5% precision and 88.6% recall.

Keywords Thermal imaging · Singular spectrum analysis · Singular value decomposition (SVD) · Region proposals · Faster R-CNN

✉ Rajiv Kapoor
rajivkapoor.dtu@gmail.com

[1] Department of Electronics and Communication Engineering, Delhi Technological University, Delhi 110042, India

[2] Computer Science and Engineering Department, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana, India

1 Introduction

With the development of high speed railways, the safety necessities [37] of the railways are increasing day by day because the rapid speed of trains lead to rise in train mishaps. The obstacle on the railway tracks is one of the main factors that are responsible for the mishaps. The obstruction are pedestrians, vehicles crossing the tracks, animals meandering on the tracks [42] and some other heavy non-living objects fallen on the track from the overpass. In both developed and developing countries, a large number of railway crossing have not proper safety means like gates and lights. These issues cause challenges for the travelers as well as lead to monetary loss to railways as train cancelations and accidental compensations to aggrieved individuals. These issues can be resolved by utilizing a warning system that can significantly diminish the train mishaps by prior detection of the obstacles on the railway tracks. Various researches have been worked to detect and recognize the obstacles on the railway tracks. Zhou et al. [49] combined object detection and background learning by introducing a unified framework named DEtecting Contiguous Outliers in the low Low-rank Representation (DECLOR) for detection of moving object.

Chen et al. [8] proposed a technique dependent on image matching and frame coupling to handle the object detection issue caused by moving camera and movement of the object. Berg et al. [5] proposed a approach for obstacle recognition on the railway track utilizing monocular thermal camera and procured a novel data collection. Sinha et al. [39] introduced an object detection method on railway tracks utilizing vibration sensors and separate out the signal from significant level acoustic noise utilizing a novel Monte Carlo based Bayesian analysis. V. Amaral et al. [2] discussed obstacle identification framework in railway level crossing utilizing 3D point clouds procured with 2D laser scanner. Wu et al. [46] introduced a coarse to fine thresholding plan on particle trajectories in video sequence of moving object captured by moving camera. Karaduman [22] proposed the obstacle identification framework utilizing laser distance meter and rail mounted camera. Manikandan et al. [29] proposed an obstacle recognition approach utilizing thermal camera and ADA boost algorithm. The rapid growth in field of machine learning and deep neural network is also contributing in the field of railway applications.

Mittal et al. [31] introduced a vision based railway track observing methodology utilizing deep learning classifier for uncontrolled real world data. Krummenacher et al. [24] introduced two machine learning strategies dependent on SVM and neural network to recognize rail wheels. Garcia et al. [13] proposed a methodology for independent train stop activity utilizing monocular vision-based methodology and Deep-Learning models. Kapoor et al. [21] demonstrated a framework using Hough transform and HSV segmentation with deep learning for recognition of railway track and obstacles on it. The scene of the railway track has something in common with the road scene; therefore the research work in the field of road information detection also helps to broaden the viewpoint. Qi Wang et al. [43] proposed a siamesed fully convolutional network (s-FCN) to segment the road regions in RGB images to improve the detection performance. Wang et al. [44] also weakly supervised adversarial domain adaption to enhance the segmentation performance from the synthetic data to real scenes using three deep neural networks. Sudha et al. [41] utilized the YOLO deep network and improved visual background extraction algorithm to detect multiple vehicles in the input videos. Zeng et al. [38] built a fusion network for segmentation of lane line and detection of road elements for autonomous driving. Nayak et al. [33] also demonstrated the use of various deep learning models for detection application particularly for COVID-19 detection in chest X-ray images

whereas in the proposed work deep detection frameworks are used for thermal images of railway track.

Other than obstacles on the railway track, the adverse weather conditions like rain, haze, darkness, cloudy weather are also big challenge for railway safety. As these conditions have different level of illumination that directly affect the visibility of the moving train at high speed. Many times low visibility conditions can cause major loss to railways. The system should also be invariant to illumination so that it can provide proper visibility in any type of adverse weather condition. Mangale et al. [28] suggested a methodology to identify object for low illumination and different weather condition thermal images using directed acyclic graph (DAG). M. Kristo et al. [23] discussed an approach for thermal object detection in adverse weather conditions using YOLO framework.

Decomposition is a fundamental stage that is imperative for providing proper information to the classifier for image recognition. Sharma et al. [34] presented a method for the automated emotion recognition using higher order statistics (i.e. third order cumulants). The sub bands of signals are applied to particle swarm optimization to remove redundant information. Then the deep learning algorithm is used to predict the human emotions. Srirangam et al. [26] utilized the Time frequency matrix of EEG signals that is evaluated by Fourier Synchro Squeezing Transform (FSST) and Wavelet Synchro Squeezing Transform (WSST) to classify focal and non focal EEG signals using deep CNN. Chaudhary et al. [6] presented an approach for decomposition using 2D Fourier Bessel Series Expansion based empirical wavelet Transform (FBSE-EWT). The sub images from 2D-FBSE-EWT are used with machine learning methods and ResNet-50 based method to detect glaucoma. Chaudhary et al. [7] also used Fourier Bessel Series Expansion-based Decomposition (FBSED) with various pre-trained deep network classifiers to diagnose the COVID-19 using X-ray and CT images. The 2D-EMD [12] approach outperforms numerous relative methods like wavelet based image decomposition [11]. In any case, the execution of 2D-EMD is truly tedious because it is an iteration based process [20]. Whereas the 2D-SSA approach requires no iterations therefore it significantly curtails computational cost with improved efficiency [47]. The 2D-SSA approach outperforms 2D-EMD and many other number of state of art approaches in terms of accuracy and also requires very less execution time than 2D- EMD [9]. The 2-D SSA algorithm is really useful for extracting more precise discriminating patterns from original data scenes [17]. There is always a requirement of an object recognition framework that is capable of achieving high safety on the railway tracks. The achievements of the deep learning (RCNN [14], Fast/Faster RCNN [15, 36], YOLO [35] & SSD [25] etc.) in the field of object recognition motivated to analyze their performance in obstacle recognition on the railway track. The performance of the 2D-SSA [18] in the field of image enhancement and image restoration [3] further motivates to utilize its capabilities along with deep network to enhance the performance of object recognition for railway safety.

In this paper, the proposed approach automatically performs railway track identification and object (obstacle) recognition on the railway track. In this approach, a system is designed using 2D-SSA that decomposes the input frame into various discriminative components (patterns). This approach is intended to identify railway tracks by using particular component (pattern) that contains the information about tracks. Then component (pattern) having the object (obstacle) is applied to the deep networks to efficiently recognizes the obstacles on the railway track. In addition, the major contributions of the paper are:

- The 2D-SSA approach is first time used to extract the railway track information to detect the railway track in railway safety applications.
- To the best of our knowledge, this is the earliest effort to use 2D-SSA with deep network classifier to enhance the object recognition performance.
- The performance of different deep networks with 2D-SSA is analyzed to choose suitable one for railway safety applications.
- This proposed approach also provides an efficient recognition performance under different illumination conditions when tested with OSU thermal pedestrian database from OTCBVS benchmark dataset collection [10].

The organization of paper in various subsections is as follows: Section 2 depicts the proposed strategy in detail including detection of railway tracks and recognition of obstacles on the railway track. The Section 3 describes the result and discussions. At the end, the conclusion and future scope provided in the Section 4.

2 Proposed method

The proposed technique is utilized to identify the railway track along with the obstacle on it from railway track thermal video (https://www.youtube.com/watch?v=xzGc71JFiBI). The basic block diagram of the proposed strategy is shown in Fig. 1. The frames are separated from the thermal video sequence and motion artifacts are compensated from the frame. Then the frames are decomposed into the 'g' distinct components having different important information of the scene using 2D- singular spectrum analysis (2D-SSA). Then the features of the railway track are extracted from 'l^{th}' component to detect the railway track. Similarly, the 'k^{th}' component having information about the obstacle is applied to the deep network (Faster R-CNN/ SSD/ Yolov2/Yolov5[27]) to recognize obstacle. The final predictions (railway track and obstacle) are combined to illustrate the final results at the output. The distinct steps of the proposed approach are illustrated in algorithm 1 and explained in the following subsections.

2.1 Compensation of motion artifacts in the frames

The thermal railway track video may have motion artifacts. The motion blur is the artifact of fundamental importance in the captured images when the camera is not in focus due to the train motion. In this work, the effect of motion blur in the frame is compensated by using wiener filtering [48] as it has shown good performance in removal of motion blur in the images. The motion artifacts in the form of blur are compensated by wiener filter as:

$$\begin{aligned} I(u,v) &= H_w(u,v) * F(u,v) \\ and \quad H_w(u,v) &= \frac{S^*(u,v)}{|S(u,v)|^2 + K} \end{aligned} \tag{1}$$

Where $F(u,v)$ is the input frame with motion artifacts, $H_w(u,v)$ is the wiener filter function, S(u,v) is the point spread function, k is the reciprocal of signal to noise ratio and $I(u,v)$ is the output of wiener filter used for further processing.

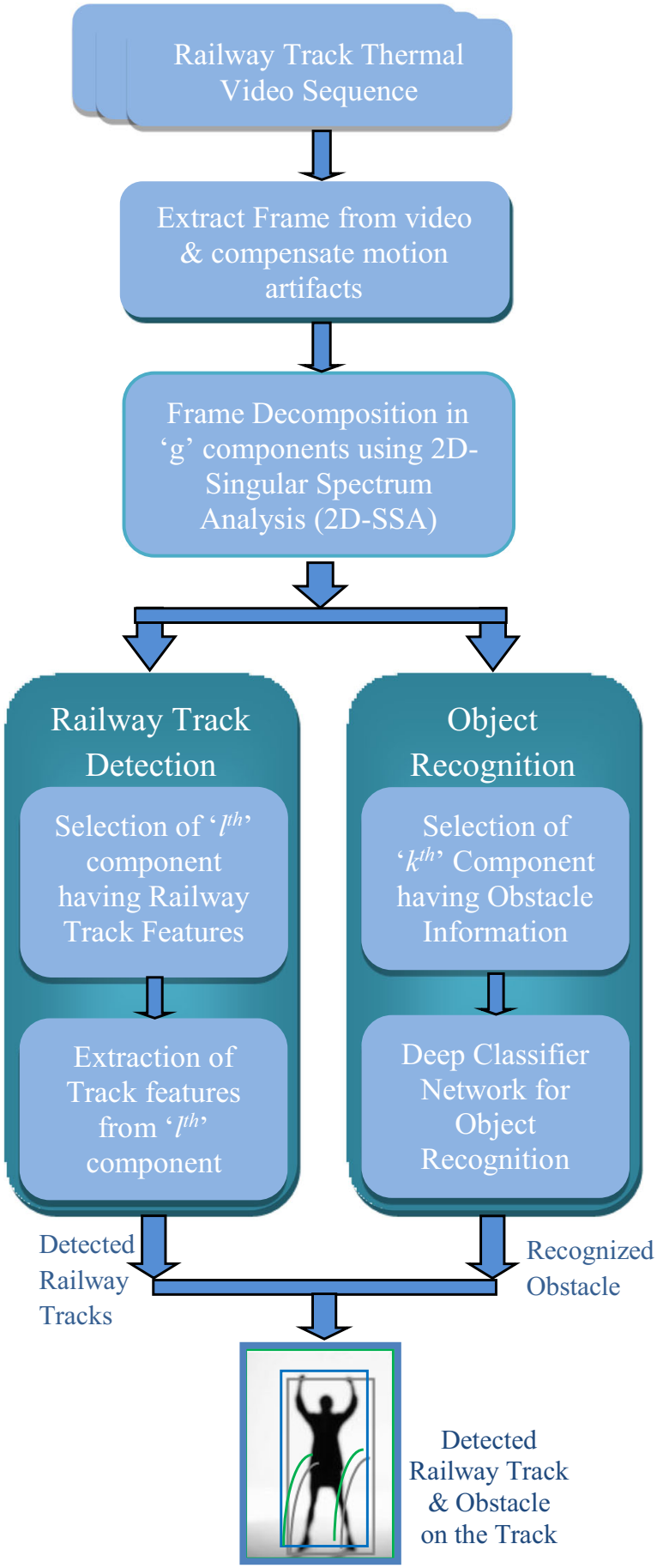


Fig. 1 Basic block diagram of the proposed method used to recognize railway track and along with object on the track

2.2 Decomposition of frames using 2D Singular Spectrum Analysis

The efficient information extracted from the thermal video frames can enhance the performance of recognition system. In this work, 2D-SSA [47] is used as an efficient decomposition tool from thermal images. When Singular Spectral Analysis (SSA) is implemented on the 2-D signal i.e. image, the steps of implementation are embedding, the Singular Value Decomposition (SVD) [40], diagonal averaging and grouping like 1-D SSA. The embedding and diagonal averaging steps have to be modified from 1-D to 2-D and rest of the stages remain identical with 1-D SSA. The steps of the 2D-SSA algorithm with mathematical explanation are discussed as below:

2.2.1 Embedding an image

In the embedding of the 2-D signal (image), the initial stage is to transform the image (2D-data) into another matrix form i.e. known as trajectory matrix [16]. The trajectory matrix not only contains all the information of an image but also retain the neighborhood information. Consider an image *'I'* of size *sxt*. To construct trajectory matrix *'T'*, a window '*B'* of size $m \times n$ is considered, where $1 \leq m \leq s$ and $1 \leq n \leq t$. Referring top left point of an image is reference point of the window. The reference points *(x, y)* ranges with $1 \leq x \leq -m = s - m + 1$ and $1 \leq y \leq -n = t - n + 1$. The path of movement of window *'$B_{x,y}$'* on image is given by reference points (x, y) as follows:

$$(x,y) \rightarrow \begin{cases} (x,y+1) if y \leq t-n \\ (x+1,1) if y = t-n+1 and x \leq s-m \end{cases} \tag{2}$$

Then, the elements of the each window *'$B_{x,y}$'* is restructured as columns by transposing rows one by one as:

$$\overrightarrow{B_{x,y}} = \left(I_{x,y}, I_{x,y+1} - - - I_{x,y+n-1} - - - I_{x+m-1,y+n-1}\right) \tag{3}$$

where $\boldsymbol{I}_{x,y}$ is the pixel value at location *(x,y)* and $\overrightarrow{B_{x,y}}$ is coloumn vector for window *'$B_{x,y}$'* with reference point *(x,y)*. Finally all the column vectors $\overrightarrow{\mathrm{B}_{\mathrm{x,y}}}$ are arranged into trajectory matrix *'T'* as:

$$T = \left[\overrightarrow{B}_{1,1}, \overrightarrow{B}_{1,2}, - - \overrightarrow{B}_{1,-n}, \overrightarrow{B}_{2,1}, - - \overrightarrow{B}_{2,-n}, \overrightarrow{B}_{3,1}, - - \overrightarrow{B}_{-m,-n}\right] \tag{4}$$

2.2.2 Singular Value Decomposition (SVD) of Trajectory matrix

The next step is to perform Singular Value Decomposition (SVD) of trajectory matrix *'T'*. The SVD expansion of *'T'* can be obtained through the eigen decomposition of the lag covariance matrix $C = TT^T$. The SVD of *'C'* can be represented as follows [30]:

$$C = TT^T = \sum_{k=1}^{K} \lambda_k E_k E_k^T \tag{5}$$

Where $\lambda_k(k = 1, 2.\mathrm{K})$ are the eigen values of TT^T and E_k are normalized eigen vectors corresponding to λ_k.

2.2.3 Process of diagonal averaging

The process of diagonal averaging is performed with the help of two step hankelization [19] process. Firstly, apply hankelization within each block and then applied between blocks. The averaging of the value belongs to the same element in the image is carried out. This enables to decompose the image into several components extracted based on SVD.

$$t^{(k)} \leftarrow \mathrm{DiagAveraging}(\boldsymbol{T_k}) \tag{6}$$

Where, $t^{(k)}$ is the 2D signal projected from T_k after diagonal averaging for $(k = 1, 2.\mathrm{K})$ and T_k is the *'k^{th}'* weighted orthogonal matrix which is given as:

$$T_k \leftarrow \lambda_k E_k E_k^T \tag{7}$$

2.2.4 Grouping of 2D projected signals

Then the next step is grouping of the *'K'* number of 2D projected signals $t^{(k)}$ into *'g'* disjoint groups using Hierarchical Cluster Analysis (HCA) [45] to obtain 'g' components of the input image as below:

$$I_i \leftarrow \mathrm{HCA}\left(\left\{\mathrm{t}^{(1)}, \mathrm{t}^{(2)}, .\mathrm{t}^{(\mathrm{K})}\right\}, \mathrm{g}\right) \tag{8}$$

The *'g'* groups of the 2D projected signals are formed as $s^{(1)}, s^{(2)} s^{(g)}$ for $1 \leq g \leq K$. The every group $s^{(i)}$ represents the decomposed component of the input image i.e. the sum of all the 2D projected signals of group *'i'* which is represented as:

$$\begin{array}{c} \textbf{For i}\leftarrow\textbf{1 to } g \textbf{ do} \\ \left|s^{(i)} \leftarrow \sum_{k \epsilon I_i} \boldsymbol{t}^{(k)}\right. \\ \textbf{End} \end{array} \tag{9}$$

The final *'g'* decomposed components of the input image are represented as 'S' which is given as:

$$S = \left\{s^{(1)}, s^{(2)}, .s^{(g)}\right\} \tag{10}$$

Once the input frame is decomposed into *'g'* components, then the each component is analyzed to identify that which component contains the desired information i.e. information about railway track and obstacles on the tracks.

2.3 Railway track detection

2.3.1 Selection of component having railway track features

The first and most significant part in the proposed work is the recognition of railway track. The frame is decomposed into various discriminative components (patterns) using 2D Singular spectrum analysis (2D-SSA). The discriminative components contain vertical, horizontal and diagonal information of the scene. The pattern of tracks most resembles with the component having diagonal features. The *'l^{th}'* component that contains the strong pattern for the railway track is selected from the *'g'* discriminative components of the particular frame as given below:

$$S_{Track} = s^{(i)}|_{i=l} \tag{11}$$

Where, $s^{(i)}|_{i=l}$ is the *'l^{th}'* discriminative component (pattern) having strong information about railway track.

2.3.2 Extraction of track features from selected component

Then the coordinates of the railway tracks are extracted from the *'l^{th}'* component using thresholding segmentation. The track features are given as:

$$S_{Seg}(i,j) = \begin{cases} 1\left(if T_{low} \le S_{Track}(i,j) \le T_{high}\right) \\ 0 else \end{cases} \tag{12}$$

Where, '1' and '0' addresses white and dark color respectively, $S_{Track}(i, j)$ is discriminative component (pattern) having strong information about railway track, $S_{Seg}(i,j)$ is output image having railway track features, T_{low} and T_{high} are the lower & upper limit threshold. These coordinate locations are highlighted on the particular frame for the representation of the railway track. Hence the railway tracks are recognized in each frame of the video sequence.

2.4 Object recognition on the railway track

2.4.1 Selection of component having railway track features

Once the railway tracks are identified, the subsequent step is to recognize objects (obstacles) on the railway tracks. Initially the *'k^{th}'* component (having horizontal information) that contains the strong pattern for the obstacles on railway track is selected from the *'g'* discriminative components of the particular frame which is given as:

$$S_{Obstacle} = s^{(i)}|_{i=k} \tag{13}$$

Where, $s^{(i)}|_{i=k}$ is the *'k^{th}'* discriminative component (pattern) having strong information of obstacles on the railway track. After the extraction of components having obstacles information, this particular component is initialized to the deep network to recognize the obstacles.

2.4.2 Deep network for object recognition

Faster Region based Convolutional Neural Network (Faster R-CNN) [36] is utilized to perceive obstacles on the railway tracks. The Faster R-CNN comprises of two modules: initial, a deep fully convolutional network used to propose regions and second is the Fast R-CNN detector [15] that identifies the objects using region proposals. In default setup, there are anchors at position of image having 3 scales and 3 ratios. At last, the Fast R-CNN network is utilized for classification which has two fully connected layers. The one layer classifies the proposals in N+ 1 distinctive class (where N class and one background class for eliminating bad proposals). Another fully connected layer is utilized for better adjustment of bounding box for 'N' classes utilizing regression prediction. The anchors those overlap the ground truth with an Intersection over Union (IOU) greater than 0.7 are classified as foreground and the anchors don't overlap any ground truth object (having IOU under 0.3) is classified as background. The loss function to be limited is given as:

$$F_{Loss}(\{m_v\},\{n_v\}) = \frac{1}{C}\sum_{v} L_c\left(m_v, m_v^*\right) + \gamma\frac{1}{R}\sum_{v} m_v^* L_r\left(n_v, n_v^*\right) \tag{14}$$

Where, m_v is the predicted probability of anchor v being an object. Vector n_v indicates the parameterized coordinates of predicted bouncing box. L_c and L_r are the classification and regression loss respectively.

The object identification strategy YOLOv2 [35] divides the image into S × S grids and predicts B bounding box and C class probability for every framework cell. Each bounding box comprises of five predictions and object confidence. The object confidence addresses the dependability of existing object in the box. Nonetheless, as an overall object identification model, YOLOv2 is appropriate to cases where there are a variety of classes to be distinguished, and the distinctions among the classes are huge. At last, the non-maximum suppression (NMS) strategy is applied to dispose of redundant bounding boxes.

The most recent model of the YOLO object recognition model is now YOLOv5 [27], released by Glenn Jocher in 2020. YOLOv5 depends on the PyTorch structure. The YOLOv5s is the smallest model, and YOLOv5m, YOLOv5l and YOLOv5x are according to the increasing size. As the network size increases, its performance may likewise increases, at the expense of extra processing times. Accordingly, the bigger models may just be helpful for complex issues where enormous datasets are available.

During preparing, SSD [25] need to figure out which default boxes compare to ground truth detection and train the network accordingly. For each ground truth box, the selection is done from default boxes that vary over location, aspect ratio, and scale. The SSD training objective is drived from the MultiBox objective yet is extended to deal with various object classes. The overall objective loss function is a weighted sum of the localization loss &; the confidence loss.

Algorithm 1: Railway Track Detection & Object Recognition

Input:
Thermal Video Sequence of railway Track, g: the number of component to extract m & n: window size (the embedding dimension).

Output:
Railway tracks and obstacles.

begin

1. Extract frames ***I(i,j)*** from the video sequence**.**
2. Decompose the frame ***I(i,j)*** into elementary components using automatic SSA.

$\mathbf{T} \leftarrow BuildTrajectoryMatrix(\boldsymbol{I, m, n})$

$R \leftarrow rank(\mathbf{T})$

Apply Sing SVD$(\mathbf{T})$

For $k \leftarrow 1$ to R **do**

$T_k \leftarrow \boldsymbol{\lambda_k E_k E_k^T}$

$t^{(k)} \leftarrow \text{DiagAveraging}(\boldsymbol{T_k})$

End

$I_i \leftarrow \text{HCA}\left(\{t^{(1)}, t^{(2)}, \ldots . t^{(R)}\}, g\right)$

For $i \leftarrow 1$ to g **do**

$s^{(i)} \leftarrow \sum_{\boldsymbol{k \epsilon I_i}} \boldsymbol{t^{(k)}}$

End

$\boldsymbol{Return}\{s^{(1)}, s^{(2)}, \ldots . . s^{(g)}\}$

Case 1: Railway Track Detection.

3. Select the 'l^{th}' component having railway track features from the 'g' output components.
4. Extract the railway track features from the 'l^{th}' component.

Case 2: Object Recognition

5. Implement the deep network to identify the object from the 'k^{th}' obstacle containing component.
6. The railway track coordinates and object information are unified to generate the final output.

End

2.4.3 Impact of braking distance of the train on the performance of object recognition

The trains have a braking distance before which they have to be alerted avoid collision with obstacles. The braking distance of a train is a distance from the point its brakes are applied to the point it comes to a stop. The braking distance 'D' of the train [4] is given as:

$$D = -u^2/2(a - g \times \tan\theta) \quad \text{and, } \tan\theta = (h_2 - h_1)/D \tag{15}$$

Where, 'u' is the speed of train at the point when brakes are applied, 'a' is train's deceleration rate, 'g' is the acceleration due to gravity and (h_2-h_1) the gradient of the track i.e. the difference in height at which deceleration began (h_1) and the its height at the stopping point (h_2). Normally Indian trains decelerate at the rate of 0.5 m/s2 to 1.2 m/s2 which may increase up to 1.5 m/s2 in case of emergency brakes [1]. In plain regions, so the gradient of track (h_2-h_1) is considered zero. The average speed of train is approx. between 80 and 100 Km/h [1]. The speed/time wise analysis along with the calculation of braking distance at speed of 100 Km/h is illustrated in Table 1.

The timely recognition of the obstacles on the track (i.e. before the braking distance) is the main aspect of the performance of the recognition system i.e. $R_D > D$. The recognition distance R_D is the distance between obstacle and train at the time of recognition. The minimum time at which the system should give an alert, must be greater than braking time 't_B' of the train. The performance of the network is improved for the timely recognition of the obstacle by optimizing the deep network parameters.

3 Result and discussion

In this section, the experimental setup for the proposed approach is discussed along with the experimental results illustrated subsequently.

3.1 Experimental setup

In this work, the frames are extracted from thermal video sequence of railway tracks. Total 749 frames of thermal video are utilized to figure out the performance of proposed work. The implementation of Faster R-CNN is performed using system with configuration Intel(R) Core(TM) i5-1035G1CPU @ 1.00 GHz, 1.19 GHz, 16 GB RAM and NVIDIA MX230 GPU.

3.2 Decomposition of for railway track and obstacle on track using 2D SSA

First of all the frames are extracted from the thermal video sequence of railway track as shown in Fig. 2(a) and motion artifacts are compensated. The results of the removal of the motion artifacts are illustrated in Fig. 2(c). Then the each frame is decomposed into several components using 2D-SSA. There are two important parameters which affect the performance of the SSA i.e. window size $m \times n$ and the number of components 'g' to be decomposed from the input frame. The performance of the proposed approach is analyzed for different values of

Table 1 Braking time and braking distance calculation

Speed of train 'u' (Km/h)	Deceleration rate 'a' (m/s^2)	Braking time (Seconds) t_B=(v-u)/a	Braking distance 'D' (meters)
100	0.5	55.54	770
100	0.75	37.03	515
100	1.0	27.77	386
100	1.2	23.15	322

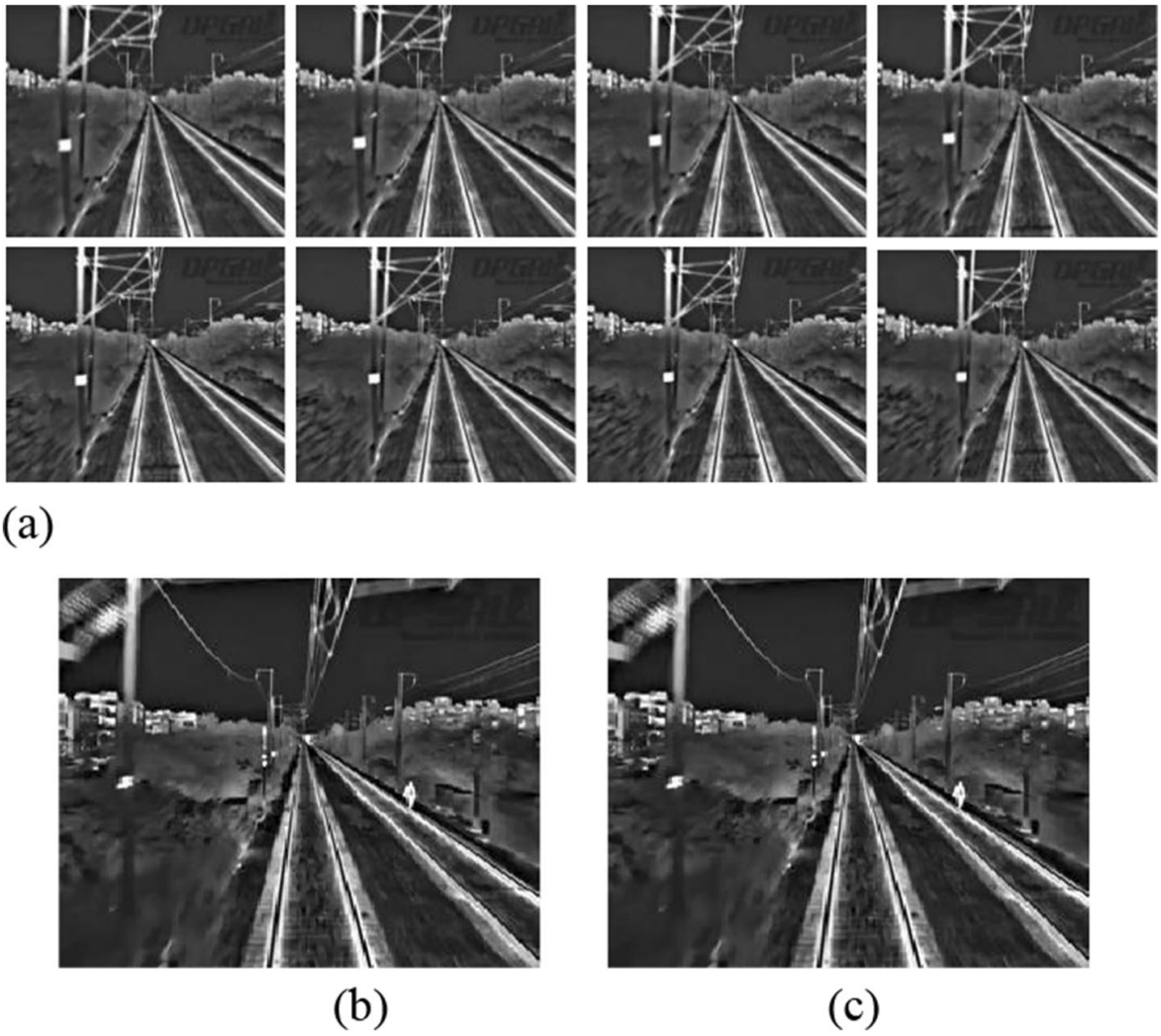


Fig. 2 (**a**) Extracted frames from the thermal video sequence railway tracks. (**b**) Input frame (**c**) Frame after motion artifact removal. (**d**) The distinct decomposed components of the input frame

$m \times n$ and g but the most informative components are decomposed at window size of $m \times n = 8 \times 8$ and number of discriminative components (patterns) $g=15$. Hence, every frame is decomposed into the 15 distinct components are shown in Fig. 2(d). The different components decomposed from input frame highlight the different information of the scene captured by the frame. In the proposed work, the desired information about railway track and obstacles are utilized for further processing.

3.3 Railway track detection

For the detection of the railway track, the every component decomposed from the each frame is analyzed, and then it is observed that the component $S_{Track} = s^{(4)}$ (4th component) of every frame contains strong information about the railway tracks. Therefore, in this work, the 4th component of each frame is used for the recognition of the railway track. Then the features of the railway tracks in the 4th component are separated and utilized for the recognition of the railway track. At last, these features are superimposed on the frame to represent the tracks. The 4th component of the frame, the extracted features from the components and the final recognized tracks are shown in Fig. 3(b), (c) and (d) respectively. Similarly, the railway track is identified in all frames and hence it enables the railway track recognition.

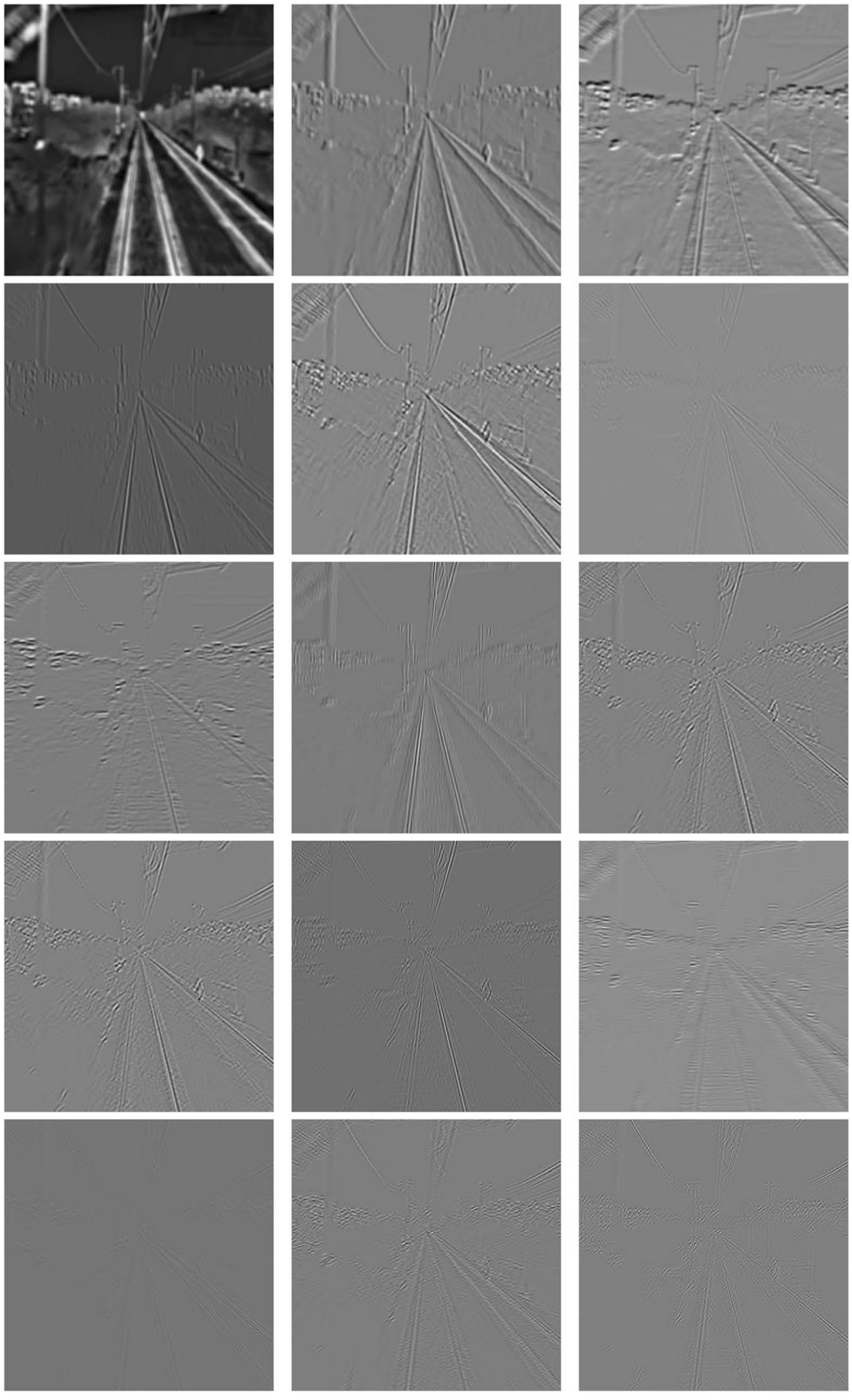


Fig. 2 continued.

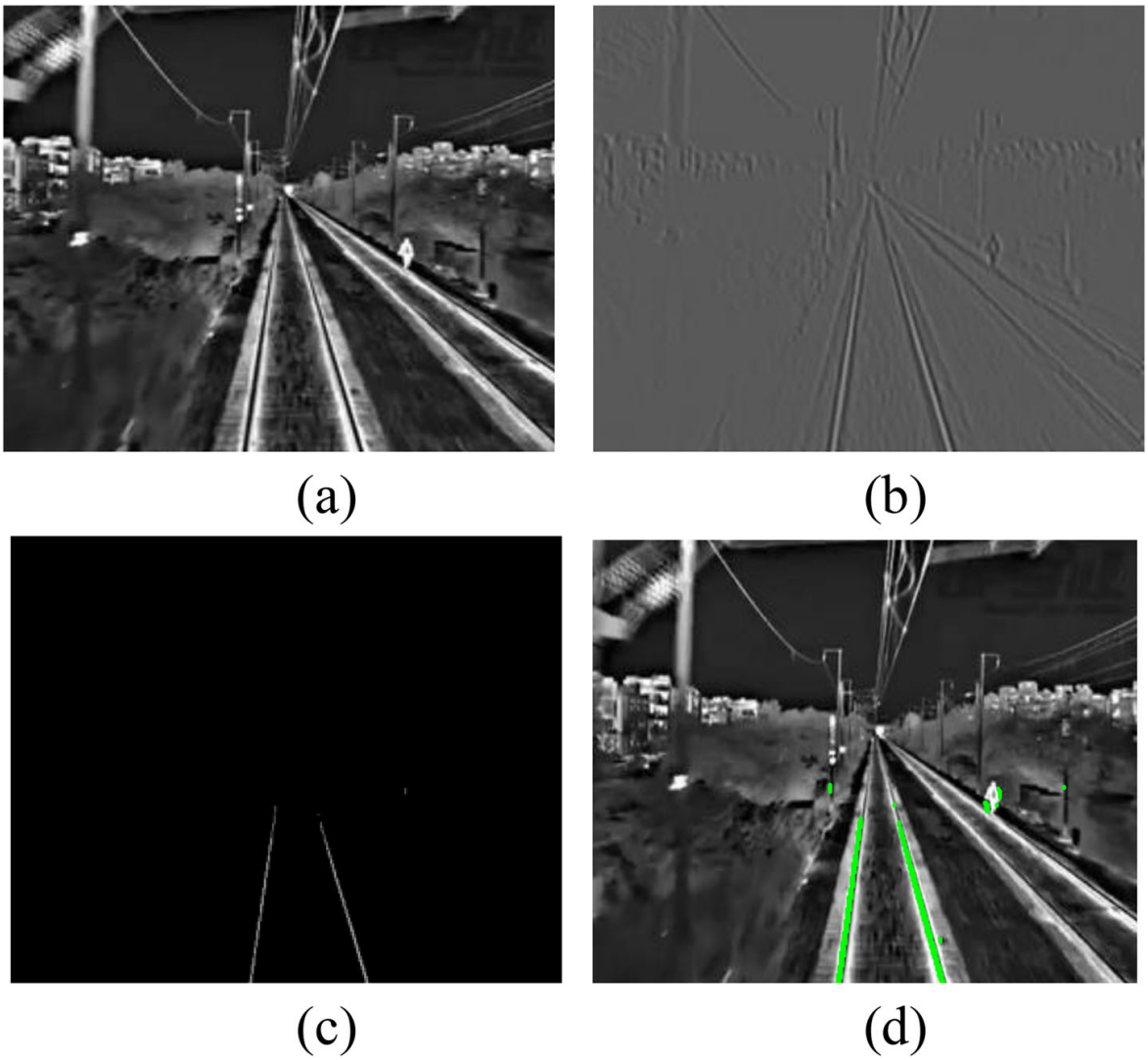


Fig. 3 (**a**) Input Frame (**b**) The 4th component of the input frame containing more dominant information of railway track (**c**) track features of the 4th component (**d**) Recognized Railway Track

3.4 Object recognition

Once the railway track is detected and visualized, the next step is to identify object (obstacles) on the railway track. The decomposed components which have dominant information about the obstacles are utilized as input to the deep network for the recognition of the obstacles on the railway track. It is observed from the analysis of every component of each frame that the component $S_{Obstacle} = s^{(2)}$ & $s^{(3)}$ (2nd & 3rd component) of the frame contains the dominant information about the obstacles on the railway tracks. The 2nd component of the frame containing strong information about object (obstacle) to be recognized is shown in Fig. 4(b).

3.4.1 Training of deep network

In this work, the transfer learning technique is applied by using the pre-trained networks. Additionally, fine-tuning of the parameters and extending the training set with the sample collection are performed to improve the performance as much as possible. Through the transfer learning approach, the training is started with the implementation of the pre-trained parameters to include the useful information gathered from a previously trained network with thermal

railway data used for obstacle detection problem. For Faster R-CNN, 360 × 450 sized image patches were used to train the model. For the RPN stage, the bounding box scales were defined as 3 aspect ratios, which ensured that the network generated anchor boxes for each location of the feature maps. The batch size was defined as 1 to prevent memory allocation errors. For the training, the weight adjusting parameter 'γ' is set to 10 and Stochastic Gradient Descent with momentum value of 0.9. The process continued until 2000 iterations, which took 8 h with the learning rate at 0.001.

For the SSD network, 360 × 450 sized images were utilized for training. The sizes and perspective proportions of the default bounding boxes of each feature map layers remained as before. The RMSProp optimization technique was utilized for gradient estimations with a 0.001 learning rate and 0.9 decay factor for 2,000 iterations. The batch size was characterized as 16 & the training process took 6 h 17 min.

The Yolo-v2 model was trained with the Adam optimizer by a learning rate of 1 × 10−5 with a decay factor of 0.1, with which the validation loss didn't diminish. 9 anchor boxes with various sizes and 3 for each phase of the network. The batch size was characterized as 16 and the entire training was done for 30 epochs. One epoch implies the feed forward and back propagation processes are finished for the entire training dataset. Training of the Yolo-v2 took around 4 h. The training procedure of Yolov5 model consisted of 15 epochs and a learning rate of 0.01. The batch size was defined as 5 which took 2 h 23 min for the thermal image dataset. Moreover, 640 of the 749 images were used for training, and 109 were used for validation.

3.4.2 Recognition of obstacles using deep networks

At last the different deep networks (such as Faster R-CNN, SSD, Yolov2 and Yolov5) are used to recognize obstacles on the railway tracks. When the deep networks are trained with training data, then 2nd components of the 2D SSA output of validation frames is applied to the network which gives recognized obstacles at the output. The recognized obstacle on the railway track in the 2nd component using Faster R-CNN along with railway track is shown in Fig. 4(c). Figure 4(d) shows the final recognized railway track & obstacle on the track using Faster R-CNN. The recognized obstacle on railway track in the 2nd component along with railway track using SSD is shown in Fig. 4(e).

Figure 4(f) shows the final recognized railway track and obstacle using SSD deep network on the frame. Figure 4(g) demonstrates the recognized obstacle on the railway track in the 2nd component along with railway track using YOLOv2. The final recognized railway track and obstacle using YOLOv2 deep network on the frame is shown in the Fig. 4(h). The recognized obstacle on the railway track in the 2nd component using Yolov5 is shown in Fig. 4(i) and the final recognized railway track & obstacle on the track using Yolov5 is shown in Fig. 4(j). The obstacle recognition using Yolov5 is also performed without 2D-SSA as shown in Fig. 4(l) to analyze the effect of 2D-SSA on the recognition performance of Yolov5.

3.5 Performance evaluation

3.5.1 Evaluation metric

The performance of the proposed technique can be assessed with the assistance of various parameters. In this work, the important parameters used to analyze the performance are recognition accuracy [32], precision [32] and recall [32] which are given as:

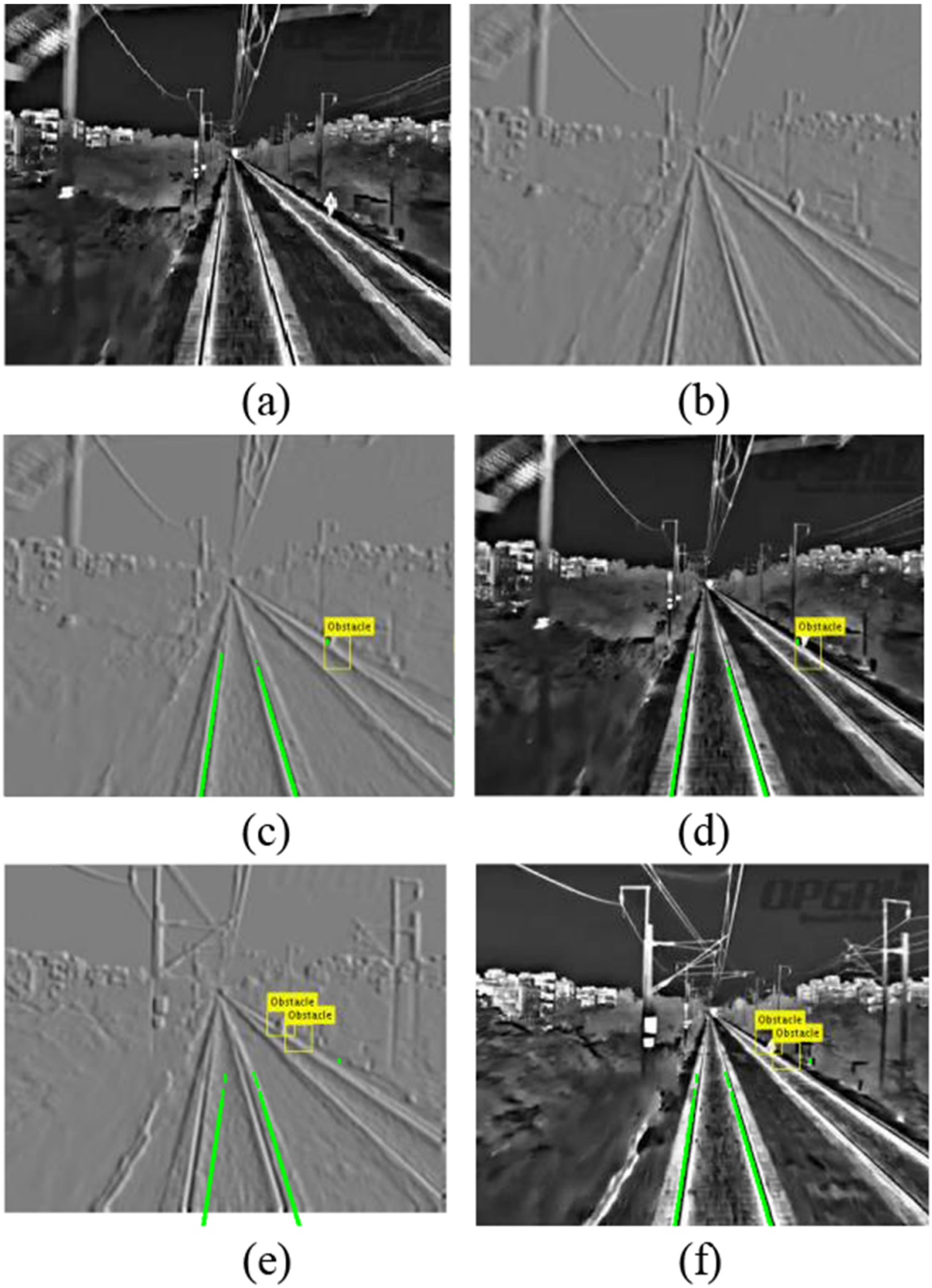


Fig. 4 (**a**) Input Frame (**b**) The 2nd component of the input frame containing more dominant information of railway track (**c**) Recognized obstacle on the railway track using Faster R-CNN (**d**) Final Recognized Railway Track and obstacles in the frame using Faster R-CNN (**e**) Recognized obstacle on the railway track using SSD (**f**) Final Recognized Railway Track and obstacles in the frame using SSD (**g**) Recognized obstacle on the railway track using Yolov2 (**h**) Final Recognized Railway Track and obstacles in the frame using Yolov2 (**i**) Recognized obstacle on the railway track using Yolov5 (**j**) Final Recognized Railway Track and obstacles in the frame using Yolov5 (**k**) Input Frame (**j**) Recognized obstacle in the frame using Yolov5 (without 2D-SSA)

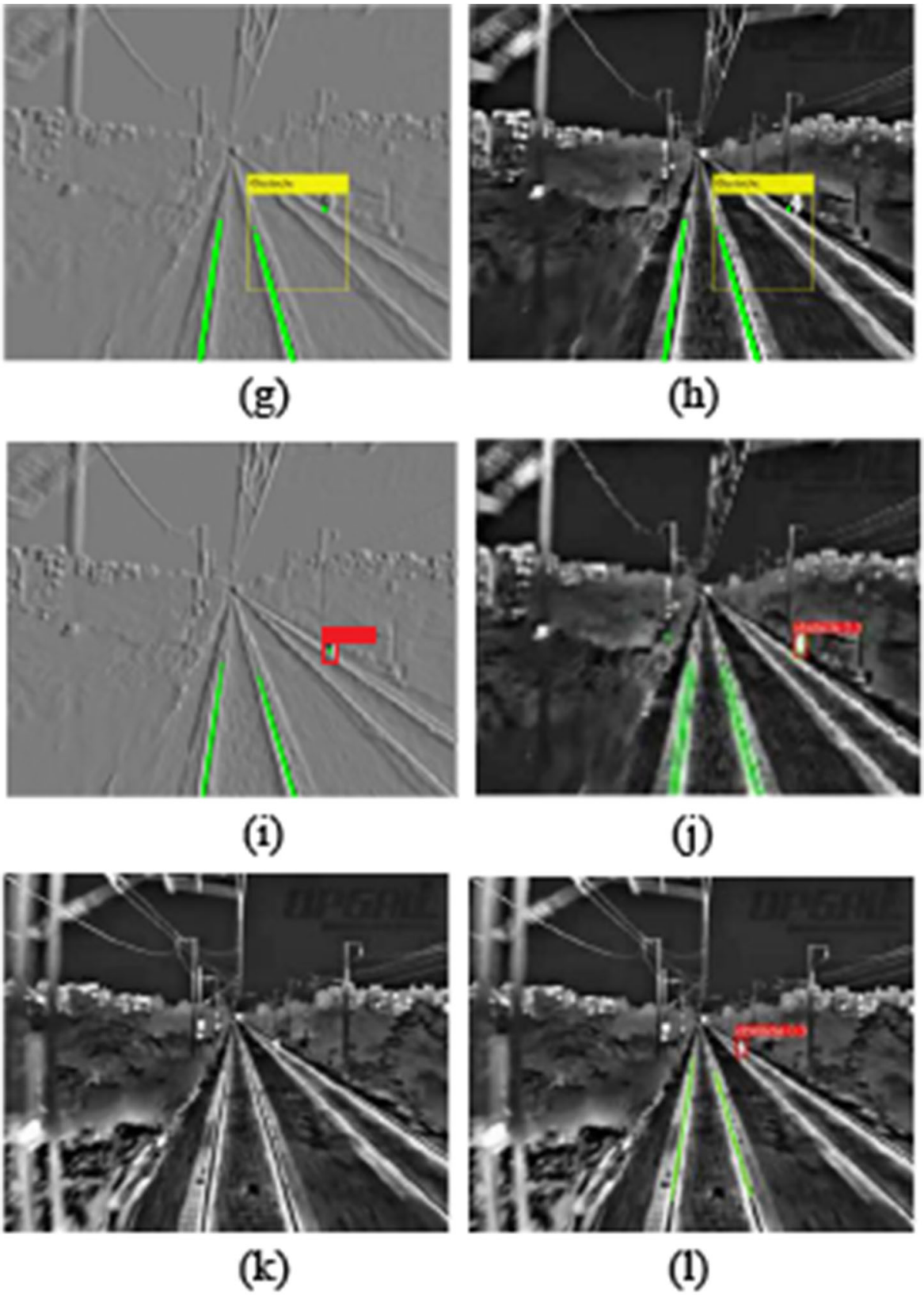


Fig. 4 continued.

$$Accuracy = \frac{(True\ Positive + True\ Negative)}{Total\ Number\ of\ images} \tag{16}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{17}$$

$$Recall = \frac{True\ Positive}{Number\ of\ images\ having\ object\ class} \tag{18}$$

3.5.2 Evaluation at different IoU of faster R-CNN

On the railway track, obstacle ought to be recognized before an adequate distance (i.e. braking distance) with the goal that the speed of the train can be controlled to avoid collision. The performance of the Faster R-CNN is improved for the timely recognition of the obstacle by optimizing the Intersection over Union (IoU) parameter. All in all, the IoU value between 0.7 to 0.9 is considered as adequate range for the precise identification utilizing Faster R-CNN Network. The initial results of the proposed algorithm are acquired by identification network utilizing threshold value 0.7 as appeared in Fig. 5(a) yet the obstacle is recognize at shorter distance from the train. The performance of the system is analyzed for different values of IoU to make a tradeoff between recognition accuracy and distance from train at the time of recognition. To take care of this issue, it is observed that at the IoU threshold value '0.5' the Faster R-CNN not only timely recognizes the obstacle on the track but also attains good accuracy. The final result of railway track identification and obstacle recognition at IoU =0.5 is appeared in Fig. 5(b).

The performance of the purposed system is analyzed for different IoU using above mentioned performance metrics.

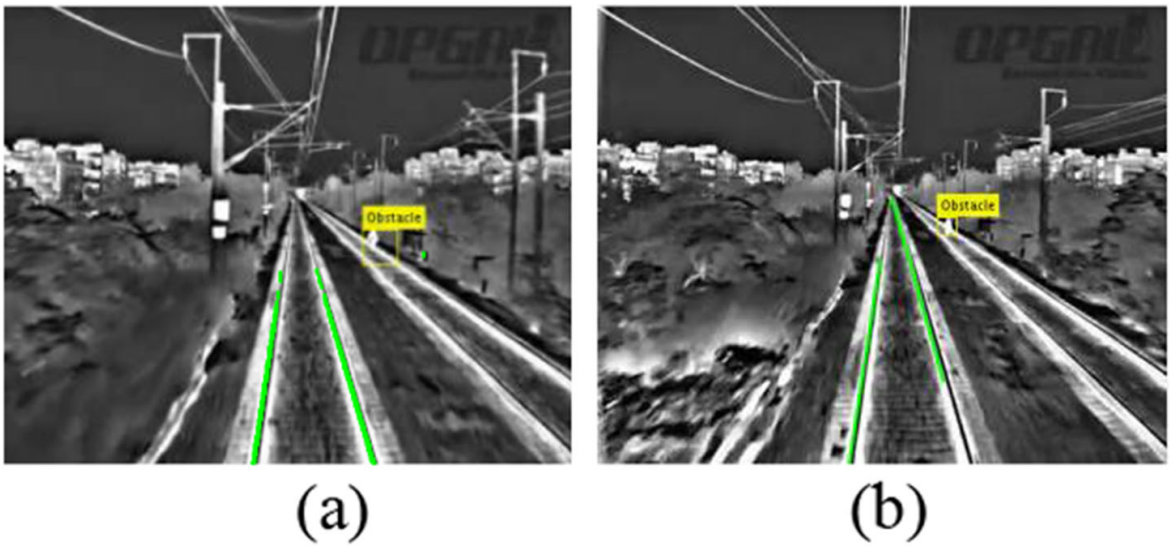


Fig. 5 (**a**) The final results of railway track detection and obstacle recognition with Faster R-CNN at IoU=0.7 (**b**) at IoU=0.5

The parameters evaluated for different values of the Intersection over Union (IoU) as shown in Table 2. The accuracy, precision and recall parameters have better value at IoU '0.5'. Figure 6 illustrates that the graphical representation of the performance of the proposed method. It is concluded from the observations that the purposed method performs better at IoU= 0.5 with accuracy of approximately 85%.

3.5.3 Comparison with state-of-the-art methods

The performance comparison of our proposed method with the other approaches used for object recognition is done for thermal railway track data with these parameters as shown in Table 3. Initially the performance of YOLOv2, SSD, Yolov5 and Faster R-CNN framework are analyzed for recognition of obstacles, and then these approaches are used along with 2D SSA for enhancement of the recognition performance for thermal railway track data. The SSD approach performs better than the YOLOv2 framework but not than others. The Yolov5 and Faster R-CNN framework has shown better performance than both YOLOv2 and SSD. When

Table 2 Performance of the proposed method for different IoUs

IoU	Classification accuracy	Precision	Recall
0.5	0.85	0.845	0.88
0.7	0.775	0.80	0.83

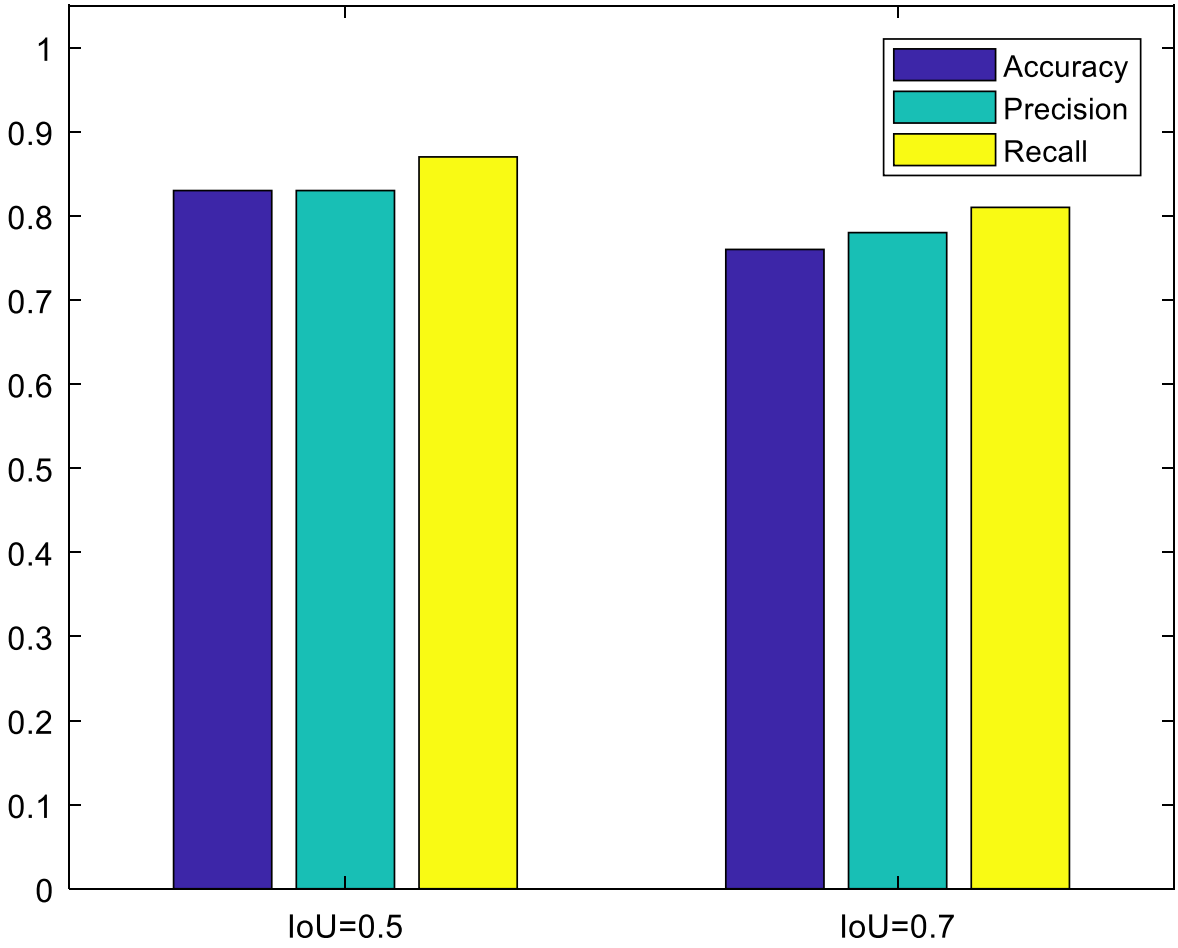


Fig. 6 The classification accuracy, precision and recall for different values of IoU

Table 3 Performance comparison of the proposed method for different state of the art methods

Method	Accuracy	Precision	Recall
YOLOv2 [21]	0.725	0.716	0.743
SSD [21]	0.738	0.727	0.751
YOLOv5	0.832	0.827	0.869
Faster R-CNN [21]	0.834	0.831	0.872
YOLOv2 + 2D SSA	0.741	0.729	0.767
SSD + 2D SSA	0.757	0.742	0.773
YOLOv5 + 2D SSA	0.849	0.843	0.881
Faster R-CNN+ 2D SSA	0.852	0.845	0.886

2D SSA is combined with these frameworks, it will boost the recognition performance. Although the performance of YOLOv2 and SSD is enhanced by utilizing 2D SSA but it unable to outperform the performance of the Faster R-CNN and Yolov5. When the Faster R-CNN and Yolov5 are used with 2D SSA, then the recognition system outperforms all the methods for thermal railway track data as discussed in comparison. The comparative plot is also shown in Fig. 7 for performance analysis of this work with other state of the art methods.

Table 4 illustrates accuracy vs. inference time analysis of the proposed work with other approaches. The performance of Faster R-CNN and Yolov5 is at par in terms of accuracy, precision and recall for thermal railway track data but.

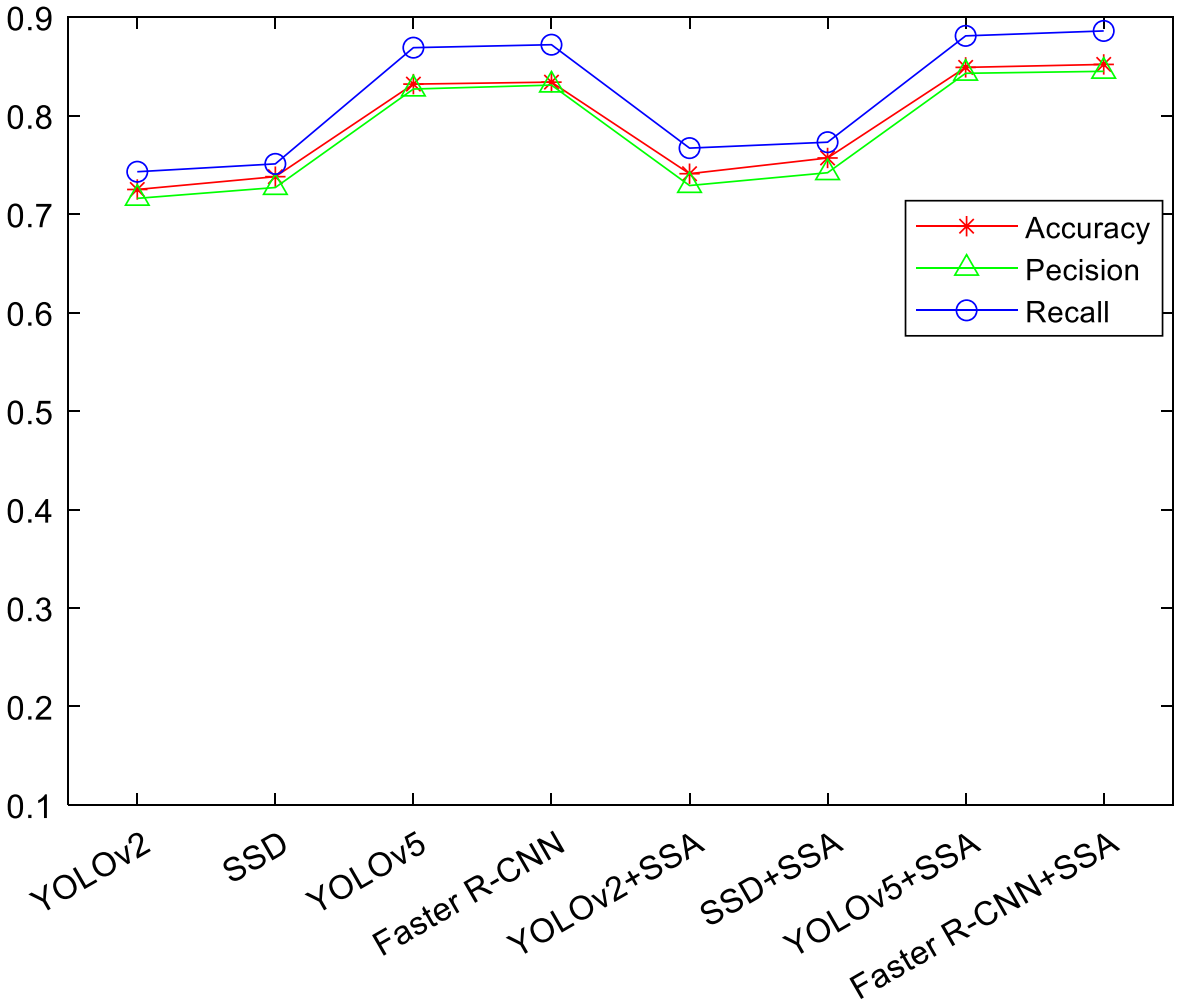


Fig. 7 The comparative plot for performance analysis of proposed work with other state of the art approaches

Table 4 Accuracy vs. inference time for object recognition methods

Method	Accuracy	Inference Time
YOLOv2 [21]	0.725	25 ms
SSD [21]	0.738	63 ms
YOLOv5	0.832	9 ms
Faster R-CNN [21]	0.834	209 ms
YOLOv2 + 2D SSA	0.741	47 ms
SSD + 2D SSA	0.757	85 ms
YOLOv5 + 2D SSA	0.849	29 ms
Faster R-CNN+ 2D SSA	0.852	231 ms

Yolov5 outperforms in terms speed. As discussed in Section 2, if the train is running at the speed of 100 Km/h, it takes 56 s. (approx.) to stop from the point when brakes are applied with deceleration of 0.5 m/s^2 and covers braking distance of 770 m. The proposed system detects the obstacle from the distance of 1Km i.e. the proposed system gives alert approx. 73 s. prior. As discussed in Table 4 the detection time required is 231 ms. The minimum processing time (i.e. sum of detection time and braking time) is 231ms + 56 s = 56.231 s. Hence the proposed system can comfortably avoid the possibility of collision between train and obstacle to enhance railway safety.

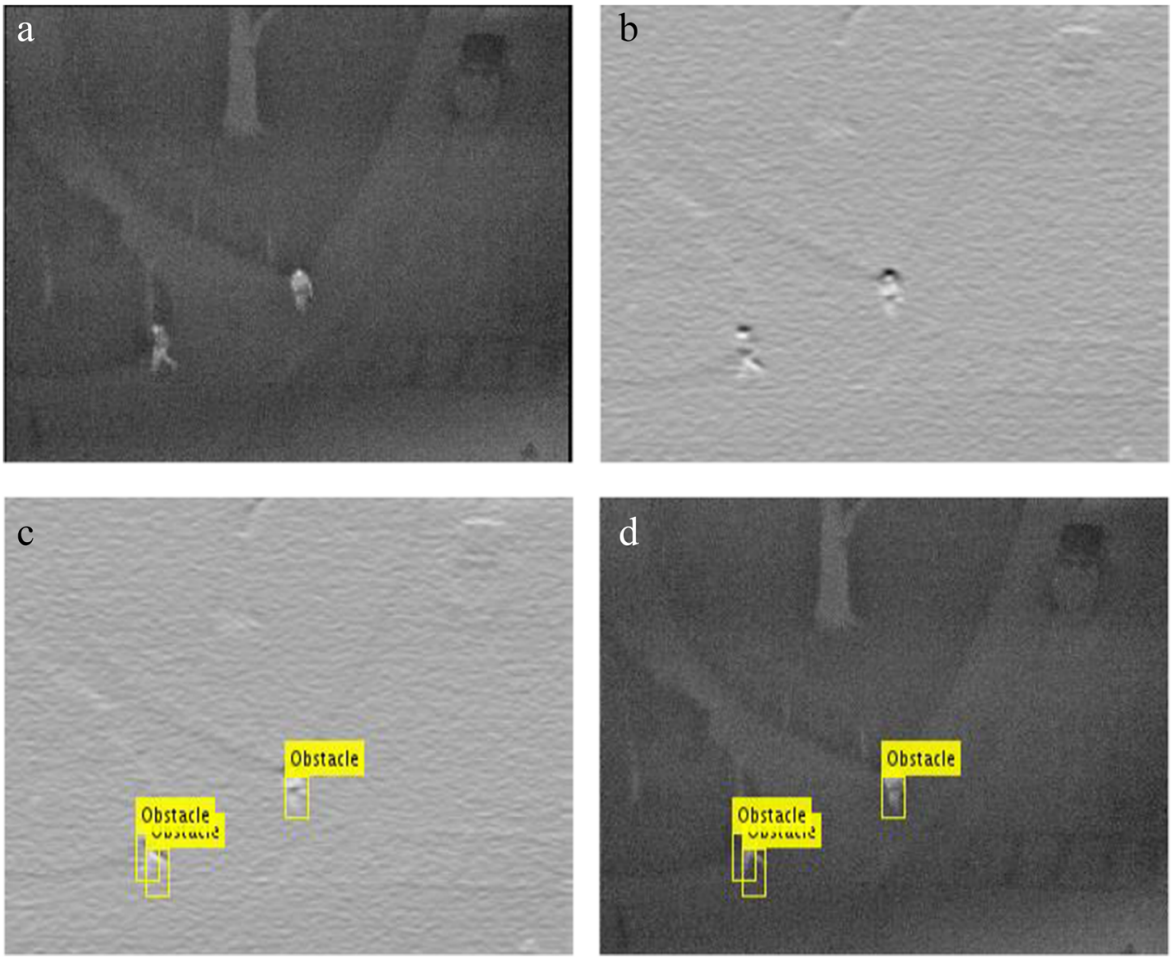


Fig. 8 (**a**) Input thermal image under light rain condition (**b**) The 2nd component of the input image containing more dominant information of pedestrians (**c**) Recognized pedestrians in the 2nd component (**d**) Final Recognized pedestrians in input thermal image

3.6 Performance of proposed work under different illumination conditions

The low illumination due to adverse environmental conditions like rain, Haze, cloudy weather and darkness can cause problems in recognition of obstacles. The proposed work is also tested for different illumination conditions to analyze its illumination invariance. The OSU Thermal Pedestrian Database from OTCBVS Benchmark Dataset Collection [10] is used to test the performance of the purposed work in different illumination conditions. The dataset contains thermal images of different weather conditions like light rain, Haze, cloudy and fair sunlight. The recognition performance is analyzed with the Faster R-

CNN network that was additionally trained on thermal image data for the object class. For the training and testing purposes, the Faster R-CNN detector architecture is used. Figure 8(a) shows the thermal image under light rain conditions in the afternoon with temperature: 68 °F, minimal dew point: 58 °F, humidity: 70% and visibility: 9.0 miles. The 2nd component is decomposed from the image using 2D SSA as shown in Fig. 8(b) because the 2nd component strongly contains the information about pedestrian. Then the 2nd component of the thermal image is applied to the Faster R-CNN for the recognition of pedestrian. Figure 8(c) represents the output of the Faster R-CNN with recognized pedestrian. The final recognized pedestrians in the input image of light rain conditions are shown in Fig. 8(d). The proposed approach performs well in the rainy conditions with thermal imaging. The thermal image captured under Haze with minimal dew point: 47 °F, humidity: 89% and visibility: 2.5 miles is shown in

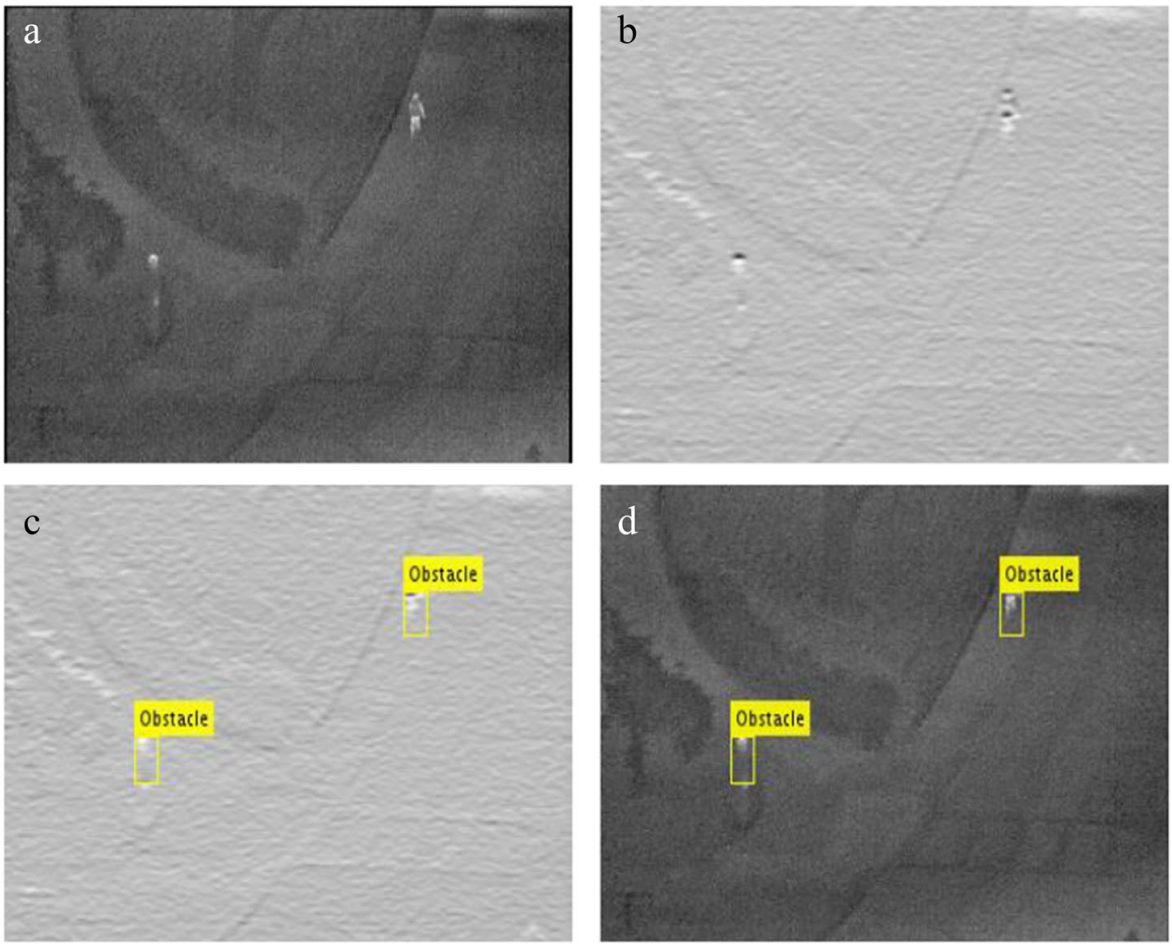


Fig. 9 (**a**) Input thermal image under Haze condition (**b**) The 2nd component of the input image containing more dominant information of pedestrians (**c**) Recognized pedestrians in the 2nd component (**d**) Final Recognized pedestrians in input thermal image

Fig. 9(a). Figure 9(b) shows the 2nd component is decomposed from the image using 2D SSA because the 2nd component strongly contains the information about pedestrian. Figure 9(c) represents the output of the Faster R-CNN with recognized pedestrian. The final recognized pedestrians in the input image of light rain conditions are shown in Fig. 9(d). It is analyzed that the proposed system can work for recognition of objects under Haze conditions. The proposed work is also tested for thermal image having scene captured under dark cloudy conditions. Figure 10(a) shows the thermal image under dense cloudy conditions in the morning with temperature: 53 °F, minimal dew point: 46 °F, humidity: 77% and visibility: 8.0 miles. The 2nd component is decomposed from the image using 2D SSA as shown in Fig. 10(b) because the 2nd component strongly contains the information about pedestrian. Figure 10(c) represents the output of the Faster R-CNN with recognized pedestrian. The final recognized pedestrians in the input image of light rain conditions are shown in Fig. 10(d). This method also shows good recognition results for low illumination in dark clouds. The next illumination condition is considered when weather is clear and sunny. The thermal image captured under fair sunny conditions in the morning with temperature: 71 °F, minimal dew point: 32 °F, humidity: 45% and visibility: 10 miles is shown in Fig. 11(a). Figure 11(b) shows the 2nd component is decomposed from the image using 2D SSA because the 2nd component strongly contains the information about pedestrian. Figure 11(c) represents the output of the Faster R-CNN with recognized pedestrian. The final recognized pedestrians in the input image of light rain conditions are shown in Fig. 11(d). At last, the illumination condition in partly cloudy weather

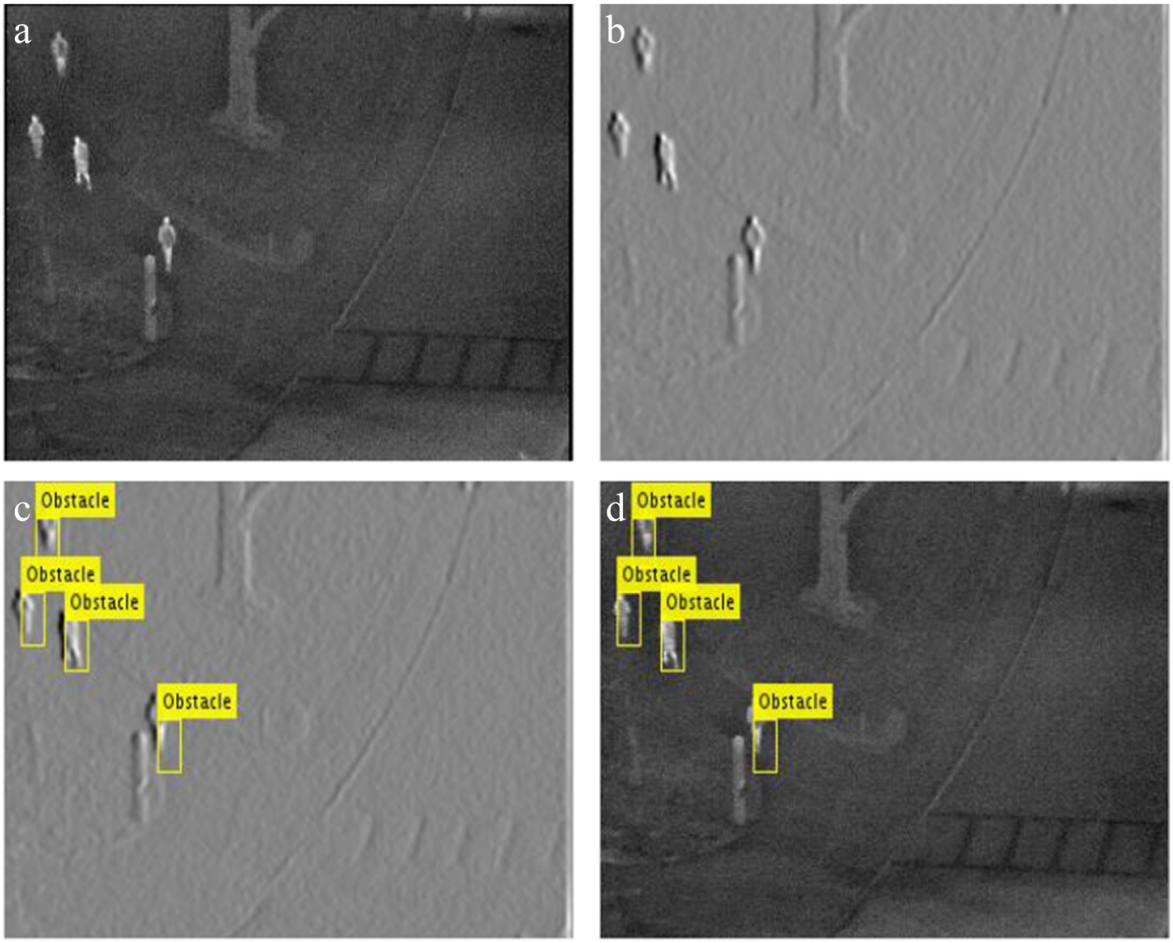


Fig. 10 (**a**) Input thermal image under dense cloudy condition (**b**) The 2nd component of the input image containing more dominant information of pedestrians (**c**) Recognized pedestrians in the 2nd component (**d**) Final Recognized pedestrians in input thermal image

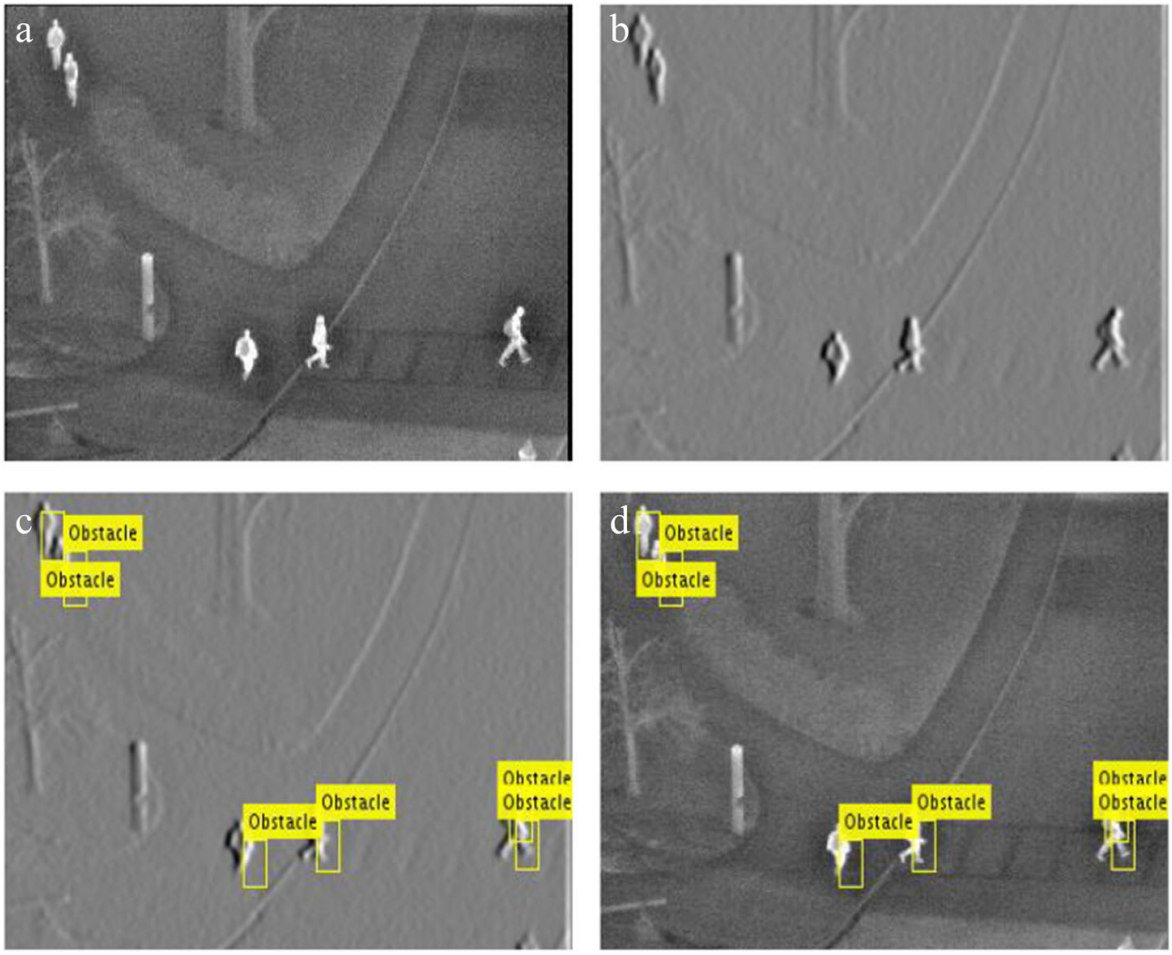


Fig. 11 (**a**) Input thermal image under fair sunny condition (**b**) The 2nd component of the input image containing more dominant information of pedestrians (**c**) Recognized pedestrians in the 2nd component (**d**) Final Recognized pedestrians in input thermal image

is considered. The thermal image captured in partly cloudy in the morning with temperature: 57 °F, minimal dew point: 37 °F, humidity: 47% and visibility: 10 miles is shown in Fig. 12(a). The 2nd component is decomposed from the image using 2D SSA because the 2nd component contains the information of pedestrian as shown in Fig. 12(b).

The output of the Faster R-CNN with recognized pedestrian is represented in Fig. 12(c). Figure 12(d) shows the final recognized pedestrians in the input image of light rain conditions. Hence the performance of the proposed work is analyzed under different illuminations in different weather conditions; it gives good illumination invariance and shows better performance in all illumination conditions.

4 Conclusion and future scope

This paper introduced a novel and efficient method for identification of railway track and obstacle recognition on the railway track. The proposed approach utilizes the 2D SSA to decompose the thermal image into distinct information carrying components, and then the deep learning network used the particular component to identify the obstacle on the railway tracks. Identification of the railway tracks using SSA is another significance of this work. The concatenation of Deep network with 2D SSA introduced a more efficient and robust approach

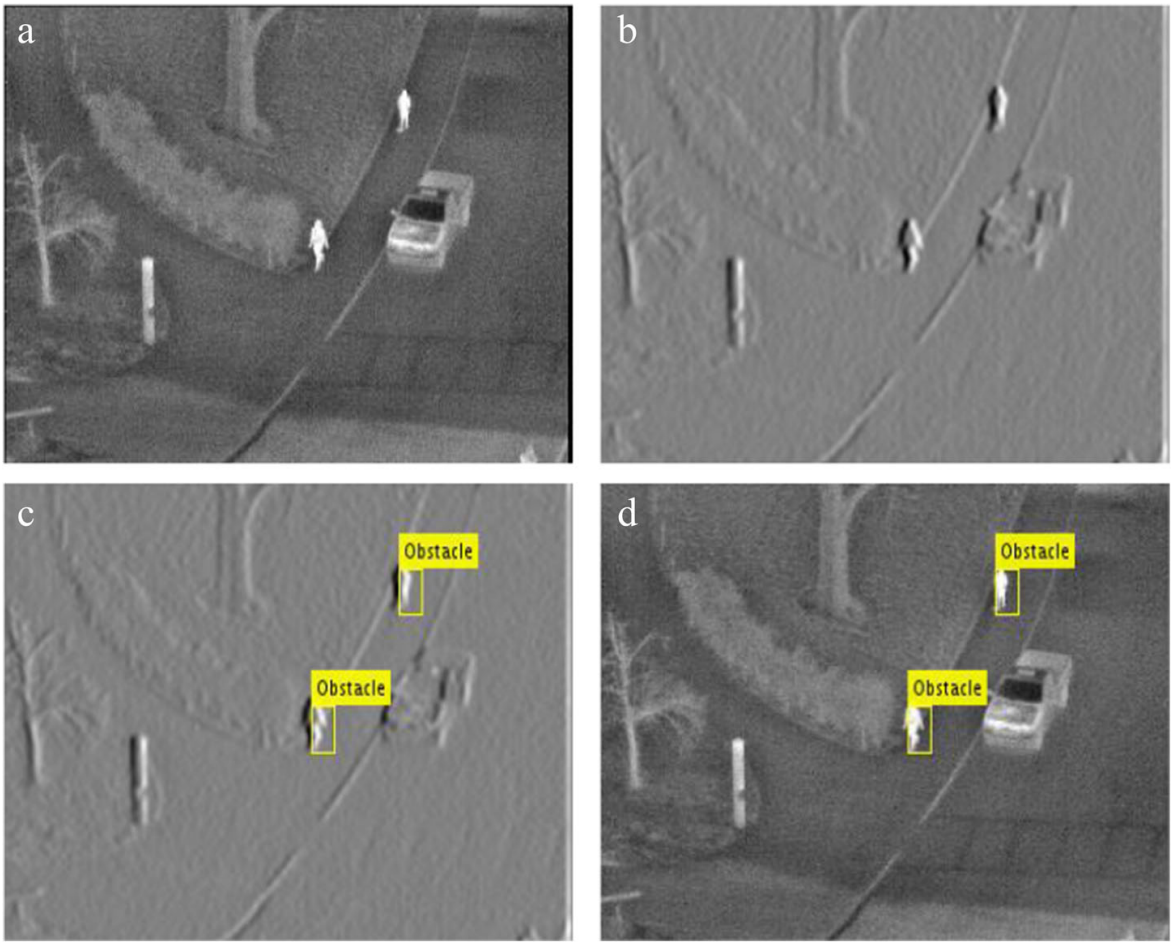


Fig. 12 (**a**) Input thermal image under partly cloudy condition (**b**) The 2nd component of the input image containing more dominant information of pedestrians (**c**) Recognized pedestrians in the 2nd component (**d**) Final Recognized pedestrians in input thermal image

to build a early warning framework to forestall railway accidents to enhance railway safety. As there is no significant modification is needed in the train infrastructure, so this system will be cost-effective. This will also diminish the monetary burden in terms of railway compensation. Additionally, this work is also shows good performance when tested for different illumination conditions using OSU thermal pedestrian database from OTCBVS benchmark dataset collection. This system will be extended for more efficient object recognition under distinct recognition challenges like occlusion, background clutter etc. as a future work to improve the adequacy of the framework.

Acknowledgements Authors are grateful to Delhi Technological University and specially Sh Anand Vardhan, CEO of MM Logic soft P Ltd for sponsoring the project.

References

1. Aakash DB, Ramachandran N, RastogiV (2019) Studies on the effects of braking loads on a Railway Wheel. In 2nd International Conference on Computational & Experimental Methods in Mechanical Engineering, IOP Conference Series: Materials Science and Engineering, 691, pp 3–5

2. Amaral V, Marques F, Lourenco A, Barata J, Santana P (2016) Laser-based obstacle detection at railway level crossings. J Sensors, Hindawi Publishing Corporation, pp 1–11
3. Aragon R, Zhigljavsky LA (2010) Singular spectrum analysis for image processing. Stat Interface 3(3):419–426
4. Barney D, Haley D, Nikandros G (2001) Calculating train braking distance. In Conference: Safety Critical Systems and Software 2001, Sixth Australian Workshop on Safety-Related Programmable Systems, (SCS2001), St Lucia, Queensland, Australia, 3, pp 23-29
5. Berg A, Ofjall K, Ahlberg J, Felsberg M (2015) Detecting rails and obstacles using a train-mounted thermal camera. In Image Analysis, Scandinavian Conference on Image Analysis (SCIA). Springer, Berlin, 9127, pp 492-503
6. Chaudhary PK, Pachori RB (2021) Automatic diagnosis of glaucoma using two-dimensional Fourier-Bessel series expansion based empirical wavelet transform. Biomed Signal Process Control 64:1–17
7. Chaudhary PK, Pachori RB (2021) FBSED based automatic diagnosis of COVID-19 using X-ray and CT images. Comput Biol Med 134:104454
8. Chen Y, Zhang Rh, Shang L (2014) A novel method of object detection from a moving camera based on image matching and frame coupling. PLoS ONE 9(10):e109809. https://doi.org/10.1371/journal.pone.0109809
9. Daubechies I, Lu J, Wu HT (2011) Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. Appl Comput Harmon Anal 30(2):243–261
10. Davis JW, Keck MA (2005) A two-stage template approach to persondetection in thermal imagery. Proc. 7th IEEE Workshops Appl. Comput. Vis. (WACV/MOTION) 1:364–369
11. Demir B, Erturk S (2008) Improved hyperspectral image classification with noise reduction pre-process. In 6th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland
12. Demir B, Erturk S (2010) Empirical mode decomposition of hyperspectral images for support vector machine classification. IEEE Trans Geosci Remote Sens 48(11):4071–4084
13. Garcia ME, Labayen M, Zamalloa M, Arexolaleiba NA (2020) Application of computer vision and deep learning in the railway domain for autonomous train stop operation. In IEEE/SICE International Symposium on System Integration (SII), pp 1-1
14. Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In IEEE Conference on Computer Vision and Pattern Recognition, pp 580–587
15. GirshickR (2015) Fast R-CNN. In Proceedings of IEEE International Conference on Computer Vision (ICCV), pp 1440-1448
16. Golyandina N (2010) On the choice of parameters in singular spectrum analysis and related subspace-based methods. Stat Interface 3:259–279
17. Golyandina NE, Usevich KD (2009) An algebraic view on finite rank in 2D-SSA. Proceedings of the 6th St. Petersburg Workshop on Simulation 1:308–313
18. Golyandina NE, Usevich KD (2010) 2D-extension of singular spectrum analysis: algorithm and elements of theory. In Matrix Methods: Theory, Algorithms, Applications World Scientific, pp 449–473
19. Golyandina N, Nekrutkin V, Zhigljavsky A (2001) Analysis of time series structure: SSA and related techniques. In: Chapman & Hall /CRC
20. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH (1998) The empirical mode decompositionand the Hilbert spectrum for nonlinear and non-stationary time seriesanalysis. Proc R Soc Lond A Math Phys Eng Sci 454(1971):903–995
21. Kapoor R, Goel R, Sharma A (2020) Deep learning based object and railway track recognition using train mounted thermal imaging system. J Comput Theor Nanosci 17(11):5062–5071
22. Karaduman M (2017) Image processing based obstacle detection with laser measurement in railways. In 10th International Conference on Electrical and Electronics Engineering (ELECO), pp 899-903
23. Kristo M, Ivasickos M, Pobar M (2020) Thermal object detection in difficult weather conditions using YOLO. IEEE Access 8:125459–125476
24. Krummenacher G, Ong CS, Koller S, Kobayashi S, Buhmann JM (2018) Wheel defect detection with machine learning. IEEE Trans Intell Transp Syst 19(4):1176–1187
25. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) SSD: single shot multibox detector. In IEEE European Conference on Computer Vision (ECCV)
26. Madhavan S, Tripathy RK, Pachori RB (2020) Time-frequency domain deep convolutional neural network for the classification of focal and non-focal EEG signals. IEEE Sens J 20(6):3078–3086
27. Malta A, Mendes M, Farinha T (2021) Augmented reality maintenance assistant using YOLOv5. Appl Sci 11(11):4758. https://doi.org/10.3390/app11114758
28. Mangale S, Tambe R, Khambete M (2017) Object detection and tracking in thermal video using Directed Acyclic Graph (DAG). ICTACT J Image Video Process 08(01):1566–1574

29. Manikandan R, Balasubramanian M, Palanivel S (2017) Vision based obstacle detection on railway track. Int J Pure Appl Math 116(24):567–576
30. Manolakis DG, Ingle VK, Kogon SM (2005) Statistical and adaptive signal processing: spectral estimation. Signal Modeling, Adaptive Filtering and Array Processing. Artech House
31. Mittal S, Rao D (2017) Vision based railway track monitoring using deep learning. Comput Sci. ArXiv
32. Murthy CB, Hashmi MF, Bokde ND, Geem ZW (2020) Investigations of object detection in images/videos using various deep learning techniques and embedded platforms—A. Compr Rev Appl Sci 10(9):3280
33. Nayak SR, Nayak DR, Sinha U, Arora V, Pachori RB (2021) Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study. Biomed Signal Process Control 64:1–12
34. Rahul S, Ram Bilas P, Pradip S (2020) Automated emotion recognition based on higher order statistics and deep learning algorithm. Biomed Signal Process Control 58:101867
35. Redmon J, Farhadi A (2016) Yolo9000: Better, faster, stronger. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 6517-6525
36. Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39(6):1137–1149
37. Sabnis OV, Lokeshkumar R (2019) A novel objects detection system for improving safety at unmanned railway crossings. In Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM). IEEE, pp 149-152
38. Shangsheng Z, Jiangzhou Z, Xiaobo C, Yanqiang L (2021) Road information detection method based on deep learning. In 6th International Conference on Electronic Technology and Information Science (ICETIS 2021), Journal of Physics: Conference Series 1827
39. Sinha D, Feroz F (2016) Obstacle detection on railway tracks using vibration sensors and signal filtering using bayesian analysis. IEEE Sens J 16(3):642–649
40. Stewart GW (1973) Introduction to Matrix Computations. Academic, Cambridge
41. Sudha D, Priyadarshini J (2020) An intelligent multiple vehicle detection and tracking using modified vibe algorithm and deep learning algorithm. Soft Comput 24:17417–17429
42. Vazquez J, MazoM, LazaroJL, Luna CA, Urena J, GarciaJJ, Cabello J, Hierrezuelo L (2004) Detection of moving objects in railway using vision. In IEEE Intelligent Vehicles Symposium, pp 872-875
43. Wang Q, Gao J, Yuan Y (2018) Embedding structured contour and location prior in Siamesed fully convolutional networks for road detection. IEEE Trans Intell Transp Syst 19(1):230–241
44. Wang Q, Gao J, Li X (2019) Weakly supervised adversarial domain adaptation for semantic segmentation in urban scenes. IEEE Trans Image Process 28(9):4376–4386
45. Ward JH (1963) Hierarchical grouping to optimize an objective function. JASA 58:236–244
46. Wu Y, He X, Nguyen TQ (2017) Moving object detection with a freely moving camera via background motion subtraction. IEEE Trans Circuits Syst Video Technol 27(2):236–248
47. Zabalza J, Ren J, Zheng J, Han J, Zhao H, Li S, Marshal S (2015) Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging. IEEE Trans Geosci Remote Sens 53(8):4418–4433
48. Zhang M, Hu X, Xu G (2018) An adaptive restoration method for motion-blurred image based on Wiener filtering. In Fifth Conference on Frontiers in Optical Imaging Technology and Applications, 10832
49. Zhou X, Yang C, Yu W (2013) Moving object detection by detecting contiguous outliers in the low-rank representation. IEEE Trans Pattern Anal Mach Intell 35(3):597–610

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

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/358768988

Analysis of Wind on Different Shape of High-Rise Structure

Conference Paper · January 2022

CITATIONS
0

READS
24

3 authors, including:



Ritu Raj Ahirwar
Delhi Technological University
23 PUBLICATIONS **60** CITATIONS
SEE PROFILE



Anbukumar Srinivasan
Delhi Technological University
22 PUBLICATIONS **25** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Wind effects on tall buildings View project

All content following this page was uploaded by Ritu Raj Ahirwar on 24 February 2022.

The user has requested enhancement of the downloaded file.

International e-Conference on Sustainable Development & Recent Trends in Civil Engineering, 4th -5th January 2022, ISBN NO: 978-93-5593-431-4

Analysis of Wind on Different Shape of High-Rise Structure

Rahul Kumar Meena
Department of Civil Engineering
Delhi Technological University
Delhi, INIDA
rahul.08dtu@gmail.com

Ritu Raj
Department of Civil Engineering
Delhi Technological University
Delhi, INIDA
rituraj@dtu.ac.in

S. Anbukumar
Department of Civil Engineering
Delhi Technological University
Delhi, INIDA
sanbukumar@dce.ac.in

***Abstract*— Availability of land is the big issue to avoid such issues different shape of high-rise structure are being constructed all around the world. There for their evolution to resist the wind effects are utmost important because of that numerical simulation is performed on different shape of high-rise structure. This paper presents the effect of wind on the different shape of tall buildings using Computational Fluid Dynamics (CFD), ANSYS CFX version 2020 R2 on a 1:200 length scale. The results show that pressure distribution is not exactly same but more or less same in nature for only windward wall of different building models. ANSYS CFX is being used. Following the simulation, near-field flow patterns around the bluff aerodynamics form are addressed, and aerodynamic mean pressure coefficients are compared to an international standard, which demonstrates good agreement with the results. The plan arrangement and size of the model have been discovered to have a direct impact on the wind pressure distribution on the windward face of all the models considered. On all three models, Face E has a substantial impact on pressure distribution. The pressure distribution on the windward and leeward faces of all the models is comparable.**

Keywords—l-shape, hexgon shape, pressure contours,CFD,Wind effect

I. INTRODUCTION

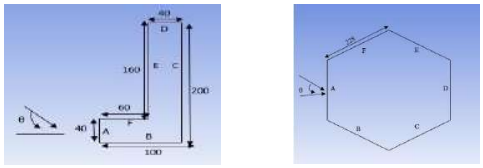
High-rise residential buildings have become more widespread in many cities as a result of increasing population and rising urbanisation, displacing broad regions of vernacular dwellings. These constructions are vulnerable to time-varying loads induced by winds, earthquakes, and other natural calamities, in addition to gravity loads. Tall buildings have a high sensitivity to wind excitation. .[1]–[5]. Chauhan and Ahuja [6] Due to the existence of two additional closely spaced tall buildings with a similar design, the height effects of interfering structures on wind pressure distribution on rectangular plan tall buildings were studied. . The configuration appears to have a significant impact on wind pressure distribution on the major building's faces, as well as the building's height. The total wind force is equal to the sum of windward pressure and leeward suction, but each of these has its own local impact. As a result, when planning for wind loading, the pressure coefficients and force coefficients for other buildings subjected to wind-induced loads, designers consult replated wind load standards. [7], [8], [17]–[20], [9]– [16]. Ahlawat and Ahuja [21] Wind tunnel studies were conducted on models of tall buildings with a "T" plan form to assess wind loads, and it was discovered that the wind flow patterns around the building are substantially influenced by the nearby buildings. Ahlawat and Ahuja [22] Using a wind tunnel test, an experimental investigation was conducted on tall structures with a "Y" plan shape, and it was discovered that the wind incidence angle affects wind pressure distribution. Paul and Dalui [23] Wind effects on cross-plan tall buildings were discussed. . Raj and Ahuja [24] Observed that base shear, base moments, and twisting moments are impacted by wind incidence angle and affected by cross- sectional shape in an experimental investigation on cross- shape tall buildings with varied cross-sectional forms. . Vafaeihosseini et al.[25] The most economical shape for tall structures is rectangle plan shape, according to researchers who employed computational fluid dynamics to investigate wind effects on high-rise buildings. Meena et al. [26] discussed the effects of wind on the bracing system used in various types of multi-storeyed steel buildings. Pal et al. [27] carried out a comparative study on square and fish plan shaped tall building of wind-induced mutual interference effect, maximum efficiency in terms of wind-induced mutual interference pressure, and base shear ix exhibited by square plan shaped model at full blockage condition. Verma et al. [28] analyzed the effects of wind load on tall building octagonal shape, using CFD simulation for 0^0,15^0 and 30^0 wind incidence angle, found that CFD can be used to predict wind-related phenomena on building and other types of structure. Nagar et al. [29] carried out an experimental study of wind-induced pressure on tall buildings of different shapes

from 0^0 to 90^0 wind incidence angle at an interval of 30^0, found that wind pressure decreases up to 60^0 wind incidence angle, "H" plan shaped building is subjected to a higher pressure than the square Model. Chandan and Kumar [30] discussed the numerical simulation of wind analysis of tall buildings using computational fluid dynamics. CFD can result in nearly wind tunnel experiments. CFD could analyze the complete domain study, better-visualized results, and cheaper than that of wind tunnel test. Bairagi and Dalui [31] discussed the aerodynamic effect on the setback of tall building using CFD simulation for 0^0 to 180^0 of wind incidence angle, high positive pressure developed in the setback of roof compared to the top roof due to high turbulence, maximum spectral density frequency developed at the extreme location of setback roof, where the turbulence is maximized. Mendis et al.[32] conducted wind tunnel testing on tall buildings discussed various parameters related to the design of a tall building. Roy et al. [33] did a critical review of wind load on high rise building with different configurations, wind pressure coefficient is maximum in case of square plan shaped building and minimum in case of circular plan shaped tall building, octagonal plan shaped building is with the sharp windward edge is more effective in reducing wind pressure coefficients than hexagonal plan shape of tall building with sharp wind ward edge. The wind forces acting on a building and the motion that arises are directly influenced by the shape of the building. By carefully designing the construction components and geometry of tall buildings, wind excitation is reduced and costs are reduced. Passive aerodynamics adjustments in the form of building shape are one of the most potent and adaptable construction approaches for considerably minimizing the impacts of wind forces by changing/altering the flow pattern around the buildings.

II. Numerical Analysis by ANSYS CFX

A. Model

For studying the influence of wind on various shape of high-rise structure two shape of building are considered in this study, the model is depicted in fig. 1



(a) (b)

Fig.1 Models with Dimension

B. Boundary Condition

Every problem which are solved using the numerical simulation need to provide the definite boundary condition that is why in this study the boundary condition are taken from Meena et al. [34]. Such a huge dimensions are applied because of that no back formation occurred in the applied domain of virtual wind tunnel. The building is placed on the ground of the virtual wind tunnel and the building model is having the particular dimensions. The side wall and top are considered as fee slip wall while the ground is no slip wall

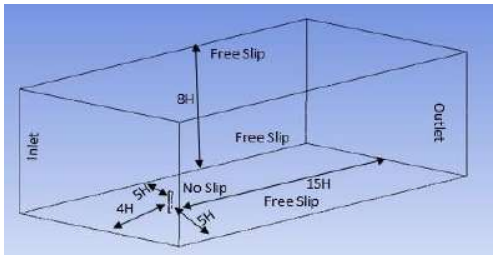
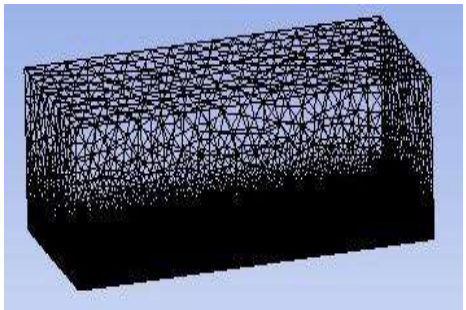


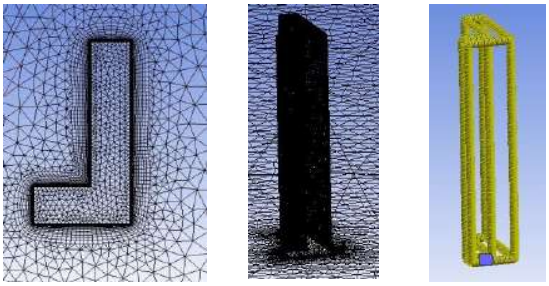
Fig.2 Computational Domain dimensions and boundary condition

C. Meshing

A good mesh may not result in an optimal solution, but a poor mesh will always result in a bad solution. Meshing is provided as shown in Fig.3, and inflation is provided. The effects of air in the separating zone are depicted in Fig.3 (b). Meshing is difficult to achieve since it is dependent on the intricacy of the problem. Because the governing equation cannot be applied to an arbitrary shape, the computer is unable to solve the simulation on the actual geometry shape. In defined volumes, mesh elements allow governing equations to be solved consistently and mathematically. (a) Full domain meshing, (b) building meshing, which is finer than the other meshing, and (d) edge meshing, which is supplied to avoid the anomalous flow. .



(a) Domain Meshing



(b) Inflation (c) Building Meshing (d) Edge meshing

Fig. 3 Meshing

III. Result and Discussion

Result of velocity profile, pressure contours and stream lines for various building shape is presented graphically in various plots in flowing sections.

A. Velocity Profile and Turbulent Profile

The Coriolis force and a pressure difference over a different section of the planet cause wind to blow. The formation of an atmospheric boundary layer near the ground is caused by the roughness of the ground. Due to ground impediments, the wind becomes extremely turbulent. The vertical profile of wind speed is determined by the degree of ground roughness and drag caused by local projections that oppose wind flow. The gradient height is the height at which the drag effects disappear, and the gradient velocity is the velocity at which the drag effects diminish. The height at which topography effects wind speed is known as the atmospheric boundary layer. The wind speed profile within the atmospheric boundary layer, as seen in Fig. 4, is determined by equation according to Power Law (1)

$$\frac{U}{U_H} = \left(\frac{Z}{Z_H}\right)^{\alpha} \quad (1)$$

Where U is the horizontal wind speed at an elevation Z; U_H is the speed at the reference elevation Z_H, which was 10 m/s; is the ground roughness parameter, which is 0.147 for terrain category 2; and Z_H is 1.0 m .

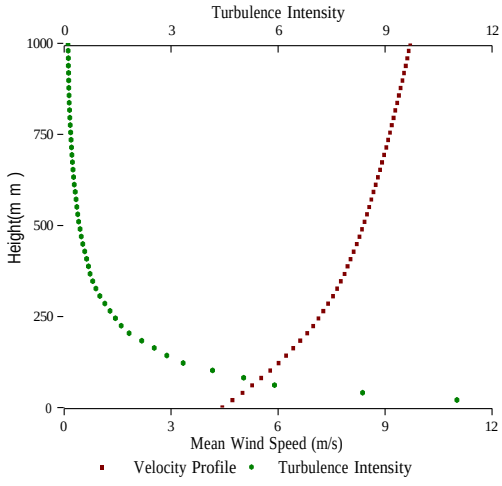


Fig.4 Velocity profile and Turbulent Intensity profile

Because of the friction between the surface and moving air, wind speed at the surface is quite low, as seen in Fig.4. This study's wind profile is identical to the wind profile described in ASCE 7 and earlier research in this area.

B. Pressure Contour

For Model A and B, the pressure on the windward face is positive. It is negative for a leeward face, as seen in Figs. 5 and 6. Figure 5 depicts the graphical representation of pressure on each face of the Model A, with the maximum pressure in the centre and decreasing in the corner region. The rest of model A's faces are under the influence of negative pressure, as shown in Fig.5. Face B's pressure distribution shows that it is lessening in magnitude as we move along the wind direction due to the suction created, while face D and face F have an almost identical pattern of pressure distribution, increasing as we move along the wind direction. Face E depicts how pressure distribution is affected by component size as illustrated in Figs. 5. Model B has the same pressure distribution pattern. Only the difference for face E is shown.

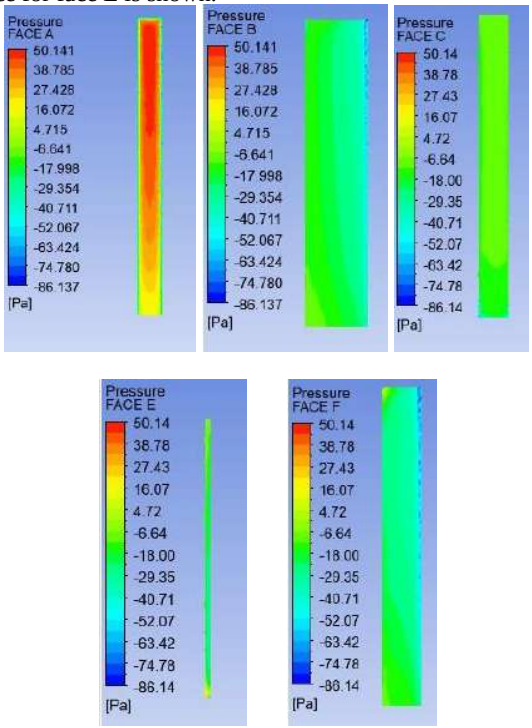


Fig.5 Pressure Contour on Building Model A

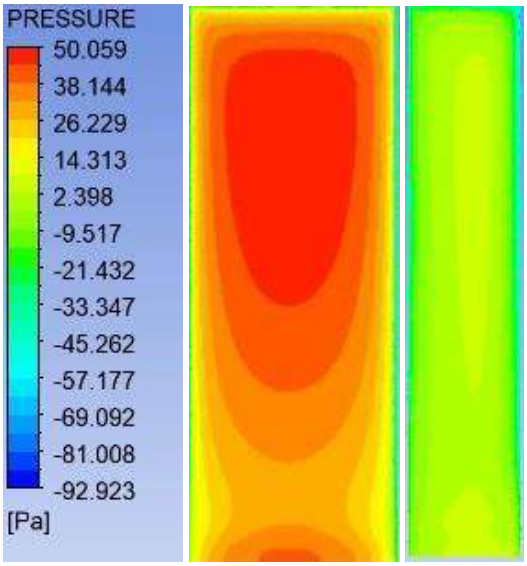


Fig.6 Pressure Contour on Building Model -B

C. Velocity Streamlines

A streamline is an imaginary line in a fluid whose tangent at any point represents the direction of the velocity of a fluid particle at that location. Because various flow patterns are observed for all models, and it is due to the variation in the plan shape of the building, the streamline for 0^0 wind incidence angle for model A and model B in the plan is shown in Fig. 7. In the wake zone, vortices are formed. On the leeward side of the building models, it will cause negative pressure. On the downstream side of building models for model A is having the vast recirculation zone of air can be seen. For a 0^0-wind incidence angle, the streamline is depicted in Fig. 7 in plan and Fig. 8 in elevation for all model buildings. The direction of streamlines is shown by the fish in Figures 7 and 8. And the colour scale indicates the velocity of streamlines, which is higher on the windward side than on the leeward side, and the vortices generated also depend on the side ratio of the building, with a larger recirculation zone on the leeward side of model A as shown in Fig. 7 and a smaller recirculation zone on the leeward side of model B as shown in Fig. 7.

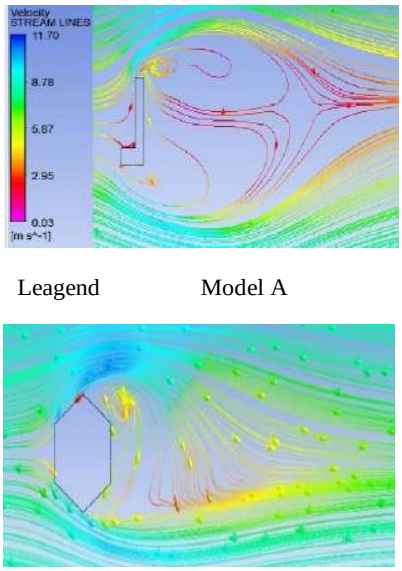
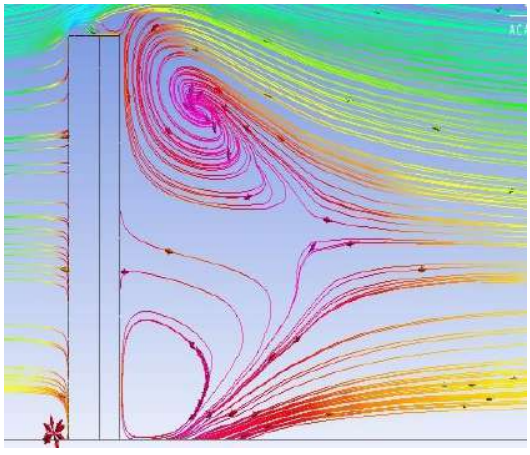
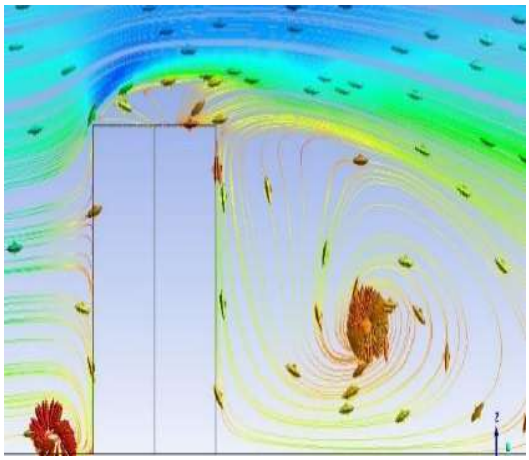


Fig. 7 wind flow patterns around the building model in plan





(B) Model B

Fig.8 Wind flow patterns around the building models in Elevation

The various types of vortex shedding are depicted in Figure 8. CFD simulation is used on tall buildings to minimise the effect of wind, which is dependent on the shape of the building. Low-pressure zones are created when the river flows through the bluff bodies. In the wake region of buildings, vortex shedding occurs. When airflow moves through the

building, it tends to diverge from the building's rear boundary. It spins into a vortex.

D. Pressure Coefficient

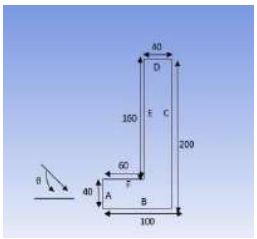
The mean pressure coefficient '$C_{p\ mean}$' is calculated from equation (2)

$$C_{p\ mean} = \frac{p - p_o}{\frac{1}{2}\rho U_H^2} \qquad (2)$$

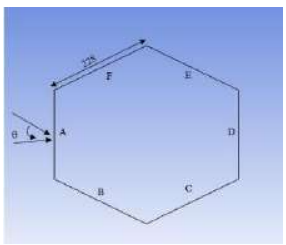
p . is the reference height static pressure, is the air density

o

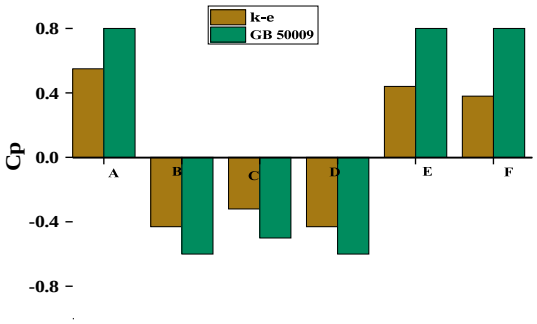
(1.225 kg/m3), and,,U_H^2. is the mean wind velocity at the building reference height.



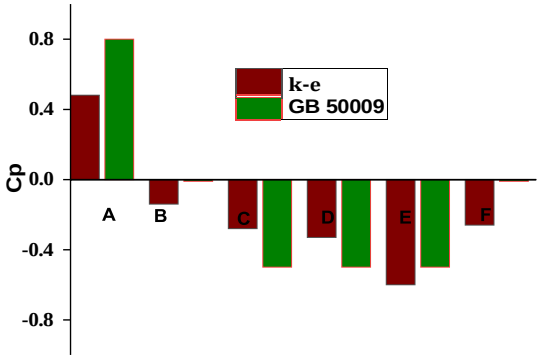
Model-A



Model-B



Pressure Coefficent on Mean Cp at model-A



Pressure Coefficent on Mean Cp at model-B

Fig. 9: Comparison of mean pressure coefficients on a different face of all Model between numerical result and GB 50009 (International Standard)

Figure 9 depicts the average Cp for all of the participants. Model, it can be deduced from Figure 9 that face A, E, and F are the only faces that are under the impact of positive pressure. The above figure illustrates that mean Cp values show a good agreement between numerical results and international standards, and model A, have nearly the same nature of pressure distribution in terms of positive and

negative pressure coefficients. While for model-B it is clearly indicated that postive nature of pressure distribution is observed on the face-A while rest of the other faces is having the negative pressure distribution.

IV. Conclusions

In this work, the pressure contour, mean pressure coefficients, and velocity streamlines for the L-shape and hexagon building model are compared at 0^0 wind incidence angles. This research is simulated using the k-ε Model. The following are the primary findings of this study

- The goal of this study is to determine the pressure distribution on all sides of the building, and the pressure contour on all sides is reported in this paper.
- The suction on the leeward and side faces of the building models is significantly affected by changes in the side ratio of the building.
- Model A has a maximum pressure coefficient of 0.7 on windward face A and a minimum pressure coefficient of 0.65 on side face F.
- For both models, Face A always has positive pressure, whilst the other face always has negative pressure.
- The side ratio of building models affects the pressure distribution on Face E, with maximum pressure distribution for model C and minimum pressure distribution for model A.
- Faces E and F are the two sides of the same coin. For all three models, the pressure distribution on the windward face is nearly identical.
- The velocity streamlines are depicted in the plan and elevation using the figure.
- As with a boundary layer wind tunnel, the precision of the results is determined by meshing the geometry model and determining the flow physics.
- The development of vorticity, which displays high turbulence, can be seen in the wind pressure distribution on the leeward face of this study.

Acknowledgment

The authors would like to express their appreciation to Dr. Manvendra Verma for his precious suggestions and motivation to complete this work on time. Delhi Technological University provided the resources and fund to conduct this study. Our sincere gratitude towards this institute.

References

[1] B. S. Taranath, *Wind and Earthquake Resistant Buildings*. 2004.

[2] G. A. Kopp, *Full-Scale Methods for Examining Wind Effects on Buildings*. .

[3] S. O. Hansen, *Wind Loading Design Codes*. 2013.

[4] M. H. Günel and H. E. Ilgin, *Tall buildings: Structural systems and aerodynamic form*. 2014.

[5] Y. Tamura and A. Kareem, *Advanced structural wind engineering*. 2013.

[6] B. S. Chauhan and A. K. Ahuja, "Height effect of interfering buildings on wind pressure distribution on rectangular plan tall buildings," *9th Asia Pacific Conference on Wind Engineering, APCWE 2017*, no. December, pp. 3–6, 2017.

[7] Hong Kong Building Department, "Code of Practice on Wind Effects in Hong Kong 2019," 2019.

[8] BNBC, "Bangladesh National Building Code (BNBC) 2020," *House Building Research Institute*, no. 2, 2020.

[9] TCVN 2737-1995, "Vietnamese standard- TCVN 2737:1995 Loads and Actions." .

[10] ASCE: 7-10(2013), *Minimum Design Loads for Buildings and Other Structures. Structural Engineering Institute of the American Society of Civil Engineering, Reston*. 2013.

[11] BIS, "National Building Code of India, 2016 Volume 1," *National Building Code of India*, vol. 80, p. 1 v. (various pagings), 2016.

[12] GB 50009-2001, *NATIONAL STANDARD OF THE PEOPLE'S REPUBLIC OF CHINA*. 2002.

[13] IS: 875 (2015), *Indian Standard design loads (other than earthquake) for buildings and structures-code of practice,part 3(wind loads)*. 2015.

[14] AS/NZS:1170.2(2011), *Structural Design Actions - Part 2: Wind actions. Standards Australia/Standards New Zealand, Sydney*. 2011.

[15] Association of Structural Engineers of the Philippines, "National Structural Code of the Philippines 2010," p. 758, 2010.

[16] MNBC, *MYANMAR NATIONAL BUILDING CODE 2020*. International relation and legal Section Department of Building Minis, 2020.

[17] NSCP C101-15, *National Structural Code of the Philippines 2015*. 2015.

[18] ETHIOPIAN STANDARD, *ES ISO 4354 (2012) (English): Wind actions on structures*, vol. 2012. 2012.

[19] TSE498, "Yapı Elemanlarının Boyutlandırılmasında Alınacak Yüklerin Hesap Değerleri," *Türk Standartları Enstitüsü*, no. 112, 1987.

[20] MVCS, *National Building Regulations*. 2018.

[21] R. Ahlawat and A. Ahuja, "Wind Loads on T Shape Tall Buildings," *Journal of Academia and Industrial Research*, vol. 24, no. 1, p. 257922, 2015.

[22] R. Ahlawat and A. K. Ahuja, “Wind loads on Y plan shape tall building,” *International Journal of Engineering and Applied Sciences*, vol. 2, no. 4, p. 257946, 2015.

[23] R. Paul, “Wind Effects on Cross Plan Shaped Tall Building Wind Effects on Cross Plan Shaped Tall Building,” no. December 2019, pp. 1–8, 2020.

[24] R. Raj and A. K. Ahuja, “Wind Loads on Cross Shape Tall Buildings,” *Journal of Academia and Industrial Research (JAIR)*, vol. 2, no. 2, pp. 111–113, 2013.

[25] P. K. R. in Ehsan Vafaeihosseini, Azadeh Sagheb, “Computational Fluid Dynamics Approach for Wind Analysis of Highrise Buildings,” no. January, 2013.

[26] R. K. Meena, G. P. Awadhiya, A. P. Paswan, and H. K. Jayant, “Effects of Bracing System on Multistoryed Steel Building,” *IOP Conference Series: Materials Science and Engineering*, vol. 1128, no. 1, p. 012017, Apr. 2021, doi: 10.1088/1757-899X/1128/1/012017.

[27] S. Pal, R. Raj, and S. Anbukumar, “Comparative study of wind induced mutual interference effects on square and fish-plan shape tall buildings,” *Sādhanā*, vol. 0123456789, 2021, doi: 10.1007/s12046-021-01592-6.

[28] D. S. K. Verma, A. . Roy, S. Lather, and M. Sood, “CFD Simulation for Wind Load on Octagonal Tall Buildings,” *International Journal of Engineering Trends and Technology*, vol. 24, no. 4, pp. 211–216, 2015, doi: 10.14445/22315381/ijett-v24p239.

[29] S. K. Nagar, R. Raj, and N. Dev, “Experimental study of wind - induced pressures on tall buildings of different shapes,” *Wind and Structures, An International Journal*, vol. 5, pp. 441–453, 2020.

[30] D. Hima Chandan and R. Pradeep Kumar, “Numerical Simulation of Wind Analysis of Tall Buildings Computational Fluid Dynamics Approach,” no. January 2011, 2014.

[31] A. K. Bairagi and S. K. Dalui, “Aerodynamic effects on setback tall building using CFD simulation,” *International Journal of Mechanical and Production Engineering Research and Development*, no. June, pp. 413–420, 2018, [Online]. Available: www.tjprc.org.

[32] P. Mendis, T. Ngo, N. Haritos, A. Hira, and J. Cheung, “Wind Loading on Tall Buildings,” *Electronic Journal of Structural Engineering*, 2007, [Online]. Available: https://opus.lib.uts.edu.au/handle/10453/5822.

[33] A. K. Roy, A. Sharma, B. Mohanty, and J. Singh, “Wind Load on High Rise Buildings with Different Configurations : A Critical Review,” *lnternational Conference on Emerging Trends in Engineering lnnovations & Technology Management*, vol. 02, no. January 2018, pp. 372–379, 2017, [Online]. Available: https://www.researchgate.net/publication/322725885_Wind_Load_on_High_Rise_Buildings_with_Different Configurations_A_Critical_Review.

[34] R. K. Meena, R. Raj, and S. Anbukumar, *Numerical Investigation of Wind Load on Side Ratio of High-Rise Buildings Numerical Investigation of Wind Load on Side Ratio of High-Rise Buildings*, no. December. Springer Singapore, 2021.

View publication stats

molecules

MDPI

Article

Biophysical, Biochemical, and Molecular Docking Investigations of Anti-Glycating, Antioxidant, and Protein Structural Stability Potential of Garlic

Mohd W. A. Khan [1,*], Ahmed A. Otaibi [1], Abdulmohsen K. D. Alsukaibi [1], Eida M. Alshammari [1], Salma A. Al-Zahrani [1], Subuhi Sherwani [2], Wahid A. Khan [3], Ritika Saha [4], Smita R. Verma [4] and Nessar Ahmed [5]

1 Department of Chemistry, College of Sciences, University of Ha'il, Ha'il 2440, Saudi Arabia; ahmed.alotaibi@uoh.edu.sa (A.A.O.); a.alsukaibi@uoh.edu.sa (A.K.D.A.); eida.alshammari@uoh.edu.sa (E.M.A.); s.alzahrane@uoh.edu.sa (S.A.A.-Z.)
2 Department of Biology, College of Sciences, University of Ha'il, Ha'il 2440, Saudi Arabia; s.sherwani@uoh.edu.sa
3 Department of Clinical Biochemistry, College of Medicine, King Khalid University, Abha 61412, Saudi Arabia; wkhan@kku.edu.sa
4 Department of Biotechnology, Delhi Technological University, Delhi 110042, India; ritikasaha_bt20a16_19@dtu.ac.in (R.S.); smitar@dtu.ac.in (S.R.V.)
5 Department of Life Sciences, Manchester Metropolitan University, Manchester M1 5GD, UK; n.ahmed@mmu.ac.uk
* Correspondence: mw.khan@uoh.edu.sa or wajidkhan11@gmail.com

Abstract: Garlic has been reported to inhibit protein glycation, a process that underlies several disease processes, including chronic complications of diabetes mellitus. Biophysical, biochemical, and molecular docking investigations were conducted to assess anti-glycating, antioxidant, and protein structural protection activities of garlic. Results from spectral (UV and fluorescence) and circular dichroism (CD) analysis helped ascertain protein conformation and secondary structure protection against glycation to a significant extent. Further, garlic showed heat-induced protein denaturation inhibition activity (52.17%). It also inhibited glycation, advanced glycation end products (AGEs) formation as well as lent human serum albumin (HSA) protein structural stability, as revealed by reduction in browning intensity (65.23%), decrease in protein aggregation index (67.77%), and overall reduction in cross amyloid structure formation (33.26%) compared with positive controls (100%). The significant antioxidant nature of garlic was revealed by FRAP assay (58.23%) and DPPH assay (66.18%). Using molecular docking analysis, some of the important garlic metabolites were investigated for their interactions with the HSA molecule. Molecular docking analysis showed quercetin, a phenolic compound present in garlic, appears to be the most promising inhibitor of glucose interaction with the HSA molecule. Our findings show that garlic can prevent oxidative stress and glycation-induced biomolecular damage and that it can potentially be used in the treatment of several health conditions, including diabetes and other inflammatory diseases.

Keywords: garlic; antioxidant; anti-glycation; glycation; AGEs; HSA

check for updates

Citation: Khan, M.W.A.; Otaibi, A.A.; Alsukaibi, A.K.D.; Alshammari, E.M.; Al-Zahrani, S.A.; Sherwani, S.; Khan, W.A.; Saha, R.; Verma, S.R.; Ahmed, N. Biophysical, Biochemical, and Molecular Docking Investigations of Anti-Glycating, Antioxidant, and Protein Structural Stability Potential of Garlic. *Molecules* **2022**, *27*, 1868. https://doi.org/10.3390/molecules27061868

Academic Editors: Tanveer A. Wani, Seema Zargar and Afzal Hussain

Received: 8 February 2022
Accepted: 9 March 2022
Published: 14 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

AGEs are a complex and highly reactive group of heterogeneous compounds produced by glycation of proteins by reducing sugars [1,2]. Protein glycation has been shown to increase the accumulation of AGEs, the production and release of reactive oxygen species (ROS), and the structural and functional alteration of proteins, as well as resulting in tissue damage [3,4]. AGEs interact with particular receptors or bind proteins, activating a number of signaling pathways implicated in diabetic complications such as nephropathy, cataracts, Alzheimer's disease, and atherosclerosis, among others [5,6]. Glycation, AGEs, and oxidative stress have all been linked to a variety of health problems [7]. The states of

total metabolic burden, chronic hyperglycemia, oxidative stress, and inflammation have been linked to an excessive buildup of AGEs [8].

In chronic inflammation, such as that observed in rheumatoid arthritis (RA), AGEs are produced and can accumulate in tissues [9]. Besides, AGEs build with age and promote pathologic stiffening of cartilage and extracellular matrix. Pentosidine is an AGE, and it is found in blood, synovial fluid, and articular cartilage of osteoarthritis patients [9]. AGEs are linked with endothelial activation and endothelial dysfunction [10]. As a result, AGEs have been suggested as an early biomarker for cardiovascular disease [11].

Intracellular ROS inhibit glucose consumption and cause oxidative alteration of intracellular proteins. Such exposure of protein can result in fragmentation, aggregation, oxidative phosphorylation, and unwanted interactions with ion channel-coupled receptors [12]. An excessive generation of free radicals/ROS leads to oxidative stress, which is involved in the development of inflammatory illnesses such as diabetes, cancer, cardiovascular disease, Parkinson's disease, Alzheimer's disease, and aging. Chronic inflammation is involved in tumor development. Tissue damage and endothelial dysfunction arise from increased ROS generation at the site of inflammation, leading to inflammatory diseases [13].

Inhibition of glycation has also been shown to be beneficial in the treatment of diabetic complications [14]. Synthetic chemicals are powerful anti-glycating agents, but they can also have serious side effects such as gastrointestinal problems, uncommon vasculitis, anemia, flu-like symptoms, nausea, and diarrhea. Thus, a lot of interest has been focused on finding natural plant phytochemicals that efficiently prevent glycation and produce fewer adverse side effects. Traditional medicine practitioners frequently employ medicinal plants and natural products in their everyday practice to treat a variety of illnesses because they are generally non-toxic, cheap, ingestible, and have fewer adverse effects. [4]. The medicinal properties of garlic (*Allium sativum*) were well documented in Sanskrit literature 5000 years ago, and its use in Chinese medicine was also reported as far back as 3000 years ago. The healing properties of garlic were well known in the ancient world and utilized by the Egyptians, Babylonians, Greeks, and Romans [15]. Louis Pasteur documented the antibacterial activity of garlic in 1858. Due to its healing potential and antiseptic properties, it was used to treat and prevent gangrene during the Second World War [16]. The overall benefits of garlic in maintaining good health and in preventing a range of health issues has shifted the focus of modern-day medicine back to a time-tested natural remedy. The therapeutic roles of garlic are supported by modern-day epidemiologic evidence, with studies indicating the benefits of garlic preparations in having antioxidant and antimicrobial effects and in reducing diabetes, cardiovascular disease, and cancer [17,18]. This study focused on investigating the anti-glycation and antioxidative stress activities of garlic extract, as well as conducting an in-depth study of the secondary structural alterations of HSA proteins and how garlic extract metabolites might inhibit the glycation reactions in addition to inhibiting secondary structure alterations. This study provides a platform for the future studies in this direction, which may potentially elucidate physiological and immunological imbalances.

2. Results

2.1. Biochemical Analysis

Phytochemical Screening

The chemical and biological properties detected in preliminary screening of aqueous solution of garlic extract are shown in Table 1. The presence of flavonoids was determined using an alkaline reagent test. $FeCl_3$ test was used to estimate phenolic content. The total phenolic compounds in garlic extract were determined to be 21.45 $\pm$ 0.02 mg gallic acid equivalent/g dry weight of the extract. Phenolic content serves as a means of protection against both infection and oxidative stress. Calculation of total flavonoid content of the extract was performed by $AlCl_3$ test, with quercetin as a reference. Thus, total flavonoid content was found to be 16.58 $\pm$ 0.03 mg quercetin equivalents (QE)/g dry weight of the extract.

Table 1. Chemical and biological studies of aqueous garlic extract.

Preliminary Screening	Garlic Extract
Weight of dry powder	50 g
Yield	5.19%
Extract	Aqueous
Flavonoids	+
Polyphenolic compounds	+
Total phenolic compounds	21.45 ± 0.02 mg gallic acid equivalent/g dry weight of extract
Total flavonoid content	16.58 ± 0.03 mg quercetin equivalent/g dry weight of the extract

+ sign indicates presence of flavonoids and polyphenolic compounds.

2.2. Antioxidant and Free Radical Scavenging Activities of Garlic Extract

2.2.1. Assay for Ferric Reducing Antioxidant Power (FRAP)

Using the FRAP test and ascorbic acid as a standard reference, we evaluated the reducing capability of garlic extract. The reduction of Fe^{+3} to Fe^{+2} by the extract is the basis for this test. At 700 nm, the solution of ascorbic acid (0–100 µg/mL) followed Beer's Law with a regression coefficient (R^2) of 0.9973 and a slope (m) of 0.004. For this figure, the intercept was 0.0216. The standard curve's equation is y = 0.004x + 0.0216 (not given). Garlic extract had FRAP values of 32.41 ± 0.86 g ascorbic acid/100 mg dry weight of extract. Garlic extract was found to have improved ferric reducing power in a dosage-dependent manner (Figure 1).

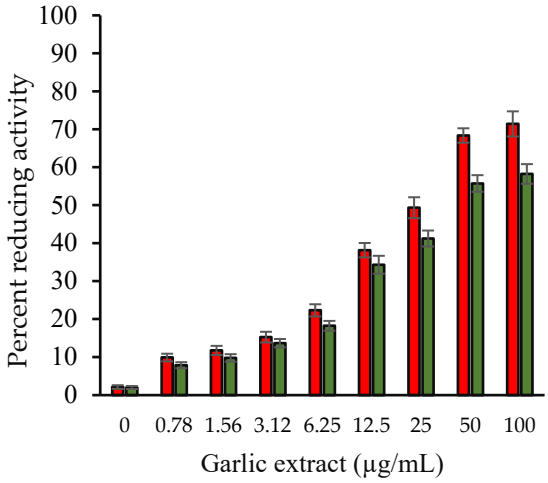


Figure 1. Percentage reducing power of ascorbic acid (red) and aqueous garlic extract (green). Samples in the histogram showed varying concentrations of ascorbic acid and garlic extract (0–100 µg/mL). The *y* axis shows the corresponding percentage reducing power. The results are presented as means ± SEM (n = 3). All the results (0.78–100 µg/mL garlic extract) were statistically significant compared with the sample without extract (0 µg/mL). Comparison between two groups was performed based on t test, and significance was defined as $p < 0.05$.

2.2.2. 2,2-Diphenyl-1-Picrylhydrazyl (DPPH) Radical Scavenging Assay

The DPPH free radical scavenging method is a widely used method for determining antioxidant capacity of various compounds. Figure 2 depicts the DPPH radical scavenging capabilities of aqueous garlic extract. Garlic extract exhibits a substantial DPPH scavenging activity that was observed to rise as extract concentrations were raised from 0–100 µg/mL. The DPPH scavenging activity of 100 µg/mL extract was 66.18%. However, ascorbic acid showed 39.59% at 100 µg/mL.

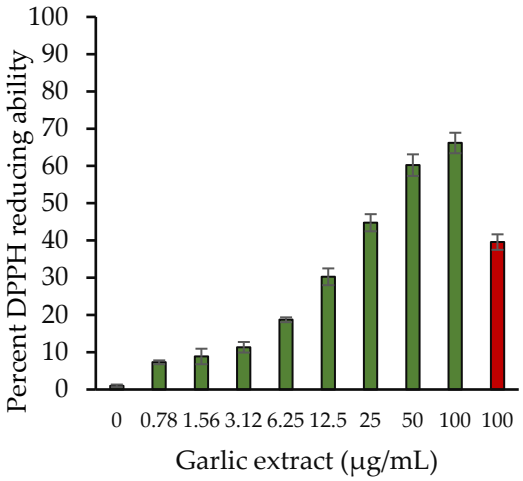


Figure 2. Percentage of free radical reduced vs. aqueous garlic extract concentrations. Various concentrations (0–100 μg/mL) of garlic extracts (green) and 100 μg/mL of ascorbic acid (red). The results are presented as means ± SEM (n = 3). All the results (0.78–100 μg/mL garlic extract) were statistically significant compared with the sample without extract (0 μg/mL). Comparison between two groups was performed based on t test, and significance was defined as $p < 0.05$.

2.3. Inhibition of Structural Changes by Garlic Extract

2.3.1. Protein Denaturation Inhibition

Inhibition in in vitro HSA protein denaturation was examined using garlic extracts. Tissue proteins have been reported to be denatured by inflammatory and oxidative reactions. Natural products were used to evaluate their potential to protect proteins from denaturation and would be employed as an anti-inflammatory dietary source. Garlic extract (50 μg/mL) inhibited heat-induced albumin denaturation at a higher percentage of 50.66% (Figure 3).

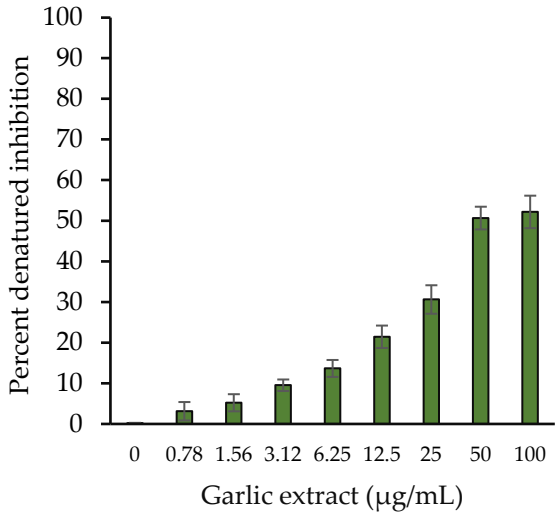


Figure 3. Percentage protection from denaturation induced by heat vs. garlic extract concentration (0–100 μg/mL). The results are presented as means ± SEM (n = 3). All the percentage denaturation inhibition results (0.78–100 μg/mL garlic extract) were statistically significant compared with the sample without extract (0 μg/mL). Comparison between two groups was performed based on t test, and significance was defined as $p < 0.05$.

2.3.2. Protein Browning Inhibition

Browning intensity is used to assess a product's capacity to defend against glycation. The effectiveness of extract to protect against glycation was assessed by measuring the

percentage intensity of browning of HSA treated with glucose. Protein glycation causes an increase in browning intensity under hyperglycemic conditions. At 420 nm, we examined the degree of browning in samples. HSA incubated with glucose without extract had the maximum browning intensity. For this sample, we used a browning intensity of 100 percent. However, when HSA was incubated with glucose in the presence of extract, there was a noticeable reduction in browning intensity (65.23%) at a concentration of 50 µg/mL (Figure 4). In addition, in the case of HSA kept without glucose and extract, some browning (5%) was seen. Internal structural changes that occur over time might be one reason for this sample. Our hypothesis that extracts prevent the development of glycated products is supported by these findings.

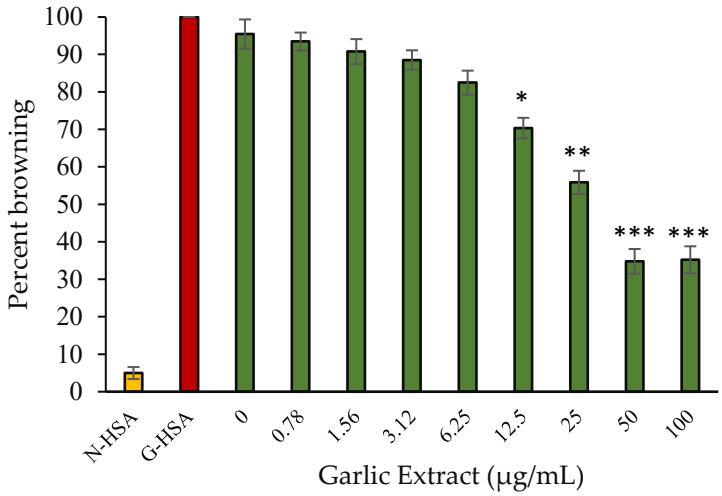


Figure 4. Percentage browning vs. concentration of garlic extract. Native HSA (N-HSA) and glycated HSA (G-HSA) were incubated for 10 weeks under similar conditions and are considered as negative and positive controls, respectively. G-HSA was co-incubated with garlic extract at various concentrations (0–100 µg/mL) of garlic extract (green). The results are presented as means ± SEM (n = 3). All the percentage browning inhibition results (0.78–100 µg/mL garlic extract) were compared with G-HSA sample. Comparison between two groups was performed based on t test, and significance was defined as * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

2.3.3. Inhibition in Protein Aggregate Formation

Glycation causes the binding of carbonyl groups to proteins, resulting in the creation of protein molecules clustering, which is also known as protein aggregation. As a result of glycation, protein aggregates formation occurs. In one of our previous studies, we showed aggregation occurs due to protein glycation [1]. Garlic extract had a beneficial effect on the inhibition of protein aggregation. The addition of extract lowered the aggregation index of G-HSA in a concentration-dependent manner (Figure 5). At a concentration of 100 µg/mL, the extract exhibited the levels of aggregation (67.77%).

2.3.4. Amyloid Structure Inhibition

Congo red (CR) dye is used to determine how much of a protein's secondary structure has been altered. CR contacts hydrophobic clefts located in between beta fibrils and has a unique ability to bind to the protein sheet structure. After binding, the CR dye has a specific absorbance of 530 nm. In Figure 6, the findings of the CR binding experiment showed reduction in cross amyloid structure formation (33.26%) compared with positive controls (G-HSA). By masking the sites of glycation and limiting the surface area accessible by solvent, garlic extract reduced HSA fibrillation and perhaps prevented the transition from α-helix to β-sheet.

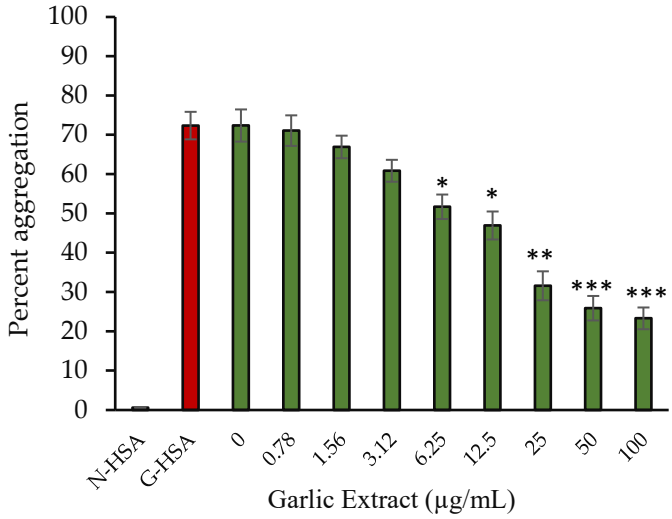


Figure 5. Percentage aggregation vs. concentration of garlic extract. Native HSA and G-HSA were incubated for 10 weeks and served as negative and positive controls, respectively. G-HSA was incubated with various concentrations (0–100 μg/mL) of garlic extract (green). The results are presented as means ± SEM (n = 3). All the percentage inhibition of protein aggregation results (0.78–100 μg/mL garlic extract) were compared with G-HSA sample. Comparison between two groups was performed based on t test, and significance was defined as * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

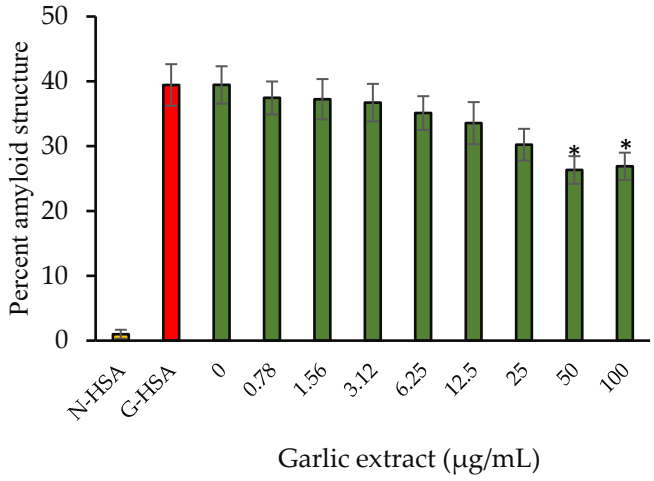


Figure 6. Percentage amyloid structure vs. concentration of garlic extract. Native HSA and G-HSA were incubated for 10 weeks as negative and positive controls, respectively. G-HSA was incubated with various concentrations (0–100 μg/mL) of garlic extract (green). The results are presented as means ± SEM (n = 3). All the percentage amyloid structure inhibition results (0.78–100 μg/mL garlic extract) were compared with G-HSA sample. Comparison between two groups was performed based on t test, and significance was defined as * $p < 0.05$.

2.3.5. Spectral Studies

UV spectral studies were conducted for all the protein samples. HSA incubated with glucose has been reported to have significant decrease in absorption peak at 280 nm as compared with the native HSA, indicating considerable hypochromicity. The observed hypochromicity of glycated sample at 280 nm might be attributable to an alteration in the protein microenvironment and the change in aromatic amino acids. Glycated HSA samples incubated with the extract (50 μg/mL) showed significantly increased UV absorption (Figure 7).

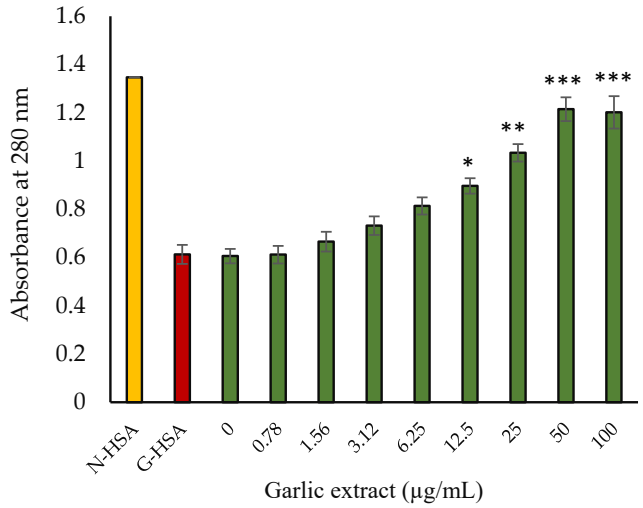


Figure 7. Absorbance of G-HSA vs. concentration of garlic extract. Native HSA and G-HSA were incubated for 10 weeks as negative and positive controls, respectively. G-HSA was incubated with various concentrations (0–100 μg/mL) of garlic extract (green). The results are presented as means ± SEM (n = 3). Change in absorbance results (0.78–100 μg/mL garlic extract) were compared with G-HSA sample. Comparison between two groups was performed based on t test, and significance was defined as * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The formation of AGEs in the glycated samples was detected using AGE pentosidine-specific autofluorescence. For each sample, the specific fluorescence of AGEs was observed to be in the range of 400–480 nm. The fluorescence intensity vs. wavelength (400–480 nm) spectra were found to be rather wide (data not shown), which might be attributable to the variety of fluorescent molecules created during the glycation process. Extract containing samples showed a steady reduction in AGE-specific fluorescence at 450 nm as the extract concentration increased in the incubated samples (Figure 8). The fluorescence intensity of HSA treated with glucose was maximum at 450 nm. Our data show that garlic extract exhibited protection against AGEs synthesis by glycation. Protection level increased as the concentration of garlic extract increased.

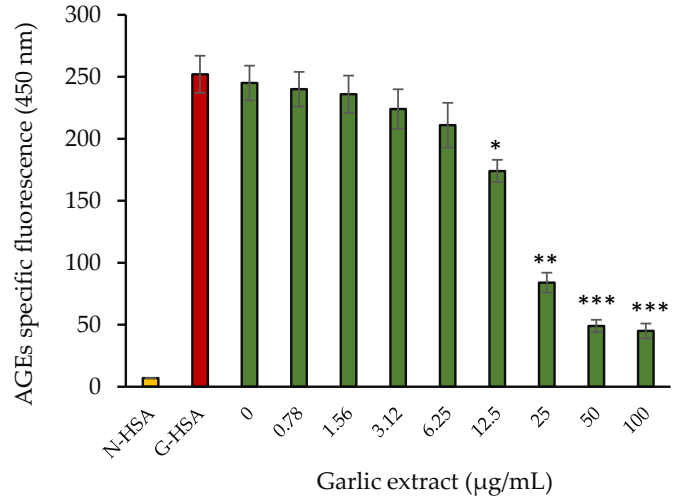


Figure 8. AGE-specific fluorescence intensity at 450 nm vs. concentration of garlic extract. Native HSA and G-HSA were incubated for 10 weeks as negative and positive controls, respectively. G-HSA was incubated with various concentrations (0–100 μg/mL) of garlic extract (green). The results are presented as means ± SEM (n = 3). Reduction in AGE fluorescence results (0.78–100 μg/mL garlic extract) were compared with G-HSA sample. Comparison between two groups was performed based on t test, and significance was defined as * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

2.3.6. Circular Dichroism

Analysis of secondary structure change with glycating agent 'glucose' alone or together with garlic extract with varying concentrations (0–100 µg/mL) was investigated using a Jasco J 810 spectropolarimeter (Table 2). The presence of secondary structural elements was estimated as relative percentages by using the Chen and Yang equation [19] via a computer data processor. CD spectra were recorded for all the samples at a similar wavelength range (200 to 280 nm). As we have shown previously, there is a significant change in α-helix (−10.5%) and β-sheet (+14.9%) structures upon glycation of HSA [1]. Inhibition in the change of both α-helix (−2.8%) and β-sheet (+4.6%) structures of glycated samples were observed when incubated with garlic extract of 100 (µg/mL). Significant decreases in all the secondary structure (α-helix ($p < 0.05$), β-sheet ($p < 0.01$), β-turns ($p < 0.01$), and random coils ($p < 0.05$)) changes in glycated HSA samples were observed at 25 µg/mL concentration of garlic extract. Furthermore, the highest reduction in all the secondary structures was achieved at 100 µg/mL concentration of extract used (Table 2).

Table 2. Secondary structure composition of N-HSA, G-HSA, and G-HSA incubated with different concentrations of garlic extract (0–100 µg/mL).

Conformation	N-HSA	G-HSA	G-HSA with Garlic Extracts (µg/mL)								
	-		0.78	1.56	3.12	6.25	12.5	25	50	100	AG
α-helix	42.7 ± 0.6	38.2 ± 0.3 (−10.5%)	38.2 ± 0.3 (−10.5%)	38.3 ± 0.3 (−10.3%)	38.4 ± 0.3 (−10.0%)	38.8 ± 0.3 (−9.1%)	39.8 ± 0.3 * (−6.8%)	40.5 ± 0.3 * (−5.2%)	40.9 ± 0.3 ** (−4.2%)	41.5 ± 0.3 *** (−2.8%)	40.8 ± 0.3 ** (−4.4%)
β-sheet	26.2 ± 0.5	30.1 ± 0.2 (+14.9%)	30.1 ± 0.2 (+14.9%)	30.0 ± 0.2 (+14.5%)	29.9 ± 0.2 (+14.1%)	29.6 ± 0.2 (+13.0%)	28.8 ± 0.2 * (+9.9%)	28.1 ± 0.2 ** (+7.3%)	27.9 ± 0.2 ** (+6.5%)	27.4 ± 0.2 *** (+4.6%)	27.8 ± 0.2 ** (+6.1%)
β-turn	18.5 ± 0.2	19.4 ± 0.3 (+4.9%)	19.4 ± 0.3 (+4.9%)	19.4 ± 0.3 (+4.9%)	19.4 ± 0.3 (+4.9%)	19.2 ± 0.3 (+4.3%)	19.1 ± 0.3 * (+3.2%)	19.0 ± 0.3 ** (+2.7%)	18.8 ± 0.3 ** (+1.6%)	18.6 ± 0.3 ** (+0.5%)	19.0 ± 0.3 ** (+2.7%)
Random coil	12.6 ± 0.5	12.3 ± 0.4 (−2.4%)	12.3 ± 0.4 (−2.4%)	12.3 ± 0.4 (−2.4%)	12.3 ± 0.4 (−2.4%)	12.3 ± 0.4 (−2.4%)	12.3 ± 0.4 (−2.4%)	12.4 ± 0.4 * (−1.6%)	12.4 ± 0.4 * (−1.6%)	12.5 ± 0.4 *** (−0.8%)	12.4 ± 0.4 * (−1.6%)

The values are in percentage. Each sample was read in triplicate. Data are mean ± standard deviation. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$ vs. control (N-HSA). Values in parentheses represent the percentage change in the secondary structure from N-HSA. Percentage decrease and increase are denoted by "−" and "+" signs. Different GE concentrations were used in µg/mL. The 5 mm of AG was used as negative control. The *t* test was adopted for the comparison between the groups.

2.4. Molecular Docking Studies for Potential Natural Product Metabolites as Inhibitors of Glycation

Some of the important phenolic compounds present in garlic extracts were analyzed for their inhibitory role in the glycation reaction. Sudlow Site I is the primary binding site for glucose in pyranose form. The interaction of the glucose ring form and other ligands (considered in this study) with the SAL subsite of the Sudlow Site I are summarized in Table 1. Figures 9 and 10 provide insight into the most stable 2D and 3D docked conformations of the ligands at the SAL subsite obtained via exhaustive molecular docking application. The orientation of the glucopyranose at the SAL subsite was obtained from the crystal of glucose-bound has submitted by Wang et. al. [20] to the protein data bank (PDB ID: 4iw2). The binding efficiency reported in Table 3 is a result of multiple interactions of the ligands with the target active site.

The blue dotted interaction of Arg257 with glucopyranose (Figure 9a) represents an altered confirmation state of glucopyranose. The 2D interaction images in Figures 9 and 10 represent the type and count of interactions specific ligand undergoes. The 3D images of docked ligands in Figures 9 and 10 represent the polarity distribution at the docking site. The pink environment in the close vicinity of the ligands represents hydrogen donors or an electronegative environment; however, the green cloud near the ligands is due to amino acid with a side chain rich in hydrogen acceptors and results in an electropositive environment. The white cloudy areas around the ligands constitute the neutral space due to the presence of hydrophobic amino acids.

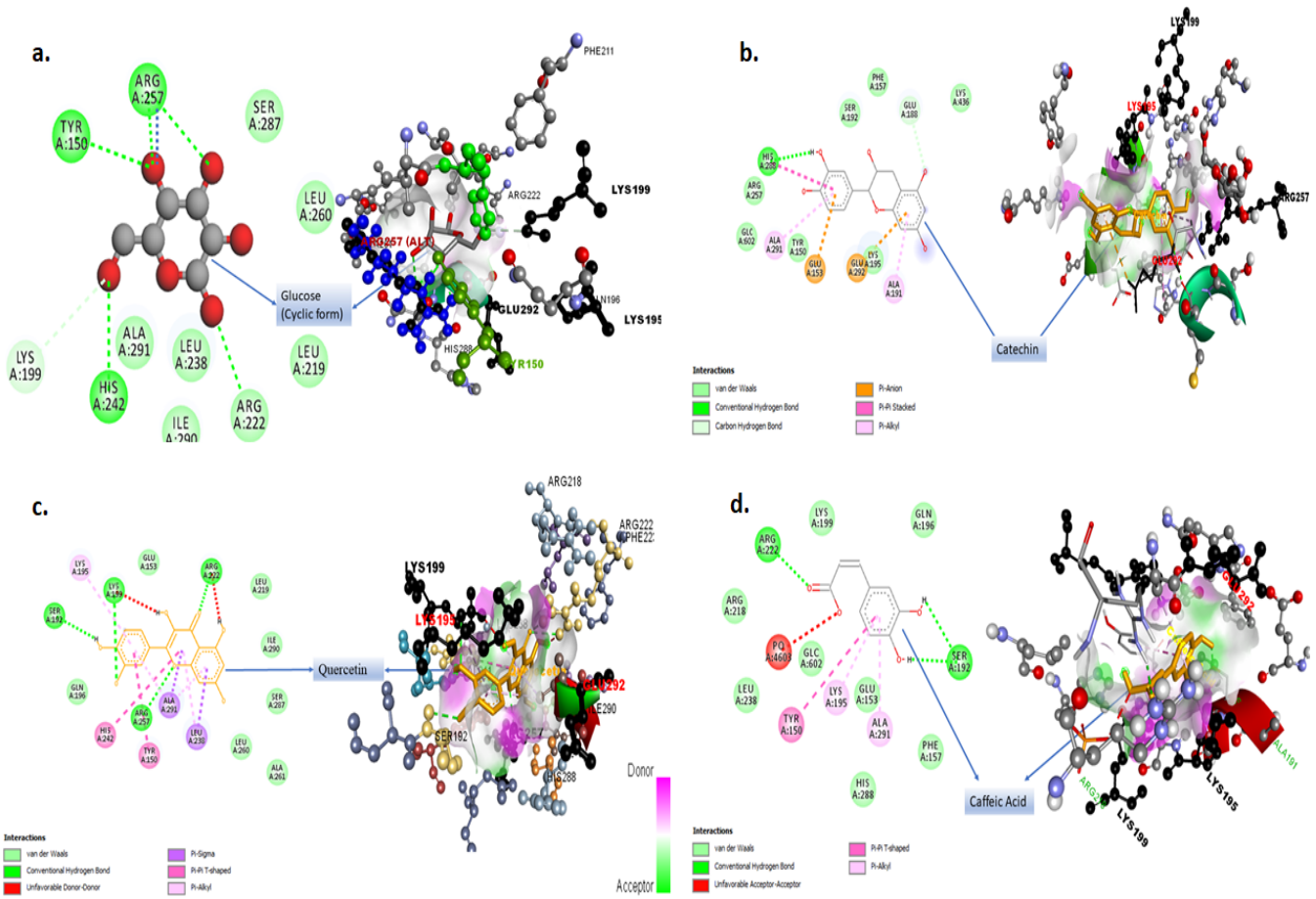


Figure 9. 2D and 3D interaction models of ligands with HSA. Ligands used are (**a**) glucose, (**b**) catechin, (**c**) quercetin, and (**d**) caffeic acid.

Table 3. Ligands. Common name refers to the compound name used in the study.

Ligand/Inhibitor Name (Common Name)	IUPAC Name, Mol Formula, and PubChem ID	Binding Affinity Kcal/Mol	Number of Hydrogen Bonds	Other Interactions *
Glucose (Cyclic form)	(3R,4S,5S,6R)-6-(hydroxymethyl)oxane-2,3,4,5-tetrol Mol formula: $C_6H_{12}O_6$ PubChem ID: 5793	−6.2	6 (3 × Arg257, 1 × Arg221, 1 × Tyr150, 1 × His242)	1 (1 × Lys199) Leu238 and Ala291 (Hydrophobic interactions)
Catechin	(2S,3R)-2-(3,4-dihydroxyphenyl)-3,4-dihydro-2H-chromene-3,5,7-triol Mol formula: $C_{15}H_{14}O_6$ PubChem ID: 73160	−6.7	1 (1 × His 288)	13 (1 × Arg 257, 1 × His 288, 1 × Glc 602, 1 × Ala 291, 1 × Tyr 150, 1 × Glu 153, 1 × Glu 292, 1 × Lys 195, 1 × Ala 191, 1 × Phe 157, 1 × Ser 192, 1 × Lys 436, 1 × Glu 188 Carbon H Bond)

Table 3. *Cont.*

Ligand/Inhibitor Name (Common Name)	IUPAC Name, Mol Formula, and PubChem ID	Binding Affinity Kcal/Mol	Number of Hydrogen Bonds	Other Interactions *
Caffeic Acid	(E)-3-(3,4-dihydroxyphenyl) prop-2-enoic acid Mol formula: $C_9H_8O_4$ PubChem ID:689043	−6.6	3 (1 × Arg 222, 2 × Ser 192)	12 (1 × Arg 218, 1 × Po 4603, 1 × Glc 602, 1 × Leu 238, 1 × Tyr 150, 1 × Lys 195, 1 × Glu 153, 1 × Ala 291, 1 × His 288, 1 × Phe 157,1 × Gln 196, 1 × Lys 199)
Gallic Acid	3,4,5-trihydroxybenzoic acid Mol formula: $C_7H_6O_5$ PubChem ID: 370	−6.4	4 (2 × Lys 199, 1 × Gln 196, 1 × Glu 153)	10 (1 × His 242, 1 × Gln 196, 1 × Ser 192, 1 × Lys 195, 1 × Tyr 150, 1 × Ala 291, 1 × Phe 157, 1 × His 288, 1 × Glu 292, 1 × Arg 257)
m-Coumaric Acid	E)-3-(3-hydroxyphenyl)prop-2-enoic acid Mol formula: $C_9H_8O_3$ PubChem ID: 637541	−6.3	4 (1 × Arg 222, 1 × PO 4603, 1 × Glc 602, 1 × Ser 192)	6 (1 × Arg 222, 1 × PO 4603, 1 × Glc 602, 1 × Ser 192, 1 × Gln 196 Unfavorable donor–donor, 1 × Gln 196 Unfavorable donor–donor)
Quercetin	2-(3,4-dihydroxyphenyl)-3,5,7-trihydroxychromen-4-one Mol formula: $C_{15}H_{10}O_7$ PubChem ID: 5280343	−8.1	4 (1 × Ser 192, 1 × Lys 199, 1 × Arg 257, 1 × Arg 222)	17 (1 × Lys 195, 1 × Lys 199, 1 × Gln 196, 1 × His 242, 1 × Tyr 150, 3 × Ala 291, 2 × Leu 238, 1 × Leu 260, 1 × Ala 261, 1 × Ser 287, 1 × Ile 290, 1 × Leu 219,1 × Arg 222, 1 × Glu 153)
Pyrogallol	benzene-1,2,3-triol Mol formula: $C_6H_6O_3$ PubChem ID: 1057	−5.4	3 (2 × Ser 192, 1 × Gln 196)	8 (1 × Tyr 150, 1 × Glu 153, 1 × Phe 157, 1 × Lys 199, 1 × Ala 291, 1 × Lys 195, 1 × His 242, 1 × Arg 257)
Dihydroxybenzoic acid	2,3-dihydroxybenzoic acid Mol formula: $C_7H_6O_4$ PubChem ID: 19	−6.2	3 (1 × Arg 222, 1 × Ser 287, 1 × Arg 257)	9 (1 × Leu 260, 1 × Ile 290, 1 × Leu 238, 1 × Ala 291, 1 × Arg 257, 1 × Leu 219, 1 × Ile 264, 1 × Ala 261, 1 × Tyr 150)

* Van der Waals, polar, pi–pi interactions, carbon–hydrogen bonds, pi–sigma, pi–alky, etc.

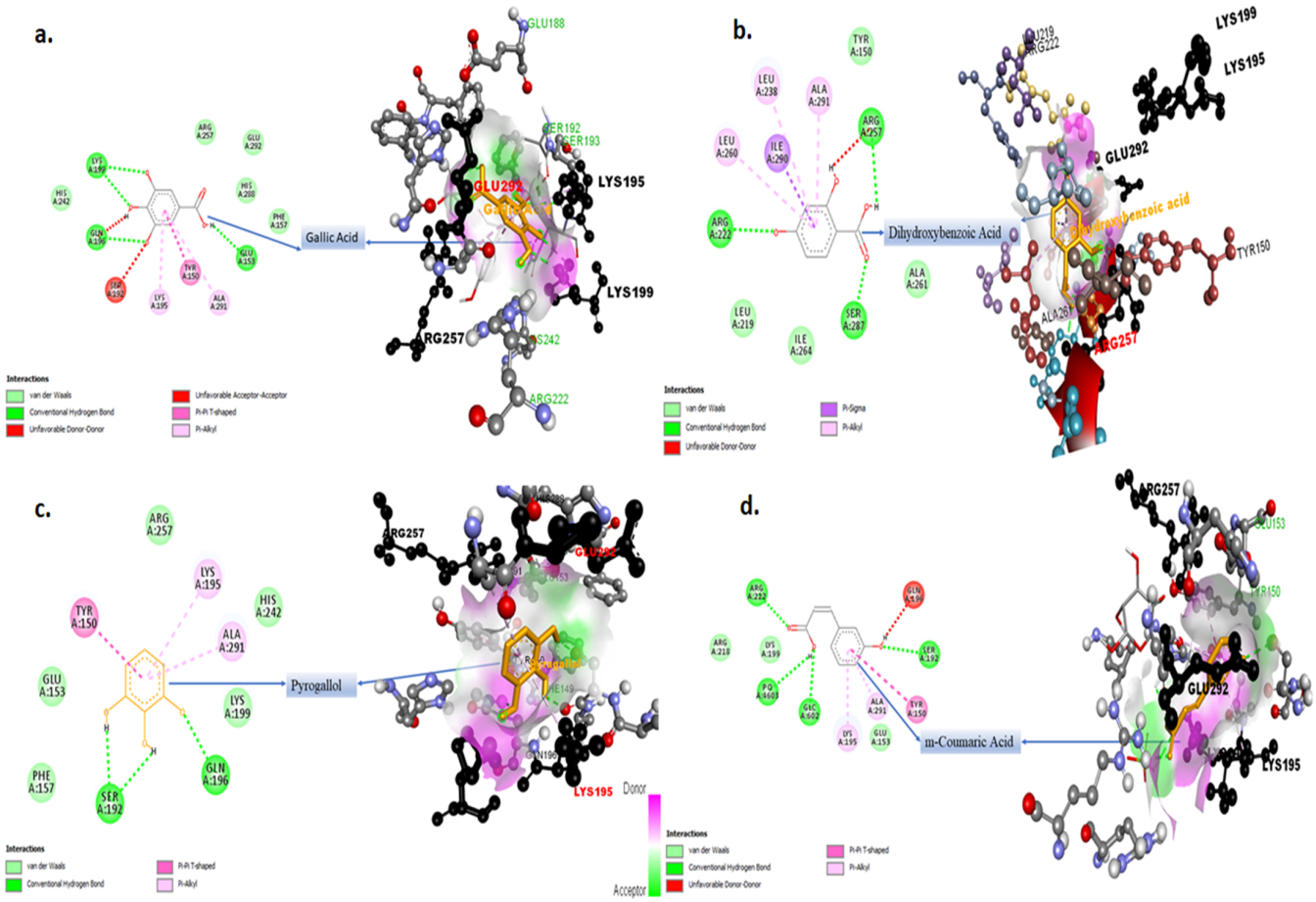


Figure 10. 2D and 3D interaction models of ligands with HSA. Ligands used are (**a**) gallic acid, (**b**) dihydroxy benzoic acid, (**c**) pyrogallol, and (**d**) m-coumaric acid.

3. Discussion

The core cause of many lethal diseases is thought to be connected to metabolic disturbances and inflammatory changes. Diabetes mellitus, characterized by hyperglycemia, is a severe health concern that affects people all over the world [21]. Long-term hyperglycemia has been linked to diabetes complications and biomolecule glycation [22–24].

Nutrition looks at how people might use their food choices to reduce their disease risk and manage their illnesses. If a person's diet lacks the right nutritional balance, they are more prone to develop a variety of health problems. When a person consumes excess or very little amounts of a nutrient, it might cause sickness. A balanced diet, according to growing data, may help you avoid problems including heart disease, cancer, osteoporosis, and type 2 diabetes [25].

Garlic has been reported to have potential therapeutic properties due to its biochemical constituents. Garlic extract can be used against glycation and AGE-related health complications linked with chronic diseases in diabetic patients due to its broad therapeutic potential [26,27]. Compounds such as quercetin, pyrogallol, caffeic acid, gallic acid, m-coumaric acid, and their derivatives are among the beneficial chemicals found in garlics. A high quantity of quercetin has been found in garlic, which is a potent antioxidant compound [28]. The antioxidant capabilities of garlic were well studied in the present investigation through various in vitro methods.

Biochemical analysis of this study showed the presence of phenolic compounds and flavonoids in a respectable amount in garlic extract [28]. According to the current findings, methanolic extract of garlic has a considerable potential to scavenge free radicals, as estimated using DPPH. Garlic also showed high reducing activity that increased with

higher concentrations of the extract. It is generally known that the breakdown of H_2O_2 produces hydroxyl free radicals in the blood. The strong antioxidant properties are attributed to the polyphenolic compounds and polysaccharides. The polysaccharides and polyphenolic compounds in garlic might prevent free radical formation, making it an effective antioxidant.

Lowering chronic inflammation may help to postpone, prevent, and possibly treat a variety of chronic illnesses, including cancer [29]. Besides, many ROS, free radicals including NO, superoxide (O_2^-), and their reaction product peroxynitrite ($ONOO^-$) are produced in excessive quantities during the host's response to infections and inflammatory circumstances [16]. Furthermore, oxidative stress and inflammation are connected in glycation pathways. Identifying alternatives to non-steroidal anti-inflammatory drugs and developing innovative, effective, and safe anti-inflammatory medicines has long been a key priority. External stress as well as several compounds may cause protein denaturation that leads to the loss of structural integrity of proteins and thus loss of their functions [30,31]. Our results suggest that garlic extract can inhibit the heat-induced denaturation of HSA.

Several diseases such as familial amyloidosis, Alzheimer's, pancreatic islet amyloidosis, etc. are caused by protein aggregation in the circulation and in organs [32]. These aggregates undergo further reactions and form amyloid fibrils that contain cross beta structures. In glycation reactions, reducing carbohydrates non-enzymatically and covalently binds to lysine and arginine groups of proteins as well as with the N terminus of polypeptides [33]. Increased levels of these aggregates may cause neurological degeneration. It has been evident that protein glycation can induce aggregation. Intense browning was observed in the polyacrylamide gel electrophoresis glycated samples of albumin [1,33].

Glycation-induced microenvironment structural alterations were observed through spectral studies that included UV spectra and AGE-specific fluorescence. These alterations were inhibited when garlic extract was present in the reaction mixture during the incubation of the glycation reaction. However, a significant inhibition in microenvironment alterations was observed at higher concentrations of the extract. Consequently, extracts lent protection from structural alterations to the protein.

Further in-depth analysis of structural alterations in glycated proteins, including the inhibitory effect of garlic extract on these changes, was investigated using CD. CD analysis of glycated HSA showed protein destabilization and reduction in α-helix structure [34]. It has been reported that changes in CD spectra of glycated HSA are based on glucose concentration. It has also been observed that upon HSA glycation there was a partial denaturation with alterations in structural integrity at different glucose concentrations (1 mg/mL and 5 mg/mL) [35]. Some authors showed that glycated albumin was more favored in a β-sheet conformation structure [36]. These previous studies showed coherence with our CD results of native and glycated HSA. Moreover, CD experimental analysis revealed that garlic extract provided protection of protein secondary structure alterations and inhibited the conversion of α-helix to β-pleated sheet structure. There are possibilities of the interaction of the metabolites or compounds present in garlic extract (Table 3) with the glycation sites on the HSA molecule, causing inhibition in secondary structure alterations (conversion of a-helix to β-sheet). These inhibitions in structural alteration of glycated HSA are highly important for their functional integrity of the protein.

Sudlow Site I is made up of three subsites: the SAL subsite, deep-seated at the bottom of the Sudlow Site I and comprised of hydrophobic residues Leu-238 and Ala-291 and the indomethacin (IMD) and 3′-azido-3′-deoxythymidine (ADT) subsites situated near the opening of the Sudlow Site I are rich in positively charged residues, Arg-218, Lys-195, and Glu-292 [20]. In blood plasma, the D-glucose found in blood plasma is a mixture of two anomers—i.e., α-D-glucopyranose and β-D glucopyranose [37]. The mutarotation between the open aldehyde chain form and ring form is quick and dependent upon medium conditions. Thus, glucose can potentially react, as an open aldehyde form or closed ring form, within a short time frame [37]. Wang Yu et al. reported the mechanism of glucose interaction with HSA [20]. Two molecules of glucopyranose were found to interact

subsequently at the Sudlow Site I [20]. The first molecule gets bottom deep-seated at the SAL subsite, and this configuration is stabilized by the hydrophobic interactions involving Leu238 and Ala291 (Figure 9a). The second glucopyranose molecule is held at the entrance of the Sudlow Site I marked by IMD and AZT subsites. This region is rich in positively charged residues Lys-195, Arg-218, and Glu-292. The oxygen atom bound to C5 (carbon 5 of the second glucopyranose molecule) is attacked (protonated) by the Lys199. This releases C1 atom, which undergoes a covalent interaction with the Lys195, providing strong stability to the second glucopyranose molecule bound to HSA. The presence of a stable ligand at the SAL subsite of the Sudlow Site I reduces the propensity of protonation of the second glucopyranose molecule by the Lys199. It can be observed that quercetin and gallic acid undergoes a conventional hydrogen bonding interaction with Lys199. In both cases the Lys199 serves as a hydrogen donor; hence, this hydrogen is not available for protonation of C5 of the second glucopyranose molecule. Therefore, no covalent interaction could be established between the second glucopyranose and Lys195. This makes the second glucose molecule, at the entrance of Sudlow Site I, more vulnerable for replacement by a more stable competitor. Since quercetin, gallic acid, catechin, and caffeic acid offer more stable interaction as compared with a non-covalently bound glucose molecule (Table 3), they can offer strong competition.

Quercetin appears to be the most promising inhibitor of glucose interaction since it efficiently prevents the formation of a covalent bond by directly interacting with the Lys199. Interaction of quercetin with the SAL subsite is stabilized by the hydrophobic interaction with Ala291 and Leu238—the same amino acids involved in the stabilization of the first glucopyranose molecule at the SAL subsite. The 3-4 dihydroxyphenyl ring bound to C2 of the deep seated 3,5,7-trihydroxychromen-4-one of quercetin at the SAL subsite (Figure 9c) extends well beyond the Lys199 wall, as described by Wang et al. [20]. This sterically hinders the settlement of the second glucopyranose molecule at the opening of the Sudlow Site I, hence inhibiting the glucose interaction. Gallic acid also prevents covalent bond formation (even more efficiently than quercetin due to two possible hydrogen bond interaction); however, due to the small size of trihydroxy benzoic acid, it remains deep-seated at the SAL subsite and well within the wall defined by the Lys199, thereby incapable of preventing the entry of the second glucopyranose molecule at Sudlow Site I.

Uncontrolled levels of blood glucose together with oxidative stress create conditions with high possibilities for the formation of intermediary metabolites of glycation and AGEs. These AGEs and AGE-related metabolites exert structural and functional alterations of blood proteins, contributing to further complications in patients with hyperglycemia as well patients with other inflammatory diseases. In our study, we proved that the natural products that are present in garlic extract have antioxidant and anti-glycation properties and lend protection against the structural destabilization of proteins such as HSA. As a result, this research will help researchers better grasp the link between glycation and natural products that could be beneficial for disease prevention mechanisms in humans.

4. Materials and Methods

4.1. Materials

Ascorbic acid, trichloroacetic acid, DPPH, Folin–Ciocalteau reagent, ferric chloride, potassium ferricyanide, gallic acid, trypsin, quercetin, and Congo red were purchased from Sigma (Saint Louis, MO, USA). Hydrochloric acid, aluminum chloride, mono sodium dihydrogen phosphate, DMSO, ethanol, methanol, disodium hydrogen phosphate, sodium hydroxide, sodium carbonate, and hydrogen peroxide were purchased from Merck (Darmstadt, Germany). All chemicals and reagents were of the highest analytical grade. All solvents used were HPLC grade.

4.2. Chemical and Biological Studies

4.2.1. Preparation of Aqueous Solution of Garlic Extract

Soft-necked varieties of garlic, i.e., *Allium sativum* var *sativum* were obtained from the local markets in Hail, KSA. Fresh aqueous extract was prepared using a previously published method with slight modifications [26]. The concentration of the aqueous solution was determined based on the total garlic used and the final volume of the extract (100 mL of aqueous extract contained 50 g of garlic, i.e., 500 mg/mL).

4.2.2. Qualitative and Quantitative Analysis for Flavonoids and Phenolics

Qualitative analysis: The presence of flavonoids (alkaline reagent test) and phenolic ($FeCl_3$ test) compounds in the aqueous garlic extract was checked using a protocol published by Alsahli et al. with slight modifications [38].

Quantitative analysis of total phenol and flavonoids content: Total phenol content was estimated using Folin– Ciocalteau reagent as described previously with slight modifications [26]. Gallic acid (0–100 µg/mL) served as a reference point. The total phenolic content in garlic extract was determined from the calibration plot and expressed as mg gallic acid equivalents (GAE). All assays were performed in triplicate. Results are expressed as mg gallic acid equivalent per g dry extract.

$$\text{Total phenolic content (TPC)} = C \times V/M$$

where C is the concentration of gallic acid in mg/mL that was obtained from gallic acid calibration curve, V is the volume of plant extract in mL, and M is the weight of pure plant extract in grams (g).

The aluminum chloride ($AlCl_3$) colorimetric method was used to determine total flavonoid content in garlic extract as described previously [12,39]. The calibration curve was plotted using quercetin (0–250 µg/mL). Garlic extract (50 µg/mL) or standard quercetin solution was added to 2% $AlCl_3$ (500 µL). The resultant solution was incubated for 1 h with occasional stirring. The absorbance of reaction mixture was estimated using a spectrophotometer at 430 nm using ethanol as blank, as 2% $AlCl_3$ solution was prepared in ethanol. Total flavonoid content was determined to be quercetin equivalent (mg/g) (mg QUE/g).

$$\text{Total flavonoid content (TFC)} = Z \times V/m$$

where Z is concentration of quercetin (mg/mL); V is volume (mL) of sample used in the extraction; m is weight of pure dried sample used (g).

4.2.3. Antioxidant and Free Radical Scavenging Activities of Extract

Antioxidant activity—reducing power: The ferric reducing antioxidant power (FRAP) technique was used to measure in vitro antioxidant activity [37]. An amount of 1 mL ascorbic acid (0–100 µg/mL) or garlic extract (0–100 µg/mL) was combined with 2.5 mL in phosphate buffer (0.1 M, pH 6.6) and 1% potassium ferricyanide (2.5 mL). After 20 min at 50 °C, 2.5 mL of trichloroacetic acid (10%) was added to the test tubes to terminate reaction. The mixtures were subsequently centrifuged at 3000 rpm for 10 min, causing the formation of supernatant. Finally, freshly made ferric chloride (0.5 mL, 0.5%) solution was mixed with the supernatant (2.5 mL) and distilled water mixture (2.5 mL). The absorbances of various samples were measured at 700 nm.

$$\text{Percentage free radical scavenging activity} = [(X_{control} - X_{sample})/X_{control}] \times 100$$

$X_{control}$ = absorbance of control sample.

X_{sample} = absorbance of sample in the presence of extract.

Analysis of free radical scavenging activity by DPPH method: Antioxidant activity of aqueous garlic extracts was estimated using 1,1 difenyl-2-picryl-hydrazyl (DPPH) as

published before [12]. Dry powder was obtained from the aqueous extract of garlic. This was dissolved in methanol. One milliliter of 0.3 mM DPPH in methanol was added to the extract solution test sample (2.5 mL) of varying concentrations (0–100 µg/mL) and incubated in the dark for 30 min at room temperature. All samples were read at 517 nm with methanol used as a blank. Inhibitory effect of DPPH was calculated according to the following equation:

$$\text{Percentage of free radical scavenging activity} = [(\text{Ac} - \text{As})/\text{Ac}] \times 100$$

where, Ac = absorbance of control, and As = absorbance in presence of extract.

4.3. Modification of HSA by Glucose

HSA glycation was carried out as described previously with minor changes [1,2]. Solutions of HSA (3 mg/mL) with and without glucose were incubated for 10 weeks under identical experimental conditions. To analyze the anti-glycation activity of garlic extract, HSA with glucose was also incubated with varying concentrations of garlic extract (0–100 µg/mL) and incubated under similar conditions. Post-incubation, all samples were dialyzed against PBS. Samples were stored at −20 °C. Protein concentrations were determined by NanoDrop™ 2000/2000c spectrophotometer (Thermo Scientific, USA).

4.4. Protein Denaturation Inhibition by Garlic Extract

The protein denaturation inhibition function of garlic extract was studied as described by Sakat et al. [40] and Pandey et al. [41]. A portion of 1% aqueous solution (500 µL) of HSA was added separately to 100 µL of varying concentrations (0–100 µg/mL) of garlic extract. First, the mixtures were incubated at 37 °C for 20 min and then heated for 20 min at 51 °C. Subsequently, the samples were cooled, and turbidity was measured at 660 nm. All samples were run in triplicate.

The percentage inhibition of protein denaturation was calculated using the following equation:

$$\text{Percentage Inhibition} = [(\text{Ac} - \text{As})/\text{Ac}] \times 100$$

where Ac = absorbance of control, and As = absorbance in presence of extract.

4.5. Measurement of Browning in Glycated Samples

The glycation process is a non-enzymatic event that can produce HSA denaturation. Glycation promotes the browning of proteins by glucose in diabetes. As a result, browning intensity can be used as a preliminary screen for glycation. The intensity of browning of glycated materials was measured by absorbance at 420 nm using a 1 cm path length cell after dilution with distilled water [12]. Experiments were performed in triplicate.

$$\text{Percentage protection from browning} = [(A_{\text{control}} - A_{\text{sample}})/A_{\text{control}}] \times 100$$

where, A_{control} = absorbance of HSA and glucose system, and A_{sample} = absorbance of HSA and glucose system incubated with extract.

4.6. Effect of Garlic Extract on Protein Aggregation Index

Glycation is a major driver of biomolecular structural changes, particularly proteins, resulting in the development of aggregates [42]. Garlic extract was tested for its ability to protect against protein aggregation caused by glycation by determining glycated sample absorbance, treated with glucose, either in the absence or presence of extract (0.78–100 µg/mL). Aggregation index was determined for each sample using their absorbance at 340 and 280 nm [12].

$$\text{Percentage of protein aggregation index} = [A_{340}/(A_{280} - A_{340})] \times 100$$

where, A_{340} = absorbance at 340 nm, and A_{280} = absorbance at 280 nm.

4.7. Spectral Studies

UV absorption measurements were taken with a Perkin Elmer Lambda 35 spectrophotometer (Waltham, MA, USA) with two beams. In the absence/presence of glucose and garlic extract, the UV spectra of HSA (0.2 mg/mL) were measured at a wavelength range of 240–500 nm. Each sample's absorbance intensity was measured at 280 nm [2,42].

AGE-specific fluorescence study: Fluorescence measurements were performed with a Shimadzu spectrofluorometer (model RF-5301PC, Tokyo, Japan). Excitation (350 nm) and emission (400–480 nm) wavelengths were used to determine the generation of fluorescent AGE products. Slit width of 3 nm was used for both excitation and emission [38].

4.8. Circular Dichroism

The CD of native and glycated samples of HSA (2.2 μM), either with or without garlic extracts (0.78–100 μg/mL), was recorded on a Jasco J 810 spectropolarimeter (Tokyo, Japan). The spectropolarimeter has a temperature-controlled sample cell holder attached to a NESLAB model RYE 110 water bath (Tokyo, Japan). [1]. Path length of cuvettes used was 1–10 mm. Each spectrum was the average of three scans. CD spectra were recorded over a wavelength range of 200 to 280 nm and sensitivity of at 5 mm/millidegree (mdeg). Sodium phosphate buffer (20 mM, pH 7.4) was used to prepare all protein solutions. Presence of secondary structural elements was estimated as relative percentage by Chen and Yang equation [19].

4.9. Structural Details of Potential Natural Product Ligands/Inhibitors of HSA Glycation

Natural products contain several phenolic compounds and their metabolites that might be involved in inhibiting the glycation reactions. Some of the competing ligands/inhibitors are given in Table 3, with the HSA molecule characterized using molecular docking studies. All the ligands' .sdf files were derived from PubChem ID (Table 3). The cyclic form of glucose bound to HSA in the crystal structure was obtained from protein data bank (PDB ID: 4iw2) and served as control. The HSA target structure was retrieved from the same PDB ID (4iw2) by removing the ligands. SAL subsite of Sudlow Site I served as the active site.

4.10. Molecular Docking and Scoring of Ligand Poses

Molecular docking of ligands on the active site (defined above) was performed by Autodock-Vina [43]. A standard docking protocol was implemented. The binding efficiency of the glucopyranose and other ligands was also determined by the prodigy-ligand webserver [44,45].

4.11. Statistical Analysis

The experimental results were calculated in triplicate and expressed as mean ± standard error. Statistical analyses were performed using OriginPro 8.5 followed by *t* test. Significance was represented by p values < 0.05.

5. Conclusions

Previous researchers have demonstrated that there is a complex network of interdependencies and correlations among qualities of various natural products, which may work together to provide the overall therapeutic capabilities of different medicinal plants. The results show that the inhibition of heat-induced denaturation of protein and prevention of glycation and AGEs formation increased with an increase in garlic extract concentration. Furthermore, garlic extract exhibited significant levels of protection in protein structural stabilization against glycation. Our research significantly supports the numerous therapeutic properties of garlic and addresses the question of the interdependence of various biological activities and their antioxidant capacity. Thus, it is suggested to include garlic into any existing preventative and treatment approach for glycation-induced health complications in diabetic patients. For this purpose, a well-designed research strategy involving animal

and clinical study is highly recommended for the verification of the therapeutic advantages of garlic extract.

Author Contributions: Conceptualization, M.W.A.K., A.A.O. and S.S.; methodology, M.W.A.K., A.A.O., S.S., W.A.K., R.S. and S.R.V.; validation, N.A., A.K.D.A. and W.A.K.; formal analysis, M.W.A.K., A.A.O., S.S., R.S. and S.R.V.; investigation, M.W.A.K., A.A.O., S.S. and W.A.K.; resources, M.W.A.K., A.A.O., S.S., S.A.A.-Z., E.M.A. and W.A.K.; data curation, M.W.A.K. and A.A.O.; writing—original draft preparation, M.W.A.K. and S.S.; writing—review and editing, N.A., E.M.A., S.A.A.-Z. and A.K.D.A.; supervision, M.W.A.K.; funding acquisition, M.W.A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been funded by the Scientific Research Deanship at the University of Ha'il–Saudi Arabia through project number RG-20 120.

Institutional Review Board Statement: This study was reviewed and approved by the Research Ethics Committee (REC) at the University of Hail dated 27/11/2020 and approved by university president letter number Nr. 20455/5/42 dated 16/04/1442 H.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: Authors are highly thankful to the Molecular Diagnostic and Personalised Therapeutics Unit, University of Ha'il, Ha'il 2440, Saudi Arabia, for providing resources and support for the completion of this research.

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: Samples of the compounds are not available from the authors.

References

1. Khan, M.W.A.; Rasheed, Z.; Khan, W.A.; Ali, R. Biochemical, biophysical, and thermodynamic analysis of in vitro glycated human serum albumin. *Biochemistry* **2007**, *72*, 146–152.
2. Khan, M.W.; Al Otaibi, A.; Al-Zahrani, S.A.; Alshammari, E.M.; Haque, A.; Alouffi, S.; Khan, W.A.; Khan, S.N. Experimental and theoretical insight into resistance to glycation of bovine serum albumin. *J. Mol. Struct.* **2021**, *1230*, 129645. [CrossRef]
3. Sirangelo, I.; Iannuzzi, C. Understanding the role of protein glycation in the amyloid aggregation process. *Int. J. Mol. Sci.* **2021**, *22*, 6609. [CrossRef]
4. Khanam, A.; Ahmad, S.; Husain, A.; Rehman, S.; Farooqui, A.; Yusuf, M.A. Glycation and Antioxidants: Hand in the glove of antiglycation and natural antioxidants. *Curr. Protein Pept. Sci.* **2020**, *21*, 899–915. [CrossRef] [PubMed]
5. Singh, V.P.; Bali, A.; Singh, N.; Jaggi, A.S. Advanced glycation end products and diabetic complications. *Korean J. Physiol. Pharmacol.* **2014**, *18*, 1–14. [CrossRef]
6. Anwar, S.; Khan, S.; Almatroudi, A.; Khan, A.A.; Alsahli, M.A.; Almatroodi, S.A.; Rahmani, A.H. A review on mechanism of inhibition of advanced glycation end products formation by plant derived polyphenolic compounds. *Mol. Biol. Rep.* **2021**, *48*, 787–805. [CrossRef] [PubMed]
7. Vlassara, H.; Uribarri, J. Advanced glycation end products (AGE) and diabetes: Cause, effect, or both? *Curr. Diab. Rep.* **2014**, *14*, 453. [CrossRef] [PubMed]
8. Nowotny, K.; Jung, T.; Höhn, A.; Weber, D.; Grune, T. Advanced glycation end products and oxidative stress in type 2 diabetes mellitus. *Biomolecules* **2015**, *5*, 194–222. [CrossRef] [PubMed]
9. de Groot, L.; Hinkema, H.; Westra, J.; Smit, A.J.; Kallenberg, C.G.; Bijl, M.; Posthumus, M.D. Advanced glycation endproducts are increased in rheumatoid arthritis patients with controlled disease. *Arthritis Res. Ther.* **2011**, *13*, R205. [CrossRef]
10. Fishman, S.L.; Sonmez, H.; Basman, C.; Singh, V.; Poretsky, L. The role of advanced glycation end-products in the development of coronary artery disease in patients with and without diabetes mellitus: A review. *Mol. Med.* **2018**, *24*, 59. [CrossRef]
11. Indyk, D.; Bronowicka-Szydełko, A.; Gamian, A.; Kuzan, A. Advanced glycation end products and their receptors in serum of patients with type 2 diabetes. *Sci. Rep.* **2021**, *11*, 13264. [CrossRef] [PubMed]
12. Anwar, S.; Almatroudi, A.; Allemailem, K.S.; Jacob Joseph, R.; Khan, A.A.; Rahmani, A.H. Protective effects of ginger extract against glycation and oxidative stress-induced health complications: An in vitro study. *Processes* **2020**, *8*, 468. [CrossRef]
13. Mittal, M.; Siddiqui, M.R.; Tran, K.; Reddy, S.P.; Malik, A.B. Reactive oxygen species in inflammation and tissue injury. *Antioxid. Redox Signal.* **2014**, *20*, 1126–1167. [CrossRef]
14. Anwar, S.; Almatoudi, A.; AlSahli, M.A.; Khan, M.A.; Khan, A.A.; Rahmani, A. Natural products: Implication in cancer prevention and treatment through modulating various biological activities. *Anticancer Agents Med. Chem.* **2020**, *20*, 2025–2040. [CrossRef]

15. Koch, H.P.; Lawson, L.D. *Garlic: The Science and Therapeutic Application of Allium Sativum L. and Related Species*, 2nd ed.; Williams & Wilkins: Baltimore, MD, USA, 1996.
16. Murray, M.T. *The Healing Power of Herbs: The Enlightened Person's Guide to the Wonders of Medicinal Plants*, 2nd ed.; Prima: Rocklin, CA, USA, 1995.
17. Colín-González, A.L.; Santana, R.A.; Silva-Islas, C.A.; Chánez-Cárdenas, M.E.; Santamaría, A.; Maldonado, P.D. The antioxidant mechanisms underlying the aged garlic extract- and S-allylcysteine-induced protection. *Oxid. Med. Cell. Longev.* **2012**, *2012*, 907162. [CrossRef] [PubMed]
18. Marrelli, M.; Amodeo, V.; Statti, G.; Conforti, F. Biological properties and bioactive components of *Allium cepa L.*: Focus on potential benefits in the treatment of obesity and related comorbidities. *Molecules* **2018**, *24*, 119. [CrossRef]
19. Chen, Y.H.; Yang, J.T. A new approach to the calculation of secondary structures of globular proteins by optical rotatory dispersion and circular dichroism. *Biochem. Biophys. Res. Commun.* **1971**, *44*, 1285–1291. [CrossRef]
20. Wang, Y.; Yu, H.; Shi, X.; Luo, Z.; Lin, D.; Huang, M. Structural mechanism of ring-opening reaction of glucose by human serum albumin. *J. Biol. Chem.* **2013**, *288*, 15980–15987. [CrossRef]
21. Safari, M.R.; Azizi, O.; Heidary, S.S.; Kheiripour, N.; Ravan, A.P. Antiglycation and antioxidant activity of four Iranian medical plant extracts. *J. Pharmacopunct.* **2018**, *21*, 82–89. [CrossRef]
22. Khan, M.W.A.; Banga, K.; Mashal, S.N.; Khan, W.A. Detection of autoantibodies against reactive oxygen species modified glutamic acid decarboxylase-65 in type 1 diabetes associated complications. *BMC Immunol.* **2011**, *12*, 19. [CrossRef]
23. Khan, M.W.A.; Qadrie, Z.L.; Khan, W.A. Antibodies against gluco-oxidative modified HSA-detected in diabetes associated complications. *Int. Arch. Allergy Immunol.* **2010**, *153*, 207–214. [CrossRef] [PubMed]
24. Khan, M.W.A.; Sherwani, S.; Khan, W.A.; Moinuddin; Ali, R. Characterization of hydroxyl radical modified GAD_{65}: A potential autoantigen in type 1 diabetes. *Autoimmunity* **2009**, *42*, 150–158. [CrossRef] [PubMed]
25. Tuso, P.J.; Ismail, M.H.; Ha, B.P.; Bartolotto, C. Nutritional update for physicians: Plant-based diets. *Perm. J.* **2013**, *17*, 61–66. [CrossRef] [PubMed]
26. Elosta, A.; Slevin, M.; Rahman, K.; Ahmed, N. Aged garlic has more potent antiglycation and antioxidant properties compared to fresh garlic extract in vitro. *Sci. Rep.* **2017**, *7*, 39613. [CrossRef] [PubMed]
27. Khan, M.W.A.; Otaibi, A.A.; Alhumaid, A.F.M.; Alsukaibi, A.K.D.; Alshamari, A.K.; Alshammari, E.M.; Al-Zahrani, S.A.; Almudyani, A.Y.M.; Sherwani, S. Garlic Extract: Inhibition of Biochemical and Biophysical Changes in Glycated HSA. *Appl. Sci.* **2021**, *11*, 11028. [CrossRef]
28. Nagella, P.; Thiruvengadam, M.; Ahmad, A.; Yoon, J.Y.; Chung, I.M. Comosition of polyphenols and antioxidant activity of garlic bulbs collected from different locations of Korea. *Asian J. Chem.* **2014**, *26*, 897–902. [CrossRef]
29. Furman, D.; Campisi, J.; Verdin, E.; Carrera-Bastos, P.; Targ, S.; Franceschi, C.; Ferrucci, L.; Gilroy, D.W.; Fasano, A.; Miller, G.W.; et al. Chronic inflammation in the etiology of disease across the life span. *Nat. Med.* **2019**, *25*, 1822–1832. [CrossRef]
30. Angel, G.R.; Vimala, B.; Nambisan, B. Antioxidant and anti-inflammatory activities of proteins isolated from eight Curcuma species. *J. Phytopharm.* **2013**, *4*, 96–105.
31. Ruiz-Ruiz, J.C.; Moguel-Ordoñez, Y.B.; Segura-Campos, M.R. Biological activity of Stevia rebaudiana Bertoni and their relationship to health. *Crit. Rev. Food Sci. Nutr.* **2017**, *57*, 2680–2690. [CrossRef]
32. Raimundo, A.F.; Ferreira, S.; Martins, I.C.; Menezes, R. Islet Amyloid Polypeptide: A partner in crime with Aβ in the pathology of Alzheimer's disease. *Front. Mol. Neurosci.* **2020**, *13*, 35. [CrossRef]
33. Bouma, B.; Kroon-Batenburg, L.M.; Wu, Y.P.; Brünjes, B.; Posthuma, G.; Kranenburg, O.; de Groot, P.G.; Voest, E.E.; Gebbink, M.F. Glycation induces formation of amyloid cross-beta structure in albumin. *J. Biol. Chem.* **2003**, *278*, 41810–41819. [CrossRef]
34. Monacelli, F.; Storace, D.; D'Arrigo, C.; Sanguineti, R.; Borghi, R.; Pacini, D.; Furfaro, A.L.; Pronzato, M.A.; Odetti, P.; Traverso, N. Structural alterations of human serum albumin caused by glycative and oxidative stressors revealed by circular dichroism analysis. *Int. J. Mol. Sci.* **2013**, *14*, 10694–10709. [CrossRef]
35. Barzegar, A.; Moosavi-Movahedi, A.A.; Sattarahmady, N.; Hosseinpour-Faizi, M.A.; Aminbakhsh, M.; Ahmad, F.; Saboury, A.A.; Ganjali, M.R.; Norouzi, P. Spectroscopic studies of the effects of glycation of human serum albumin on l-trp binding. *Protein Pept. Lett.* **2007**, *14*, 13–18. [CrossRef] [PubMed]
36. GhoshMoulick, R.; Bhattacharya, J.; Roy, S.; Basak, S.; Dasgupta, A.K. Compensatory secondary structure alterations in protein glycation. *Biochim. Biophys. Acta.* **2007**, *1774*, 233–242. [CrossRef] [PubMed]
37. Miwa, I.; Maeda, K.; Okuda, J. Anomeric compositions of D-glucose in tissues and blood of rat. *Experientia* **1978**, *34*, 167–169. [CrossRef] [PubMed]
38. Alsahli, M.A.; Anwar, S.; Alzahrani, F.M.; Almatroudi, A.; Alfheeaid, H.; Khan, A.A.; Allemailem, K.S.; Almatroodi, S.A.; Rahmani, A.H. Health promoting effect of Phyllanthus emblica and Azadiractha indica against advanced glycation end products formation. *Appl. Sci.* **2021**, *11*, 8819. [CrossRef]
39. Ordonez, A.A.L.; Gomez, J.D.; Vattuone, M.A.; Lsla, M.I. Antioxidant activities of Sechium edule (Jacq.) Swartz extracts. *Food Chem.* **2006**, *97*, 452–458. [CrossRef]
40. Sakat, S.S.; Juvekar, A.R.; Gambhire, M.N. In vitro antioxidant and anti-inflammatory activity of methanol extract of Oxalis corniculata linn. *Int. J. Pharm. Pharm. Sci.* **2010**, *2*, 146–155.
41. Pandey, A.K.; Kashyap, P.P.; Kaur, C.D. Anti-inflammatory activity of novel Schiff bases by in vitro models. *Bangladesh J. Pharmacol.* **2017**, *12*, 41–43. [CrossRef]

42. Anwar, S.; Younus, H. Inhibitory effect of alliin from Allium sativum on the glycation of superoxidedismutase. *Int. J. Biol. Macromol.* **2017**, *103*, 182–193. [CrossRef]
43. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [CrossRef] [PubMed]
44. Vangone, A.; Schaarschmidt, J.; Koukos, P.; Geng, C.; Citro, N.; Trellet, M.E.; Xue, L.C.; Bonvin, A.M. Large-scale prediction of binding affinity in protein–small ligand complexes: The PRODIGY-LIG web server. *Bioinformatics* **2019**, *35*, 1585–1587. [CrossRef] [PubMed]
45. Kurkcuoglu, Z.; Koukos, P.I.; Citro, N.; Trellet, M.E.; Rodrigues, J.P.G.L.M.; Moreira, I.S.; Roel-Touris, J.; Melquiond, A.S.J.; Geng, C.; Schaarschmidt, J.; et al. Performance of HADDOCK and a simple contact-based protein-ligand binding affinity predictor in the D3R Grand Challenge 2. *J. Comp. Aid. Mol. Des.* **2018**, *32*, 175–185. [CrossRef] [PubMed]

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/359209006

Breast Cancer Mortality Trends and Predictions to 2030 and Its Attributable Risk Factors in East and South Asian Countries

Article *in* Frontiers in Nutrition · March 2022

DOI: 10.3389/fnut.2022.847920

CITATIONS
0

READS
49

7 authors, including:



Sumaira Mubarik
Wuhan University
70 PUBLICATIONS **995** CITATIONS
SEE PROFILE



Mujahid Iqbal
Wuhan University
1 PUBLICATION **0** CITATIONS
SEE PROFILE



Nawsherwan .....
Wuhan University
20 PUBLICATIONS **270** CITATIONS
SEE PROFILE



Xiaoxue Liu
Department of Epidemiology and Biostatistics•School of Public Health Wuhan Uni…
33 PUBLICATIONS **197** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

The mode of Chinese death and related statistical model View project

Breast cancer incidence and mortality in Asia View project

All content following this page was uploaded by Sumaira Mubarik on 14 March 2022.

The user has requested enhancement of the downloaded file.

frontiers in Nutrition

ORIGINAL RESEARCH
published: 14 March 2022
doi: 10.3389/fnut.2022.847920

Breast Cancer Mortality Trends and Predictions to 2030 and Its Attributable Risk Factors in East and South Asian Countries

*Sumaira Mubarik[1], Rajesh Sharma[2], Syeda Rija Hussain[3], Mujahid Iqbal[4], Nawsherwan[5], Xiaoxue Liu[1] and Chuanhua Yu[1]**

[1] Department of Epidemiology and Biostatistics, School of Public Health, Wuhan University, Wuhan, China, [2] University School of Management and Entrepreneurship Delhi Technological University Delhi, New Delhi, India, [3] Department of Health Sciences, Rawalpindi Medical University, Rawalpindi, Pakistan, [4] Department of Psychology, School of Philosophy, Wuhan University, Wuhan, China, [5] Institute of Cardiovascular Diseases, Xiamen University, Xiamen, China

OPEN ACCESS

Edited by:
Mohd Razif Shahril,
National University of Malaysia, Malaysia

Reviewed by:
Hairong He,
The First Affiliated Hospital of Xi'an Jiaotong University, China
Zhen Zhai,
Second Affiliated Hospital of Xi'an Jiaotong University, China
Ming Lu,
Shandong University, China

***Correspondence:**
Chuanhua Yu
yuchua@whu.edu.cn
orcid.org/0000-0002-5467-2481

Specialty section:
This article was submitted to Nutritional Epidemiology, a section of the journal Frontiers in Nutrition

Received: 03 January 2022
Accepted: 07 February 2022
Published: 14 March 2022

Citation:
Mubarik S, Sharma R, Hussain SR, Iqbal M, Nawsherwan, Liu X and Yu C (2022) Breast Cancer Mortality Trends and Predictions to 2030 and Its Attributable Risk Factors in East and South Asian Countries. Front. Nutr. 9:847920. doi: 10.3389/fnut.2022.847920

Background: Amidst the rising breast cancer burden in Asia, we aim to predict the future mortality risk due to breast cancer and identify the risk-attributable deaths for breast cancer among East and South Asian countries.

Methods: We used country-level data to predict the trends in the next decade relating to female breast cancer mortality by employing data from 1990 to 2019 from the Global Burden of Disease 2019 study. We used the stochastic mortality modeling and prediction techniques to forecast the age-specific and risk-attributable breast cancer mortality trends at the regional and national levels of East and South Asia.

Results: The number of deaths caused by the breast cancer is predicted to increase in East and South Asian countries in the next decade (2020–2030). Age-standardized death rate (ASDR) of breast cancer is predicted to increase by 7.0% from 9.20/100,000 (95% CI: 6.04–12.12) in 1990 to 9.88/100,000 (95% CI: 7.12–11.4) in 2030 in East Asia, and about 35% increase from 13.4/100,000 (95% CI: 9.21–16.02) in 1990 to 18.1/100,000 (95% CI: 13.23–21.10) in 2030 in South Asia. At the national level, the highest percent change in ASDR between 1990 and 2030 was reported in Pakistan (a 62% increase) and Nepal (a 47% increase). The highest percent change in breast cancer mortality between 2020 and 2030 for females of age group 80–84 years was observed in Pakistan [21.6, (95% CI, 20.6–94.7)], followed by Afghanistan [13.3 (4.0–80.8)], and Nepal [36.6 (11.1–125.7)] as compared to the other countries. In the females of aged 50–80 years, the predicted death rates were associated with high body mass index, high-fasting plasma glucose, and diet high in red meat, across the majority of countries under study. Furthermore, reductions in percent change in mortality rates occurred in several countries with increases in sociodemographic index (SDI), notably across high SDI countries.

Conclusion: Breast cancer mortality risk varies substantially across East and South Asian countries with higher mortality risk in low/middle SDI countries. Early detection using screening, awareness among females and health workers, and cost-effective and timely treatment of patients with breast cancer is vital in stemming the tide of breast cancer in the next decade.

Keywords: breast cancer mortality, prediction, Lee–Carter model, life expectancy, risk factors

INTRODUCTION

Developing countries and countries undergoing a westernized lifestyle in Asia may be facing a breast cancer epidemic in the near future due to rising breast cancer cases and deaths (1, 2). Breast cancer has been considered a disease of the Western World in the past but the situation has changed considerably in the past two decades and the disease is not a concern of developed countries alone, but is equally or even more damaging to several Asian countries. In Asia, breast cancer is the most common cancer and the second leading cause of cancer-related deaths among women, with Asia accounting for 39% of all breast cancer incident cases diagnosed globally (2). This rising breast cancer mortality across Asia is posing stress on the oncologic system and hampering an individual's quality of life (2). Previous literature has found better breast cancer survival rates in developed countries than in developing or least-developed ones (3). In 2012, deaths due to breast cancer amounted to 40.8% of the all-cancer deaths in females, ranking second behind lung cancer, illustrating breast cancer as a major health issue with alarming situations in Asia. The age-standardized mortality rate of breast cancer in Asia is 10.2/100,000, which is lower than the global level (12.9/100,000) and matching countries with a medium human development index (2). There were also large differences in breast cancer mortality rates across Asia, ranging from 8.6/100,000 in East Asia to 15.0/100,000 in South Asia in 2017 (4). It has been observed that the mortality-to-incidence ratio (MIR) of breast cancer is high in South-Central Asia and low in Eastern Asia (5).

Globally, an increase in breast cancer has been reported partly due to aging and population expansion (6, 7). Several modifiable risk factors such as lifestyle factors, obesity, excess alcohol consumption, smoking, no or reduced physical activity, and also unhealthy or insufficient diet, if addressed, might help in cancer control and prevention (8, 9). Westernized lifestyle and infection-related etiological factors also have contributions toward increased breast cancer incidence and mortality in developing and underdeveloped countries (2, 10).

Overall, Asia has forty-eight countries with almost 32% of the global female population and most of them are middle- and low-income countries with poor diagnostic and treatment facilities (2). Asian countries constitute a diverse mixture of cultures, geography, ethnicities, and socioeconomic disparities with greatly varied healthcare systems along with large variations in incidence and mortality of breast cancer (2, 4, 11). The majority of Asian countries are developing countries, which does not have robust public health system and healthcare facilities to provide cost-effective diagnostic and therapeutic capacity for patients with cancer and 20% of all the Asian countries have population-based cancer registries (PBCRs), of which only a few countries (e.g., South Korea, Japan, and Singapore) cover the entire population (5). Therefore, due to the lack of representative information, it is pertinent to understand the breast cancer burden using available estimates, which can serve as a valuable tool to develop cancer control policies.

In this study, we examine the geographical disparities in breast cancer mortality at the regional and national levels of East and South Asia, from 1990 to 2019 for women aged 20–84 years. We predicted the breast cancer mortality trends in these regions and countries for the period 2020–2030 using Lee–Carter (LC) model, which accounts for spatial variability and predicts future trends in each country and also allows for mortality risk, longitudinal age trends, and age-adjusted trends over time. To assess the risk-attributable breast cancer mortality in each country under study, we also predicted the mortality trends by risk factors. We conclude by discussing possible underlying risk factors for breast cancer mortality and identifying specific areas of public health concern. To the best of our knowledge, this is the first study using advanced statistical methods to predict breast cancer-related mortality trends by age and risk factors for the next decade at the regional and national levels of East and South Asia.

MATERIALS AND METHODS

Study Population and Data Source

This study used publicly available data from the Global Health Data Exchange (GHDx) (12). As we only used country-specific rates, this study was exempt from ethical review and informed consent according to Code of Federal Regulations [45 CFR 46.101(b)]. This study complies with strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline for observational studies.

The GHDx is a health-related database available at the Institute for Health Metrics Evaluation (IHME) and IHME produces estimates of the burden of diseases, including breast cancer. The estimates in this database are created from a repository of input data, which includes data directly obtained from cancer registries, surveillance data, administrative records, and vital registration. It provides a systematic assessment of 369 injuries and diseases, 87 risk factors, and 284 reported reasons of death among 204 countries, twenty-one regions, and seven super-regions. This study covered

the female population diagnosed with breast cancer as the per International Classification of Diseases (ICD) version 10 code C50 (11). Countries' corresponding population estimates by age group and calendar year from 1990 to 2019 were obtained from the population estimates 1990–2019 via online GHDx (13).

The GHDx database produced as a result of the global burden of disease 2019 study presents the latest estimates, which are produced via innovative estimation methodologies that have been detailed elsewhere (8, 14, 15). Breast cancer mortality estimates were generated using Cause of Death Ensemble Model (CODEm). The CODEm models the mortality estimates for breast cancer based on all available data and covariates consisting of smoking, education, lagging distribution income, sociodemographic index (SDI), and alcohol consumption (16, 17).

The SDI is a composite indicator of a country's lag-distributed income per capita, average years of schooling among persons above 15 years of age, and total fertility rate (TFR) in females under the age of 25 years (18). The composite SDI is calculated as the geometric mean of these three indices for a given location year. The cutoff values used to determine quintiles for analysis were then calculated using country-level estimates of SDI for the year 2019. This index (SDI) ranges from 0 to 1, i.e., least developed to the most developed ones. The cutoff points for low, low-middle, middle, high-middle, and high SDI countries in 2019 are 0.4547, 0.6076, 0.6895, 0.7905, and 0.8051, respectively.

Attributable Risk Factors

The comparative risk assessment method has already been integrated into the Global Burden of Disease (GBD) study (19) to measure the burden of multiple causes and impairments of 87 environmental, occupational, metabolic, and behavioral risk factors. Briefly, we chose the mortality component to model the attributable burden of risk factors, following an assessment of the causal evidence in each risk-outcome pair (20).

Statistical Analysis

In this analysis, death rates (per 100,000) of female breast cancer, corresponding to age groups 20–84 years with 5-year age intervals, and successive thirty calendar years from 1990 to 2019, were considered for modeling analysis. LC model (21) was applied to breast cancer death data of each East and South Asian country to project the death rates for the forthcoming 11 years (2020–2030). The model was applied to the two age groups, 20–84 and 50–84 years, by risk factors from 1990 to 2019. The data includes the number of deaths $D_{x,t}$ and exposure to death $E_{x,t}$ for a person of age x last birthday during year t. The central mortality rate $m_{x,t}$ is calculated using the formula $m_{x,t} = \frac{D_{x,t}}{E_{x,t}}$. The data to be used are annual age-specific death rates (22).

The LC-based modeling frameworks are viewed in the current literature as among the most efficient and transparent methods of modeling and projecting mortality improvements. Previously, scholars have used this approach to forecast all-cause and cause-specific mortality in different countries (22–26). The general expression of the model is expressed as: $m_{x,t} = e^{a_x+b_x+k_t+\varepsilon_{x,t}}$, where, $m_{x,t}$ is the central mortality rate at age x and year t, and a_x is the average (over time) log-mortality at age x, and b_x measures the response at age x to changes in the overall level of mortality over time, and k_t represents the overall level of mortality in year t, and $\varepsilon_{x,t}$ is the residual term. The Singular Value Decomposition (SVD) algorithm was used to forecast the general mortality index for the period 2020 to 2030, for each country separately.

Based on age-standardized mortality estimates of 2019, we ranked the risk factors of breast cancer for East and South Asian countries. The four risk factors that were ranked first were selected for further prediction analysis (**Supplementary Figure 1**). The spatial/regional disparity in mortality trends attributable to high body mass index (BMI), high fasting plasma glucose (FPG), alcohol use, and diet high in red meat were predicted from 1990 to 2030 among the female population of age 50–84 years. In addition, the autoregressive integrated moving average (ARIMA) method (27, 28) was employed to predict the SDI values from 1990 to 2030 for each country.

The mortality modeling methodology was implemented using publically available R programming functions with package ilc (29), which facilitates both the preparation of mortality data, fitting, and analysis of the given log-linear modeling structures. The package also incorporates methods to produce future mortality rates forecasts and compute the corresponding future life expectancy. Further, the forecast package in R (30) was used to predict the future SDI values for each country.

This study adopted the following steps of analysis. First, we predicted the age-specific mortality rates for 11 years from 2020 to 2030 and then the change between 2019 and 2030 in estimated age-specific mortality rates was expressed as a percentage (the difference between morality rates at two points divided by the initial mortality value and multiplied by 100). Age-standardized mortality rates were computed using the direct method based on the WHO World Standard Population (31). The direct calculation of the age-standardized mortality rate was as follows (32):

$$\textit{Age standardised mortality rate} = \frac{\sum \textit{Age specific rate} \times \textit{Standard population in corresponding age group}}{\sum \textit{Standard population}}$$

We also predicted risk attributable mortality rates by potential breast cancer risk factors among females of aged 50–84 years. Finally, we projected the SDI value for the same prediction period and assessed the association between SDI and mortality changes using the spearman correlation coefficient test during the entire period from 1990 to 2030. Additionally, to test the sensitivity of the method, we evaluated the performance of the LC model in comparison with an alternative stochastic mortality model called Booth–Maindonald–Smith (BMS) model (33) and predicted the breast cancer mortality trends by using both the models. A detailed description of about BMS method has been provided elsewhere (22, 33).

RESULTS

Fitting of the Model to Breast Cancer Mortality

The results of estimation of parameters in the LC model, for twelve locations (2 regions and 10 countries) are described in **Supplementary Figure 2**. The estimated values of age-dependent parameters α_x and β_x, and the time-dependent parameter k_t are reported for each country. Applying SVD to the matrix $M_{x,t}$, we obtained an explained variance of 83.2, 82.8, 81.7, 90.9, 94, 73.9, 85.5, 74.5, 96.8, 94.5, 89, and 76% by fitted LC model for East Asia, China, Japan, North Korea, South Korea, Mongolia, South Asia, India, Pakistan, Afghanistan, Nepal, and Bangladesh, respectively. These results indicate that the LC model fitted breast cancer-related mortality data reasonably well. Further, the forecasted breast cancer death rates from 2020 to 2030 had

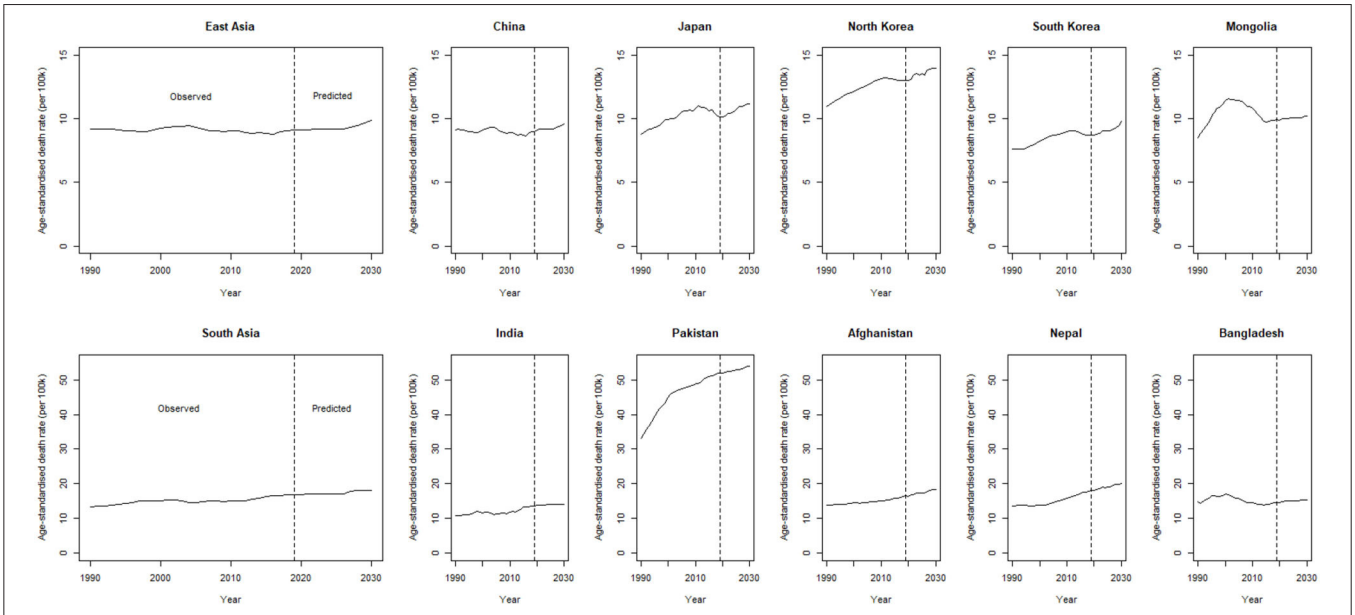


FIGURE 1 | Trends in observed and predicted age-standardized death rate (ASDR) (per 100,000) of breast cancer, from 1990 to 2030 at a regional and national level of East and South Asia.

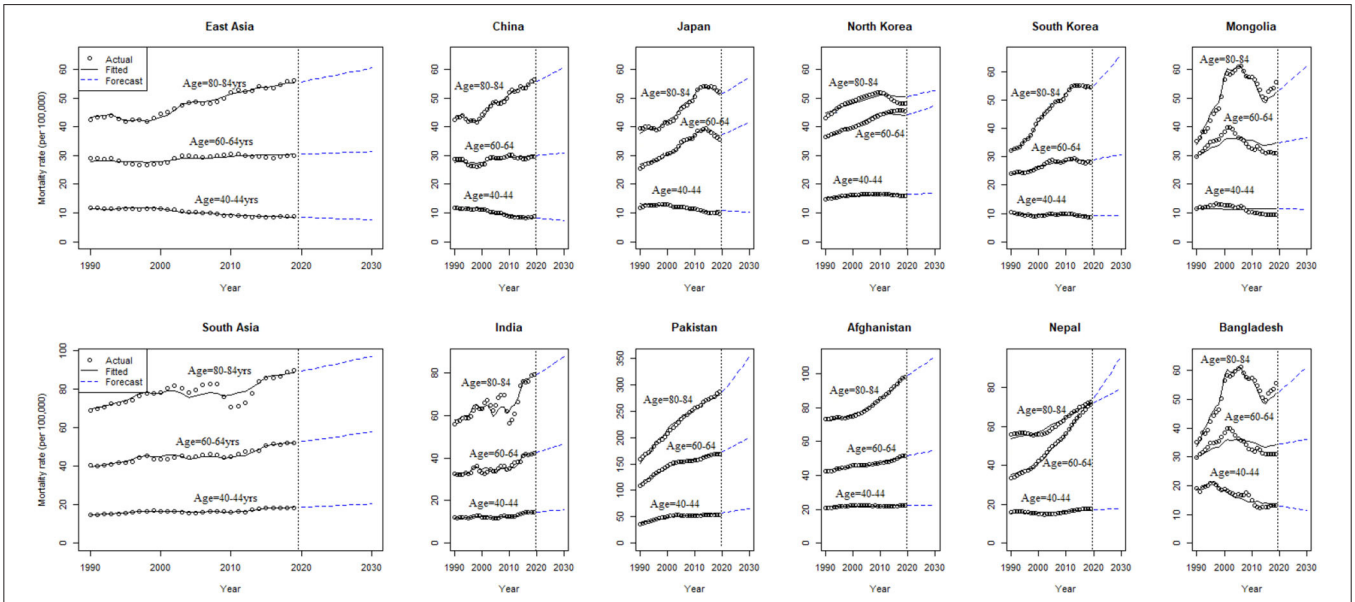


FIGURE 2 | Age-specific observed, estimated, and predicted breast cancer mortality rates (per 100k) from 1990 to 2030, at the regional and national level of East and South Asia.

low values of mean squared error (MSE), Akaike's Information Criteria (AIC), and Bayesian Information Criteria (BIC).

Trends and Changes in Age-Standardized Breast Cancer Death Rates From 1990 to 2030

The trends in observed and predicted age-standardized death rates (ASDRs) (per 100,000) of female breast cancer at regional and national levels from 1990 to 2030 are depicted in **Figure 1**. Overall ASDR of breast cancer is predicted to increase by 7.0% from 9.20 (95% CI: 6.04–12.12) per 100,000 in 1990 to 9.88 (95% CI: 7.12–11.4) per 100,000 in 2030 in East Asia, and by about 35% from 13.4 (95% CI: 9.21–16.02) per 100,000 in 1990 to 18.1 (95% CI: 13.23–21.10) per 100,000 in 2030 in South Asia. At the national level, the highest percent change in ASDR between 1990 and 2030 is predicted in Pakistan (a 62% increase) and Nepal (a 47% increase) (**Figure 1**, **Supplementary Table 1**).

The age-specific breast cancer mortality rates (per 100,000) in East and South Asian regions from 1990 to 2030 are shown in **Figure 2**. Mortality rates are predicted to increase with age, particularly for the age group 80–84 years in all the East and South Asian regions. A significantly higher percent change in breast cancer mortality between 2019 and 2030 for females of age group 75–79 years is observed in South Korea [21.0 (95% CI: 10.9–71.2)] and North Korea [9.4 (6.9–61.8)] among East Asian regions. The predicted breast cancer mortality rates are greater in South Asian regions than East Asian and significantly higher mortality percent change between 2019 and 2030 is predicted in Pakistan [21.6 (20.6–94.7)], Afghanistan [13.3 (4.0–80.8)], and Nepal [36.6 (11.1–125.7)] for age group 80–84 years (**Table 1**).

Breast Cancer Mortality and Attributable Risk Factors

The high BMI- and high FPG-related mortality rates (per 100,000) are increasing in both the regions, with the highest increase observed in Pakistan and Afghanistan during 2020–2030. The alcohol association to increase mortality risk is expected in Japan, Mongolia, South Korea, and Nepal during the next decade than other countries. Six out of ten countries (China, Mongolia, South Korea, Pakistan, Afghanistan, and Nepal) are predicted to have high mortality attributable to a diet high in red meat during 2020–2030 (**Figure 3**). The age-specific mortality rates (per 100,000) by different risk factors are projected from 2020 to 2030 and are presented in **Figure 4**. Overall increasing mortality trend is predicted from 2020 to 2030 for the females in the 75-plus age groups who had high FPG and high BMI. The risk-attributable mortality rates are higher in South Asian countries when compared with the East Asian countries.

Additionally, we also reported life expectancy trends of patients with breast cancer of age 50–84 years (all-cause) (**Supplementary Figure 3**). We observed that overall in East Asia, China, and South Korea, the life expectancy of a female patient with breast cancer might increase during the prediction period 2020–2030; whereas for other locations, it is expected to decrease or remain constant during the same period (**Supplementary Figure 3**).

Trend Prediction in SDI and Its Association With Reduced Mortality Risk

The country- and region-wise trends in SDI for 1990 to 2030 are given in **Figure 5**. We observe an increasing SDI trend during the entire study duration among all the Asian regions and countries, including East and South Asia, with comparatively high SDI predicted in Japan and South Korea.

Figure 6 presents the association between SDI and mortality rates from 1990 to 2030. A significantly high negative correlation ($r \geq 0.83$, $p \leq 0.001$) was observed between SDI and mortality rate across all the countries in both the prediction and forecasted periods (1990, 2000, 2010, 2020, and 2030). **Figure 6F** shows the association between percent change in SDI between 1990 and 2030 and percent change in mortality rates in East and South Asian countries. Decreases in percent change in mortality rates did occur in many locations with increases in percent change in SDI, notably, for locations belonging to the high SDI quintiles (Japan and South Korea). The largest increases in mortality rate are expected to occur in low-middle-SDI and low-SDI countries led by Pakistan.

As a sensitivity check, we estimated the mortality change using both models (LC and BMS; **Supplementary Figure 4**) and evaluated the model's performance by explaining variance (%) (**Supplementary Table 2**) and error measures (**Supplementary Table 3**) including MSE. We predicted the death rates from 2020 to 2030 using both the models (**Supplementary Figure 5**) and almost similar results were obtained by these methods.

DISCUSSION

This study estimated the future trends of breast cancer-related mortality in East and South Asian countries. We also predicted the future SDI trends of regions and countries from 1990 to 2030 and analyzed the relation between SDI and breast cancer mortality risk. There was an inverse association between SDI and mortality rates of breast cancer within East and South Asian countries. The largest percent increase in mortality rate occurred in low-middle-SDI and low-SDI countries led by Pakistan. These findings suggest the contribution of socioeconomic development in reduced risk of breast cancer mortality.

Sociodemographic factors, lack of beliefs in preventive health, health literacy, cultural influences, socioeconomic status as well as lack of access to healthcare are the challenges to breast cancer management and care in the countries under this study (34–37). Breast cancer mortality is predicted to increase continuously; however, if effective control and prevention are not implemented, mortality is expected to continue to surge even more in the coming decades because of substantial demographic changes, which may create considerable public health challenges for these Asian countries. Previous studies have also found that breast cancer-related mortality has been increasing more rapidly in Asia than the global average (38, 39). These increasing trends have many possible reasons apart from demographic changes, which includes risk factors such as smoking, alcohol consumption, physical inactivity, low fiber and high-fat diet, obesity (40, 41),

TABLE 1 | Estimated age-specific mortality rates (per 100 k), percent change between 2019 and 2030, for breast cancer, in East and South Asian regions.

	2019 Mortality rate (95% CI)	2030 Mortality rate (95% CI)	%Change, between 2019 and 2030
East Asia			
Age 20–24	0.2 (0.2, 0.3)	0.1 (0.1, 0.2)	−32.9 (−66.9, 39.4)
25–29	0.7 (0.5, 0.9)	0.5 (0.3, 0.7)	−27.5 (−62.6, 43.8)
30–34	2.3 (1.8, 3)	1.6 (1.1, 2.5)	−30 (−64.1, 39.7)
35–39	5.2 (4, 6.6)	4.1 (3.1, 5.6)	−20.6 (−53.3, 38.4)
40–44	8.9 (7, 11.1)	7.5 (5.8, 9.8)	−15.4 (−47.6, 38.5)
45–49	12.4 (9.8, 15.5)	9.7 (6.4, 14.5)	−21.8 (−58.5, 48.4)
50–54	19.7 (15.6, 24.5)	18.7 (16.3, 21.4)	−5.2 (−33.4, 37.4)
55–59	26.7 (21.4, 33.4)	29.2 (27.2, 31.3)	9.1 (−6.3, 27.1)
60–64	30 (24.2, 36.8)	32.1 (29.4, 35.1)	6.9 (−4.6, 21.4)
65–69	35 (28.6, 42.6)	37.7 (32.9, 43.2)	7.7 (1.3, 15)
70–74	41.2 (33.5, 50.1)	43.9 (37.6, 51.3)	6.5 (2.3, 12.2)
75–79	50.1 (41.4, 60.3)	55.2 (44, 69.2)	10.1 (6.3, 14.7)
80–84	56 (45.4, 66.8)	61.1 (49.5, 75.4)	9 (8.1, 13)
China			
Age 20–24	0.2 (0.1, 0.3)	0.1 (0.1, 0.2)	−33.9 (−72.8, 65.4)
25–29	0.7 (0.5, 0.9)	0.5 (0.3, 0.8)	−28.2 (−68.3, 66.6)
30–34	2.3 (1.7, 2.9)	1.6 (0.9, 2.8)	−31.1 (−69.7, 61.7)
35–39	5.1 (3.9, 6.5)	4 (2.7, 6)	−21.4 (−58.8, 55.7)
40–44	8.7 (6.7, 11)	7.3 (5.1, 10.4)	−16.2 (−53.5, 54.4)
45–49	12 (9.3, 15.2)	9.1 (5.2, 16.1)	−23.5 (−66, 73.5)
50–54	19.3 (15.1, 24.2)	18.1 (14.8, 22)	−6.4 (−38.8, 45.8)
55–59	26.2 (20.6, 33.3)	28.7 (26.2, 31.5)	9.4 (−5.3, 26.8)
60–64	29.6 (23.5, 36.7)	31.7 (28.2, 35.6)	7.1 (−3.1, 19.8)
65–69	34.8 (28.1, 42.8)	37.2 (31.2, 44.3)	7.0 (3.7, 11.0)
70–74	41.1 (33.3, 50.3)	43.6 (35.6, 53.3)	6.0 (5.9, 7.2)
75–79	50.3 (41.2, 61)	55.3 (40.9, 74.7)	10 (−0.7, 22.5)
80–84	56.3 (45.3, 67.6)	61.4 (46.3, 81.4)	9.1 (2.3, 20.5)
Japan			
Age 20–24	0.1 (0.1, 0.1)	0.1 (0.1, 0.1)	10.2 (4.8, 16.1)
25–29	0.6 (0.5, 0.6)	0.6 (0.6, 0.6)	4.2 (0, 8.1)
30–34	2.1 (1.9, 2.2)	2.1 (1.9, 2.3)	2.1 (−0.8, 4.1)
35–39	4.9 (4.6, 5.3)	5.1 (4.6, 5.6)	3.6 (−0.5, 5.7)
40–44	9.8 (9.3, 10.4)	10.2 (9.6, 10.9)	4.9 (3.5, 5.3)
45–49	16.6 (15.7, 17.5)	17.9 (17.4, 18.5)	8.3 (5.7, 10.6)
50–54	25.5 (24.2, 26.9)	28.5 (28, 29.1)	12 (3.8, 20.1)
55–59	32.2 (30.4, 34.2)	37.3 (34.9, 39.8)	15.8 (2.1, 30.9)
60–64	35.5 (33.3, 37.8)	40.3 (36.4, 44.6)	13.3 (−3.7, 34)
65–69	37.8 (34.6, 40.3)	41 (37.2, 45.2)	8.6 (−7.6, 30.9)
70–74	39.2 (34.6, 42.5)	42.3 (38.6, 46.4)	7.9 (−9.1, 34.3)
75–79	46.4 (38.4, 51.2)	50.2 (45.3, 55.7)	8.3 (−11.5, 45.2)
80–84	52.4 (38, 60.4)	55.6 (50.5, 61.3)	6.2 (−16.5, 61.5)
North Korea			
Age 20–24	0.3 (0.2, 0.5)	0.3 (0.3, 0.3)	−0.7 (−44.1, 86.4)
25–29	1.1 (0.6, 2)	1.1 (1.1, 1.1)	−0.1 (−44.9, 101.2)
30–34	3.7 (1.8, 6.5)	3.7 (3.7, 3.7)	−1 (−44.2, 98.8)
35–39	8.9 (4.6, 15.7)	9.2 (9.1, 9.3)	3.4 (−41.8, 102)
40–44	16 (8.8, 26.5)	16.9 (16.6, 17.1)	5.1 (−37.3, 94.5)
45–49	26.9 (15.9, 42.9)	28.5 (27.8, 29.3)	5.9 (−35.2, 84.6)

(Continued)

TABLE 1 | Continued

	2019 Mortality rate (95% CI)	2030 Mortality rate (95% CI)	%Change, between 2019 and 2030
50–54	35.7 (21.2, 55.1)	37.8 (36.6, 39)	6 (−33.5, 84)
55–59	44.8 (28.1, 69.4)	46.7 (44.9, 48.5)	4.1 (−35.3, 72.4)
60–64	45.4 (29.2, 67.1)	47.1 (45.4, 48.9)	3.8 (2.4, 67.5)
65–69	45.4 (29.8, 65.6)	48.1 (46.3, 50)	5.9 (4.4, 67.5)
70–74	43.7 (30.8, 61.2)	47.2 (45.7, 48.7)	8 (5.4, 58)
75–79	46.5 (32.3, 65)	50.9 (49.5, 52.3)	9.4 (6.9, 61.8)
80–84	48.1 (32.5, 68.8)	52.4 (51.3, 53.5)	8.9 (2.4, 64.9)
South Korea			
Age 20–24	0.1 (0.1, 0.1)	0.1 (0.1, 0.1)	−14 (−36.7, 21.5)
25–29	0.5 (0.3, 0.7)	0.4 (0.4, 0.5)	−14 (−35.2, 14.9)
30–34	2.2 (1.6, 2.9)	2.1 (1.9, 2.2)	−5.7 (−23.5, 20.1)
35–39	5.1 (3.8, 6.6)	5.6 (5.6, 5.6)	10.9 (−14.4, 46.1)
40–44	8.7 (6.7, 10.8)	9.2 (9.2, 9.3)	6.3 (−13.6, 36.8)
45–49	13.6 (10.9, 16.6)	13.8 (13.6, 14)	1.7 (−15.8, 25.1)
50–54	18.2 (14.5, 21.9)	18.5 (18.4, 18.7)	2.1 (−14.5, 26.4)
55–59	24.2 (19.6, 29.1)	25.6 (25.2, 26.1)	5.8 (−13.4, 32.8)
60–64	28.1 (22.7, 34.3)	30.5 (29.5, 31.4)	8.3 (−13.9, 38.5)
65–69	33.6 (27.2, 41.1)	36.6 (34.8, 38.5)	8.9 (5.2, 41.3)
70–74	38.7 (31.1, 47.3)	43.5 (40.8, 46.4)	12.4 (10.6, 49.2)
75–79	48 (37.5, 58.9)	58 (52.5, 64.2)	21 (10.9, 71.2)
80–84	54.6 (39.9, 71.3)	63.7 (58.2, 69.7)	16.6 (13.3, 75)
Mongolia			
Age 20–24	0.2 (0.1, 0.4)	0.3 (0.3, 0.3)	10.5 (−37.3, 111)
25–29	0.6 (0.3, 1)	0.7 (0.6, 0.9)	20.1 (−36.4, 146.5)
30–34	2.8 (1.6, 4.6)	3.3 (2.8, 3.8)	15 (−38.5, 129.1)
35–39	5.8 (3.6, 8.9)	6.8 (6.1, 7.6)	16.2 (−32.2, 111.9)
40–44	9.5 (5.8, 14.2)	11.9 (10.6, 13.4)	25.4 (−25.7, 130.4)
45–49	15.6 (9.9, 23.8)	19.4 (16.6, 22.6)	24.4 (−30.2, 128.4)
50–54	21.8 (13.8, 32.5)	26.6 (23.1, 30.7)	21.8 (−29, 122.1)
55–59	27.3 (17.5, 39.8)	32.7 (28, 38.2)	19.6 (−29.7, 118.6)
60–64	30.9 (20.7, 45.3)	35.6 (31.7, 40)	15.2 (−30.1, 93.8)
65–69	38.1 (25.4, 56.2)	41.1 (36.7, 46.1)	8 (−34.8, 81.9)
70–74	43.2 (29.2, 61.6)	43 (40.5, 45.7)	−0.3 (−34.2, 56.7)
75–79	48.5 (32.6, 68.5)	49.2 (43, 56.2)	1.4 (−37.1, 72.5)
80–84	55.6 (36.8, 81.1)	53.5 (47, 61)	−3.7 (−42.1, 66)
South Asia			
Age 20–24	1 (0.8, 1.3)	1.2 (1.1, 1.4)	18.8 (−17, 77.3)
25–29	2.4 (1.9, 3)	2.9 (2.6, 3.2)	18.8 (−14.7, 68.5)
30–34	4.7 (3.7, 5.8)	5.1 (4.8, 5.5)	8.8 (−17.3, 46.8)
35–39	9.3 (7.5, 11.6)	10.2 (9.5, 10.8)	8.9 (−18.1, 45)
40–44	18.4 (14.5, 22.4)	20.5 (19, 22.1)	11.3 (−15.3, 52.1)
45–49	26.9 (21.7, 32.9)	29.9 (28.2, 31.7)	11.3 (−14.1, 45.8)
50–54	45.9 (35.3, 57.3)	53.6 (47.9, 59.8)	16.6 (−16.4, 69.5)
55–59	50.5 (40.2, 61.6)	58.2 (53, 64)	15.4 (14, 59.1)
60–64	52.2 (42.9, 62.4)	58.8 (53.7, 64.3)	12.5 (−14, 49.8)
65–69	58.7 (48.1, 69.5)	64.1 (59.6, 68.9)	9.1 (4.3, 43.2)
70–74	70.1 (57.6, 83.5)	75.5 (70.6, 80.6)	7.7 (−15.5, 40)
75–79	75.3 (61.8, 89.2)	75.8 (73.6, 78.1)	0.7 (−17.5, 26.4)
80–84	89.7 (71.3, 108.1)	96.6 (89.8, 103.9)	7.7 (5.9, 45.7)

(Continued)

TABLE 1 | Continued

	2019 Mortality rate (95% CI)	2030 Mortality rate (95% CI)	%Change, between 2019 and 2030
India			
Age 20–24	0.6 (0.5, 0.8)	0.6 (0.6, 0.6)	−1.4 (−27.6, 39)
25–29	1.6 (1.2, 2.1)	1.7 (1.5, 1.9)	6.7 (−25.8, 58.2)
30–34	3.7 (2.8, 4.8)	3.7 (3.5, 3.8)	−0.2 (−26.5, 38)
35–39	8.2 (6.3, 10.6)	8.6 (8.1, 9.2)	5 (−23.4, 47)
40–44	14.4 (10.8, 18.7)	16 (14.4, 17.7)	11.5 (−22.7, 64.5)
45–49	20.5 (15.7, 26.3)	22.6 (20.9, 24.3)	10.3 (−20.4, 55.1)
50–54	36.7 (25.5, 49.2)	47 (37.9, 58.3)	28.1 (−23.1, 128.2)
55–59	40.3 (29.5, 53.1)	46.6 (40.6, 53.5)	15.7 (3.5, 81.4)
60–64	42.4 (32.5, 54.1)	47.6 (42, 53.9)	12.2 (9.4, 65.9)
65–69	48 (38.2, 60.3)	51.7 (47.3, 56.5)	7.7 (5.7, 47.9)
70–74	57.4 (44.9, 72.6)	60.9 (56.1, 66.1)	6.2 (4.7, 47.2)
75–79	64.2 (50.2, 80.5)	64.9 (62.1, 67.9)	1.1 (−22.9, 35.2)
80–84	79.3 (59.5, 102)	87.9 (77, 100.2)	10.8 (2.5, 68.3)
Pakistan			
Age 20–24	3.8 (2.4, 5.7)	4.9 (4.5, 5.3)	28 (−21.1, 121.4)
25–29	8.7 (5.7, 12.4)	10.9 (10.2, 11.6)	25 (−17.6, 102.2)
30–34	14.1 (9.2, 20.6)	17.1 (16.2, 18.2)	21.4 (20.6, 98.1)
35–39	21.7 (14.8, 32.5)	25.6 (24.4, 26.8)	17.8 (12.9, 81.7)
40–44	54 (35.7, 80.2)	66.1 (63.1, 69.2)	22.3 (−21.4, 94.1)
45–49	83 (56.7, 122.7)	98.6 (95.4, 101.9)	18.8 (10.3, 79.7)
50–54	123.4 (85.9, 175.2)	146.3 (140.8, 151.9)	18.5 (14.6, 76.7)
55–59	151 (106.1, 214.9)	181.2 (173.2, 189.5)	20 (19.4, 78.6)
60–64	169 (118.3, 232.5)	200.8 (191.1, 211.1)	18.8 (17.8, 78.4)
65–69	194 (137.8, 269.2)	231.9 (219.8, 244.6)	19.5 (18.4, 77.5)
70–74	238.5 (168.4, 327.3)	284.3 (267.8, 301.7)	19.2 (18.2, 79.2)
75–79	251.7 (172.8, 352.9)	298.6 (281.3, 316.9)	18.6 (17.3, 83.3)
80–84	286.3 (191.8, 419.6)	348.3 (324.8, 373.5)	21.6 (20.6, 94.7)
Afghanistan			
Age 20–24	0.9 (0.5, 1.4)	0.8 (0.8, 0.8)	−4.4 (−40.4, 62.5)
25–29	2.1 (1.2, 3.5)	2 (2, 2)	−7.4 (−42.5, 60.3)
30–34	5.9 (3.4, 9.6)	5.4 (5.4, 5.5)	−7.8 (−42.9, 57.4)
35–39	11.6 (7.2, 18.1)	11.3 (11.2, 11.3)	−3.1 (−37.8, 56.4)
40–44	22.2 (13.9, 32.8)	22.5 (22.5, 22.6)	1.6 (−31.5, 63.2)
45–49	27.2 (17, 40.8)	27 (26.9, 27.2)	−0.6 (−34, 59.5)
50–54	33.2 (20.7, 50)	34.2 (33.8, 34.6)	3.1 (2.4, 66.8)
55–59	47.1 (30.3, 68.6)	49.7 (48.8, 50.5)	5.5 (4.9, 66.7)
60–64	51.7 (33.2, 77.2)	55.1 (54, 56.1)	6.5 (3.2, 69.0)
65–69	51.2 (33.3, 72.9)	56.6 (55.1, 58.1)	10.5 (4.4, 74.5)
70–74	59.2 (39.7, 86.6)	65.8 (63.8, 67.8)	11.1 (6.3, 70.9)
75–79	84 (55.4, 120.2)	94.9 (91.7, 98.2)	12.9 (7.7, 77.2)
80–84	98.1 (63.7, 143)	111.2 (107.3, 115.2)	13.3 (4.0, 80.8)
Nepal			
Age 20–24	0.2 (0.1, 0.3)	0.2 (0.2, 0.2)	6 (−34.9, 91.3)
25–29	0.6 (0.3, 1)	0.6 (0.6, 0.6)	−7.4 (−41.5, 67.5)
30–34	2.2 (1.2, 3.5)	2.2 (2.2, 2.2)	0.6 (−38, 89.6)
35–39	6.3 (3.7, 9.8)	6.3 (6.2, 6.3)	−0.5 (−37.2, 70.8)
40–44	17.8 (10.9, 27.1)	17.8 (17.6, 18)	0 (−35, 65.1)
45–49	24.4 (15.2, 38.1)	22.6 (22.4, 22.8)	−7.2 (−41, 50.5)

(Continued)

TABLE 1 | Continued

	2019 Mortality rate (95% CI)	2030 Mortality rate (95% CI)	%Change, between 2019 and 2030
50–54	45.6 (28.6, 68.1)	45.6 (44.9, 46.4)	0.2 (−34.1, 62.3)
55–59	87.8 (57.8, 126.9)	101.1 (98.8, 103.5)	15.2 (12.1, 78.9)
60–64	72.9 (47.6, 103.7)	79.6 (77.5, 81.7)	9.2 (6.3, 71.8)
65–69	70.2 (45.8, 103.6)	79.1 (76.5, 81.8)	12.6 (10.2, 78.7)
70–74	64.5 (43.7, 95.1)	76.5 (73.3, 79.8)	18.6 (14, 82.4)
75–79	68.1 (45.1, 96.6)	85 (80.5, 89.8)	24.9 (16.7, 99.1)
80–84	71.5 (46.4, 102.4)	97.6 (91.1, 104.7)	36.6 (11.1, 125.7)
Bangladesh			
Age 20–24	0.9 (0.5, 1.4)	1 (0.6, 1.7)	14.6 (−57.3, 223.4)
25–29	2 (1.2, 3.1)	2.6 (1.1, 5.9)	29.2 (−64.9, 386.3)
30–34	3 (1.8, 4.5)	4.7 (1.3, 16.1)	55.5 (−70.5, 795.6)
35–39	5.6 (3.5, 8.2)	8.2 (2.7, 25)	48.2 (−67.1, 616.5)
40–44	13.3 (8.6, 19.2)	19 (6.4, 56.3)	43 (−66.7, 551.7)
45–49	22.4 (15.2, 32)	28.6 (15.3, 53.5)	27.9 (−52.2, 251.5)
50–54	51.7 (35.5, 73.1)	56.5 (40.1, 79.7)	9.3 (−45.1, 124.6)
55–59	44.9 (30.9, 61.7)	49 (34.3, 69.9)	9 (−44.4, 126.3)
60–64	41.7 (29.4, 59.2)	53.9 (26.2, 110.7)	29.1 (−55.7, 276)
65–69	53.7 (37.1, 75.6)	51.2 (52.9, 49.6)	−4.6 (−30, 33.6)
70–74	68.2 (48.2, 93.5)	62 (73.4, 52.4)	−9.1 (−21.5, 8.8)
75–79	57.9 (39.5, 83.9)	53.9 (53.6, 54.2)	−6.9 (−36.1, 37.1)
80–84	65.9 (43.9, 95.3)	53.2 (90, 31.4)	−19.3 (−5.6, −28.4)

ambient air pollution, and air pollution due to the burning of solid fuels in houses (42).

Breast cancer treatment depends upon molecular subtypes which are generally characterized on the basis of the status of estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2), breast cancer which is negative in ER, PR, and HER2 is termed as triple-negative breast cancer (TNBC) (43). TNBC is generally more aggressive and has a poor prognosis (44). Previous studies have shown that patients with TNBC in Pakistan, China, and India are generally of younger age and present with large invasive tumors, high node positivity, and high histological grade (45–47). Late-disease presentation is another common issue in Asian patients with breast cancer, which results in poor prognosis (48). Among cultural factors, in several Asian countries, females tend to hide tumor formation on breast from family and husband, which also delays the diagnosis leading to poor prognoses (1).

The mortality trend from breast cancer is predicted to increase in the next decade compared with the past decades in both East and South Asian countries. Japan officially recommended in 2004 that women should start mammographic screening at the age of 40 years, and once biannually after that (49). However, there is no clear guideline for the termination of the age of screening, and elderly women do not pay enough attention to breast screening and self-examination, which also contributes to the high mortality rate. Clinical breast examinations, as well as mammograms, are widely available in public primary health care institutions in Japan, which may also account for the high incidence rates of breast cancer along with lower mortality rates and higher 5-year survival rates in Japan (49).

The South Asian region demonstrated increasing mortality trends for all its countries with Pakistan, India, Nepal, and Afghanistan demonstrating a steady upward trend in mortality in the next decade. The population aging is a non-modifiable risk factor for developing breast cancer leading to mortality in Asian women (50–52). Correspondingly, this study shows that mortality is predicted to increase with age, particularly among 80–84 years old in all the East and South Asian countries. Previous studies have also identified mortality risk increases with age with a peak of breast cancer-related mortality at approximately 85 years (1, 4, 24, 53).

Most of the developed and developing countries of Asia are increasingly adopting a Westernized lifestyle, such as a diet high in fat and processed meat, low in fruits and vegetables, and lack of (or low) physical activity leading to obesity, which is linked with breast cancer apart from several other health risks of obesity (54). On average, obesity increases breast cancer-related mortality risk by ∼30–40% (55–58). In this study, breast cancer mortality trends related to high BMI, high FPG, alcohol use, and a diet high in red meat are also predicted for the period 2020–2030 among females of age 50–84 years. The high BMI- and high FPG-related breast cancer mortality rates were high in both the Asian regions and country-wise, these were the highest in Pakistan and Afghanistan, whereas alcohol-related mortality rate was highest in Japan, Mongolia, South Korea, and Nepal.

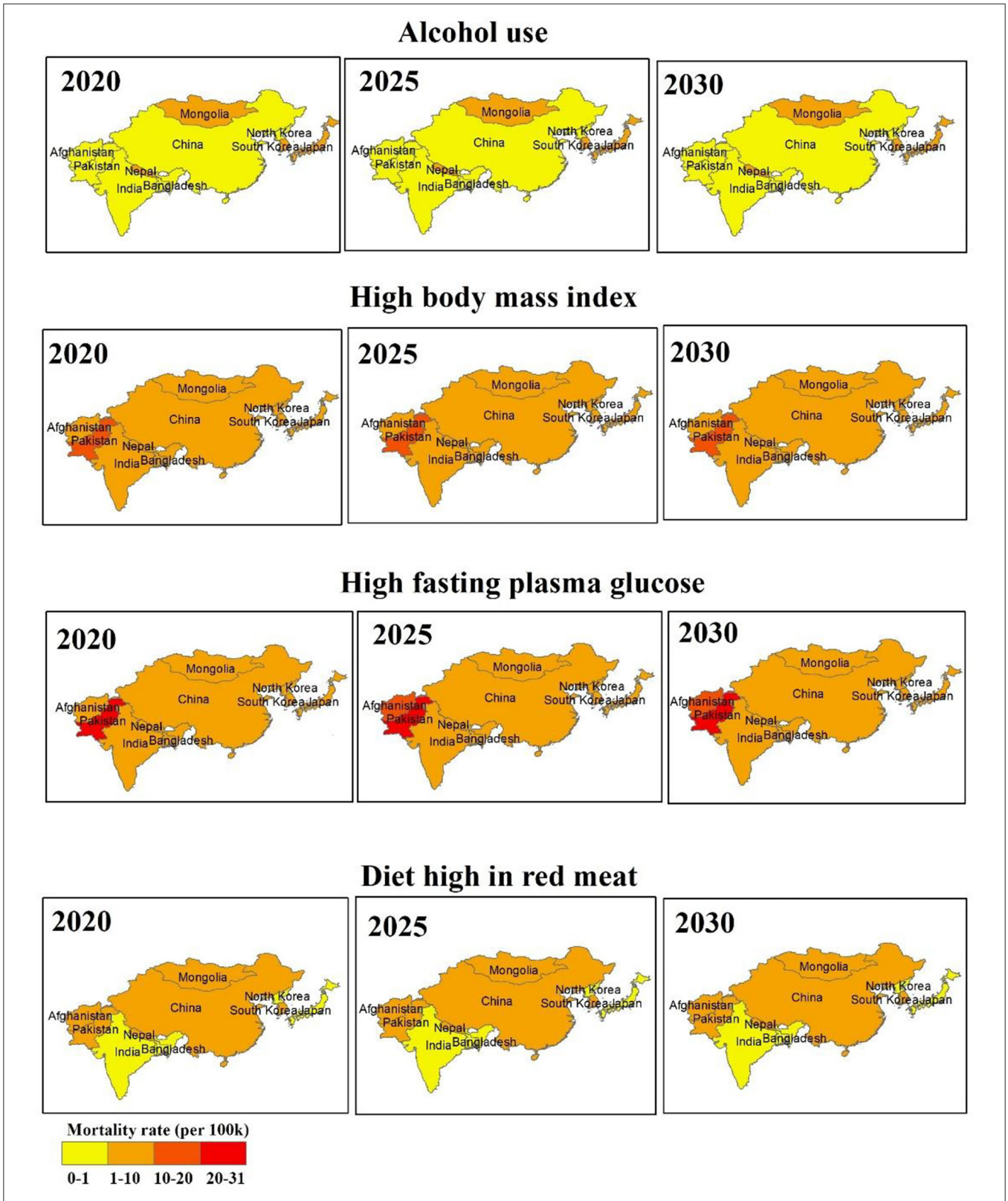


FIGURE 3 | Regional disparity of mean predicted risk attributable mortality rates (per 100 k) for the female of aged 50–84 years old, for the year 2020, 2025, and 2030.

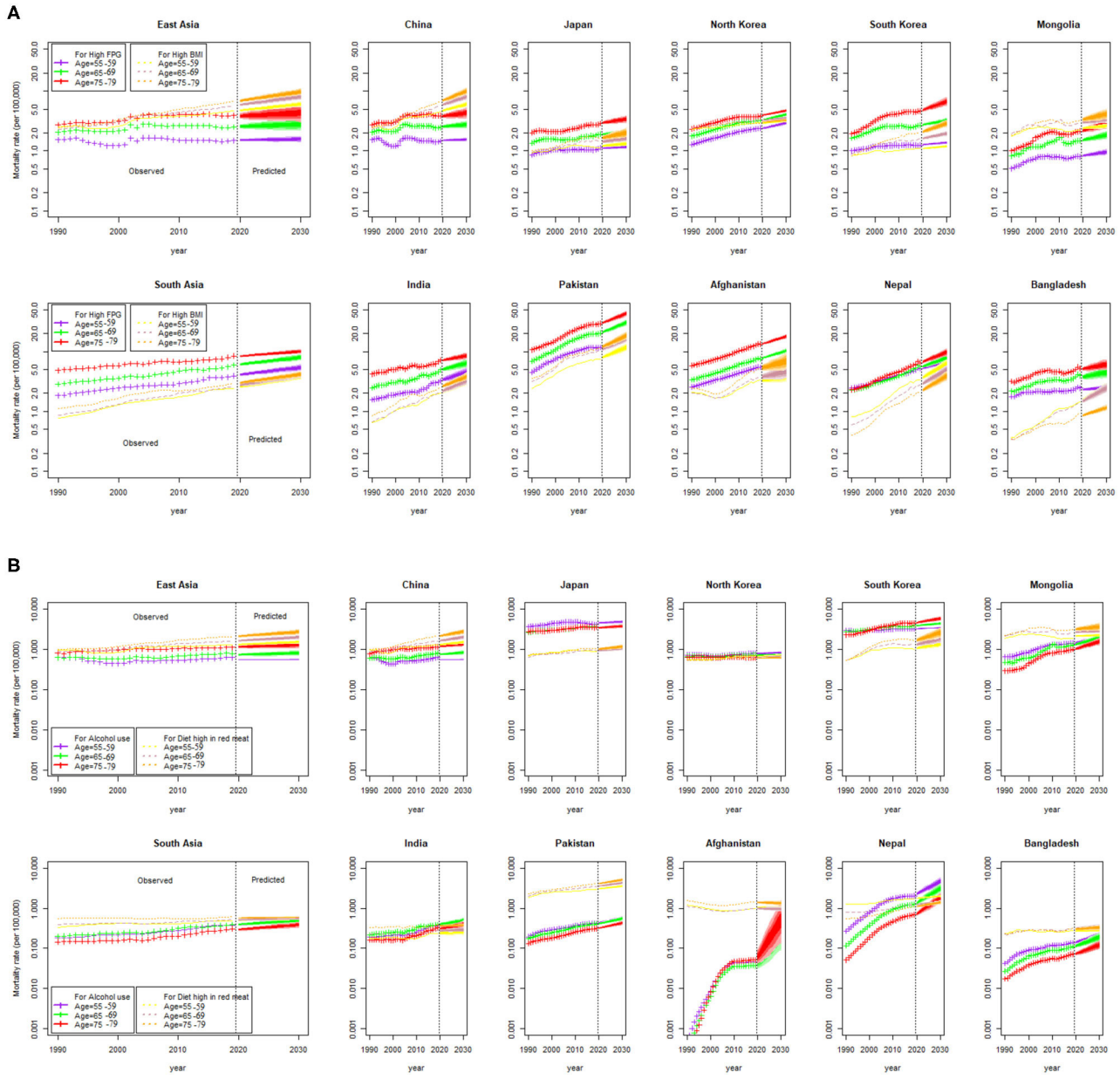


FIGURE 4 | Age-specific risk attributable breast cancer mortality rate (per 100,000) from 1990 to 2030, at a regional and national level. Shades in the fan represent prediction intervals at 80% and 90% level, **(A)** Predicted mortality rates attributable to high fasting plasma glucose (FPG) and high body mass index (BMI), and **(B)** Predicted mortality rates attributable to alcohol use and for a diet high in red meat.

Among 10 countries, six countries (China, Mongolia, South Korea, Pakistan, Afghanistan, and Nepal) have shown a high-predicted mortality rate attributable to a diet high in red meat. Another study has shown that higher red meat consumption, especially processed red meat, is associated with an increased risk of breast cancer-related mortality, which has a relationship between red meat consumption and mortality risk (59). A meta-analysis of prospective studies investigated the association between red meat intake and risk of breast cancer and has shown that the pooled relative risk of breast cancer for the highest vs. the lowest categories of red meat intake were 1.10 (95% CI, 1.02–1.16) and each additional intake of 120 g of red meat per day was associated with 11% additional risk of breast cancer (59). The potential biological mechanisms include heme iron and non-heme iron pro-oxidant activity and the carcinogenic by-products compounds produced in the

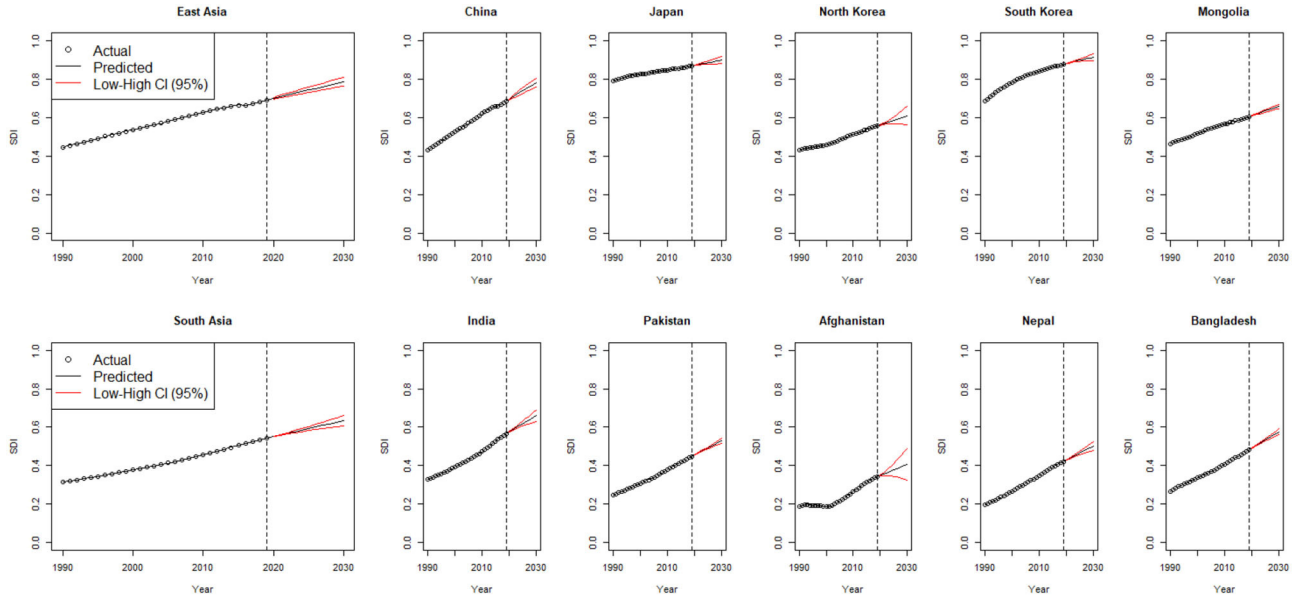


FIGURE 5 | Observed, estimated, and predicted trends in SDI from 1990 to 2030, by region and countries; SDI, sociodemographic index; CI, confidence interval.

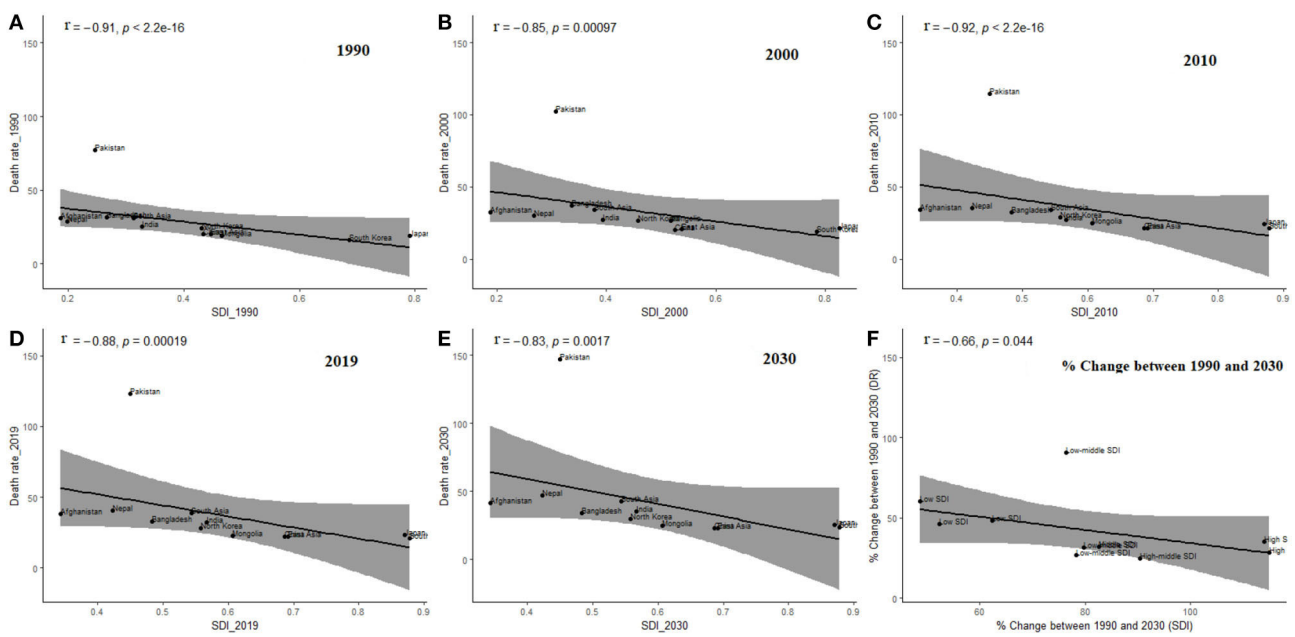


FIGURE 6 | Relationship between death rate (DR) and sociodemographic index (SDI) within Asian region including East and South Asia, in different prediction period, **(A)** 1990, **(B)** 2000, **(C)** 2010, **(D)** 2019, **(E)** 2030, **(F)** % Change between 1990 and 2030; The black dots represent the breast cancer related DR of each country, the black lines are the fitting lines of correlation trend, and grey shaded area is the 95% confidence interval for fitting results. DR, mean death rate (per 100k) from age 20 to 84 years.

high-temperature cooking of red meat (60). The consumption of red meat has increased tremendously in the East Asian countries in the last few decades and is expected to be the region with one of the highest red meat consumption by 2030 (61).

This study predicted life expectancy trends among patients with breast cancer of age 50–84 years. We observed that the life expectancy of a female patient with breast cancer in East Asia, China, and South Korea, might increase while for other countries, it is expected to decrease or remain

constant during the prediction period 2020–2030. Increases in life expectancy might be due to changing lifestyles, provision of better hospitalization, and fulfillment of required nutritious standards in these countries (62). In addition, population-based breast cancer screening programs are also available in some of the East Asian countries, include China and South Korea (63–66).

Our research results are estimated by integrating data from multiple sources in the GBD network using a constantly updated modeling process; there are, however, few limitations in this study. First, although GBD estimates of the breast cancer burden are based on multiple sources of data, due to lack of quality PBCRs in low-resource countries of East and South Asia, the estimates are produced using spatiotemporal modeling, because of which, these estimates may be biased. Moreover, the uncertainty intervals may be too wide for a few countries. Second, we have described only the overall trends for each country, but did not describe the subadministrative regions of each country. Third, the estimates may also be biased because of the lack of diagnostic facilities, or healthcare workers expertise in proper stage and grade detection of tumors. Therefore, these estimates cannot be considered a substitute for data collection using quality PBCRs. Moreover, we predicted the trends by modeling the age-period pattern in breast cancer mortality, which can be further confirmed by incorporating the cohort effects using other stochastic mortality methods. Though statistical findings in this study may differ from the actual situation, it provides a rough estimate of disease diagnosis and possible detrimental health effects. These estimates call for policymakers and governments' attention across these regions and countries to take appropriate measures to combat the disease burden.

CONCLUSION

This study predicted the trends of breast cancer-related mortality in East and South Asian countries over 40 years from 1990 to 2030, showing that breast cancer-related mortality has been increasing and is expected to remain a significant public health challenge in most of these Asian countries in the next decade. Several breast cancer deaths are attributable to multiple modifiable risk factors, mainly dietary and metabolic risk factors. The increase in breast cancer-related mortality in countries with weaker public health systems and poor awareness levels may result in the reduction of life expectancy among patients with breast cancer. Therefore, improving early detection, increasing breast cancer awareness and cost-effective treatments are warranted to reduce breast cancer mortality in these Asian countries. Moreover, systematic and accurate epidemiological data collection through cancer registries is obligatory to observe the trends and assess the efficacy of respective interventions, which require collaborations between national and international organizations. Cancer awareness and community health workers education can significantly boost early detection, control, and prevention in low- and middle-income Asian countries.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: Institute for Health Metrics and Evaluation (IHME): http://ghdx.healthdata.org/gbd-results-tool.

AUTHOR CONTRIBUTIONS

SM: conceptualization, data curation, formal analysis, methodology, software, validation, visualization, writing—original draft, and writing—review and editing. RS: software, validation, visualization, and writing—review and editing. SRH: software, validation, visualization, and investigation. MI and XL: data curation, validation, visualization, and investigation. N: methodology, validation, visualization, and investigation. CY: investigation, resources, validation, funding acquisition, project administration, and supervision. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the National Key Research and Development Program of China (Grant No. 2018YFC1315302) and the National Natural Science Foundation of China (Grant Nos. 82173626 and 81773552).

ACKNOWLEDGMENTS

We would like to thank the funding foundations for their support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnut.2022.847920/full#supplementary-material

REFERENCES

1. Sharma R. Examination of incidence, mortality and disability-adjusted life years and risk factors of breast cancer in 49 Asian countries, 1990–2019: estimates from global burden of disease study 2019. *Jpn J Clin Oncol.* (2021) 51:826–35. doi: 10.1093/jjco/hyab004
2. Fan L, Goss PE, Strasser-Weippl K. Current status and future projections of breast cancer in Asia. *Breast Care.* (2015) 10:372–8. doi: 10.1159/000441818
3. Kamangar, F. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J Clin Oncol.* (2006) 24:2137–50. doi: 10.1200/JCO.2005.05.2308
4. Liu H, Shi W, Jin Z, Zhuo R, Dong J, Lao Q, et al. Global, regional, and national mortality trends of female breast cancer by risk factor, 1990–2017. *BMC Cancer.* (2021) 21:459. doi: 10.1186/s12885-021-08217-5
5. Ng CJ, Teo CH, Abdullah N, Tan WP, Tan HM. Relationships between cancer pattern, country income and geographical region in Asia. *BMC Cancer.* (2015) 15:613. doi: 10.1186/s12885-015-1615-0
6. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and

mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2021) 71:209–49. doi: 10.3322/caac.21660
7. Schlipköter U, Flahault A. Communicable diseases: achievements and challenges for public health. *Public Health Rev.* (2010) 32:90–119. doi: 10.1007/BF03391594
8. James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. *Lancet.* (2018) 392:1789–858. doi: 10.1016/S0140-6736(18)32279-7
9. Li N, Deng Y, Zhou L, Tian T, Yang S, Wu Y, et al. Global burden of breast cancer and attributable risk factors in 195 countries and territories, from 1990 to 2017: results from the global burden of disease study 2017. *J Hematol Oncol.* (2019) 12:140. doi: 10.1186/s13045-019-0828-0
10. Anand P, Kunnumakara AB, Sundaram C, Harikumar KB, Tharakan ST, Lai OS, et al. Cancer is a preventable disease that requires major lifestyle changes. *Pharm Res.* (2008) 25:2097–116. doi: 10.1007/s11095-008-9661-9
11. World Health Organization. *International Classification of Diseases, 10th Revision (ICD-10)*. Available online at: https://www.who.int/standards/classifications/classification-of-diseases (accessed June 10, 2021).
12. *Institute for Health Metrics and Evaluation Cancer Registry Database*. Available online at: http://ghdxhealthdataorg/about-ghdx/our-information-sources (accessed March 2, 2021).
13. Wang H, Abbas KM, Abbasifard M, Abbasi-Kangevari M, Abbastabar H, Abd-Allah F, et al. Global age-sex-specific fertility, mortality, healthy life expectancy (HALE), and population estimates in 204 countries and territories, 1950–2019: a comprehensive demographic analysis for the global burden of disease study 2019. *Lancet.* (2020) 396:1160–203. doi: 10.1016/S0140-6736(20)30977-6
14. Roth GA, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the global burden of disease study 2017. *Lancet.* (2018) 392:1736–88. doi: 10.1016/S0140-6736(18)32203-7
15. Kyu HH, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. *Lancet.* (2018) 392:1859–922. doi: 10.1016/S0140-6736(18)32335-3
16. Foreman KJ, Lozano R, Lopez AD, Murray CJ. Modeling causes of death: an integrated approach using CODEm. *Popul Health Metrics.* (2012) 10:1. doi: 10.1186/1478-7954-10-1
17. Wang H, Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, et al. Global, regional, and national under-5 mortality, adult mortality, age-specific mortality, and life expectancy, 1970–2016: a systematic analysis for the global burden of disease study 2016. *Lancet.* (2017) 390:1084–50. doi: 10.1016/S0140-6736(17)31833-0
18. Hu K, Ding P, Wu Y, Tian W, Pan T, Zhang S. Global patterns and trends in the breast cancer incidence and mortality according to sociodemographic indices: an observational study based on the global burden of diseases. *BMJ Open.* (2019) 9:e028461. doi: 10.1136/bmjopen-2018-028461
19. Murray CJ, Lopez AD. On the comparable quantification of health risks: lessons from the global burden of disease study. *Epidemiology Baltimore.* (1999) 10:594–605. doi: 10.1097/00001648-199909000-00029
20. Gakidou E, Afshin A, Abajobir AA, Abate KH, Abbafati C, Abbas KM, et al. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: a systematic analysis for the global burden of disease study 2016. *Lancet.* (2017) 390:1345–422. doi: 10.1016/S0140-6736(17)32366-8
21. Girosi F, King G. *Understanding the Lee-Carter Mortality Forecasting Method*. Gking Harvard Eduation (2007). Available online at: https://tinyurl.com/y2vdglrj
22. Lyu P, Waegenaere AD, Melenberg B. A Multi-population approach to forecasting all-cause mortality using cause-of-death mortality data. *North Am Actu J.* (2020) 25:S421–56. doi: 10.1080/10920277.2019.1662316
23. Mubarik S, Hu Y, Yu C. A multi-country comparison of stochastic models of breast cancer mortality with P-splines smoothing approach. *BMC Med Res Methodol.* (2020) 20:299. doi: 10.1186/s12874-020-01187-5
24. Mubarik S, Wang F, Fawad M, Wang Y, Ahmad I, Yu C. Trends and projections in breast cancer mortality among four Asian countries (1990–2017): evidence from five stochastic mortality models. *Sci Rep.* (2020) 10:5480. doi: 10.1038/s41598-020-62393-1
25. Booth H, Hyndman RJ, Tickle L, De Jong P. Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demogr Res.* (2006) 15:289–310. doi: 10.4054/DemRes.2006.15.9
26. D'Amato V, Piscopo G, Russolillo M. The mortality of the italian population: smoothing techniques on the lee—carter model. *Ann Appl Stat.* (2011) 5:705–24. doi: 10.1214/10-AOAS394
27. Nelson BK. Time series analysis using autoregressive integrated moving average (ARIMA) models. *Accad Emerg Med.* (1998) 5:739–44. doi: 10.1111/j.1553-2712.1998.tb02493.x
28. Asteriou D, Hall SG. ARIMA models and the Box–Jenkins methodology. *Appl Econometr.* (2011) 2:265–86. doi: 10.1057/978-1-137-41547-9_13
29. Butt Z, Haberman S, Shang HL. *The Ilc Package*.
30. Hyndman RJ, Khandakar Y. Automatic time series forecasting: the forecast package for R. *J Stat Softw.* (2008) 27:1–22. doi: 10.18637/jss.v027.i03
31. Ahmad OB, Boschi-Pinto C, Lopez AD, Murray CJ, Lozano R, Inoue M. *Age Standardization of Rates: A New WHO Standard*. Geneva: World Health Organization (2001).
32. Li N, Wu P, Shen Y, Yang C, Zhang L, Chen Y, et al. Predictions of mortality related to four major cancers in China, 2020 to 2030. *Cancer Commun.* (2021) 41:404–13. doi: 10.1002/cac2.12143
33. Booth H, Maindonald J, Smith L. Applying Lee-Carter under conditions of variable mortality decline. *Popul Stud.* (2002) 56:325–36. doi: 10.1080/00324720215935
34. Wang Q, Li J, Zheng S, Li JY, Pang Y, Huang R, et al. Breast cancer stage at diagnosis and area-based socioeconomic status: a multicenter 10-year retrospective clinical epidemiological study in China. *BMC Cancer.* (2012) 12:122. doi: 10.1186/1471-2407-12-122
35. Xue QY. Socioeconomic disparities in breast cancer survival: relation to stage at diagnosis, treatment and race. *BMC Cancer.* (2009) 9:364. doi: 10.1186/1471-2407-9-364
36. Mohaghegh P, Yavari P, Akbari ME, Abadi A, Shormeij Z. Relationships between family levels of socioeconomic status and distribution of breast cancer risk factors. *Iran J Cancer Prev.* (2015) 8:53–9.
37. Kwate N. Re: Race, socioeconomic status, and breast cancer treatment and survival. *J Natl Cancer Instit.* (2002) 94:1254; author reply doi: 10.1093/jnci/94.16.1254
38. Wei W, Zeng H, Zheng R, Zhang S, He J. Cancer registration in China and its role in cancer prevention and control. *Lancet Oncol.* (2020) 21:e342–9. doi: 10.1016/S1470-2045(20)30073-5
39. Hai-Rim Shin, Marie Clem Carlos, Varghese C. Cancer control in the Asia Pacific region: current status and concerns. *Jpn J Clin Oncol.* (2012) 42:867–81. doi: 10.1093/jjco/hys077
40. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Willumsen JF. World health organization 2020 guidelines on physical activity and sedentary behaviour. *Br J Sports Med.* (2020) 54:1451–62. doi: 10.1136/bjsports-2020-102955
41. Suzuki Y, Tsuno Da H, Kimura T, Yamauchi H. BMI change and abdominal circumference are risk factors for breast cancer, even in Asian women. *Breast Cancer Res Treat.* (2017) 166:919–25. doi: 10.1007/s10549-017-4481-4
42. Danaei G, Hoorn SV, Lopez AD, Murray CJ, Ezzati M. Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. *Lancet.* (2005) 366:1784–93. doi: 10.1016/S0140-6736(05)67725-2
43. Perou CM, Sørlie T, Eisen MB, Van De Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature.* (2000) 406:747–52. doi: 10.1038/35021093
44. Wang C, Kar S, Lai X, Cai W, Arfuso F, Sethi G, et al. Triple negative breast cancer in Asia: an insider's view. *Cancer Treat Rev.* (2018) 62:29–38. doi: 10.1016/j.ctrv.2017.10.014
45. Sajid MT, Ahmed M, Azhar M, Mustafa Q, Shukr I, Kamal Z. Age-related frequency of triple negative breast cancer in women. *J Coll Physicians Surg Pak.* (2014) 24:400–3.

46. Sharma M, Sharma JD, Sarma A, Ahmed S, Kataki AC, Saxena R, et al. Triple negative breast cancer in people of North East India: critical insights gained at a regional cancer centre. *Asian Pac J Cancer Prevent.* (2014) 15:4507–11. doi: 10.7314/APJCP.2014.15.11.4507
47. Li C-Y, Zhang S, Zhang X-B, Wang P, Hou G-F, Zhang J. Clinicopathological and prognostic characteristics of triple-negative breast cancer (TNBC) in Chinese patients: a retrospective study. *Asian Pac J Cancer Prevent.* (2013) 14:3779–84. doi: 10.7314/APJCP.2013.14.6.3779
48. Yip C. *Breast cancer in Asia: Methods in Molecular Biology*. Clifton, NJ; Totowa, NJ: Human Press (2009). doi: 10.1007/978-1-59745-416-2_3
49. Uchida K, Ohashi H, Kinoshita S, Nogi H, Kato K, Toriumi Y, et al. Breast cancer screening and the changing population pyramid of Japan. *Breast Cancer.* (2015) 22:172–6. doi: 10.1007/s12282-013-0470-6
50. Kotepui M, Piwkham D, Chupeerach C, Songsri A, Charoenkijkajorn L. Epidemiology and histopathology of benign breast diseases and breast cancer in southern Thailand. *Euro J Gynaecol Oncol.* (2014) 35:670–5.
51. Chen JJ, Yan W, Xue JY, Chen Y, Chen YL, Qin X, et al. A clinicopathological study of early-stage synchronous bilateral breast cancer: a retrospective evaluation and prospective validation of potential risk factors. *PLoS ONE.* (2014) 9:e95185. doi: 10.1371/journal.pone.0095185
52. Virani S, Sriplung H, Rozek LS, Meza R. Escalating burden of breast cancer in southern Thailand: analysis of 1990–2010 incidence and prediction of future trends. *Cancer Epidemiol.* (2014) 38:235–43. doi: 10.1016/j.canep.2014.03.003
53. Mubarik S, Yu Y, Wang F, Malik SS, Liu X, Fawad M, et al. Epidemiological and sociodemographic transitions of female breast cancer incidence, death, case fatality and DALYs in 21 world regions and globally, from 1990 to 2017: an age-period-cohort analysis. *J Adv Res.* (2021). doi: 10.1016/j.jare.2021.07.012
54. Popkin BM. Global nutrition dynamics: the world is shifting rapidly toward a diet linked with noncommunicable diseases. *Am J Clin Nutr.* (2006) 2:289–98. doi: 10.1093/ajcn/84.2.289
55. Tapan TK, Iyigun ZE, Ilgun S, Ozmen V. Evaluation of the eating habits of breast cancer patients. *Pak J Med Sci.* (2020) 36:2368. doi: 10.12669/pjms.36.7.2368
56. Vance V, Mourtzakis M, McCargar L, Hanning R. Weight gain in breast cancer survivors: prevalence, pattern and health consequences. *Obes Rev.* (2011) 12:282–94. doi: 10.1111/j.1467-789X.2010.00805.x
57. Mubarik S, Liu X, Malik SS, Wang L, Yu Y, Yu C. Evaluation of lifestyle risk factor differences in global patterns of breast cancer mortality and DALYs during 1990–2017 using hierarchical age-period-cohort analysis. *Environ Sci Pollut Res.* (2021) 28:49864–76. doi: 10.1007/s11356-021-14165-1
58. Mubarik S, Wang F, Malik SS, Shi F, Wang Y, Nawsherwan, et al. A hierarchical age–period–cohort analysis of breast cancer mortality and disability adjusted life years (1990–2015) attributable to modified risk factors among Chinese women. *Int J Environ Res Public Health.* (2020) 17:1367. doi: 10.3390/ijerph17041367
59. Guo J, Wei W, Zhan L. Red and processed meat intake and risk of breast cancer: a meta-analysis of prospective studies. *Breast Cancer Res Treat.* (2015) 151:191–8. doi: 10.1007/s10549-015-3380-9
60. Zhang Y, Tian X, Jiao Y, Liu Q, Li R, Wang W. An out of box thinking: the changes of iron-porphyrin during meat processing and gastrointestinal tract and some methods for reducing its potential health hazard. *Crit Rev Food Sci Nutr.* (2021) 13:1–16. doi: 10.1080/10408398.2021.1963946
61. Nam K-C, Jo C, Lee M. Meat products and consumption culture in the East. *Meat Sci.* (2010) 86:95–102. doi: 10.1016/j.meatsci.2010.04.026
62. Lofterød T, Frydenberg H, Flote V, Eggen AE, Thune I. Exploring the effects of lifestyle on breast cancer risk, age at diagnosis, and survival: the EBBA-Life study. *Breast Cancer Res Treat.* (2020) 182:215–27. doi: 10.1007/s10549-020-05679-2
63. Song QK, Wang XL, Zhou XN, Yang HB, Li YC, Wu JP, et al. Breast cancer challenges and screening in China: lessons from current registry data and population screening studies. *Oncologist.* (2015) 20:773–9. doi: 10.1634/theoncologist.2014-0351
64. Sun L, Legood R, Sadique Z, dos-Santos-Silva I, Yang L. Breast cancer screening programme in China: does one size fit all? A cost-effectiveness analysis based on a Markov model. *Lancet.* (2018) 392:S2. doi: 10.1016/S0140-6736(18)32631-X
65. Ma GX, Yin L, Gao W, Tan Y, Liu R, Fang C, et al. Workplace-based breast cancer screening intervention in china. *Cancer Epidemiology Biomark Prevent.* (2012) 21:358–67. doi: 10.1158/1055-9965.EPI-11-0915
66. Li T, Mello-Thoms C, Brennan PC. Descriptive epidemiology of breast cancer in China: incidence, mortality, survival and prevalence. *Breast Cancer Res Treat.* (2016) 159:395–406. doi: 10.1007/s10549-016-3947-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Mubarik, Sharma, Hussain, Iqbal, Nawsherwan, Liu and Yu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

View publication stats

arXiv:2203.12881v1 [cs.CL] 24 Mar 2022

Can Unsupervised Knowledge Transfer from Social Discussions Help Argument Mining?

Subhabrata Dutta*
Jadavpur University
subha0009@gmail.com

Jeevesh Juneja*
Delhi Technological University
creativityinczenyoga@gmail.com

Dipankar Das
Jadavpur University
dipankar.dipnil@gmail.com

Tanmoy Chakraborty
IIIT-Delhi
tanmoy@iiitd.ac.in

Abstract

Identifying argument components from unstructured texts and predicting the relationships expressed among them are two primary steps of argument mining. The intrinsic complexity of these tasks demands powerful learning models. While pretrained Transformer-based Language Models (LM) have been shown to provide state-of-the-art results over different NLP tasks, the scarcity of manually annotated data and the highly domain-dependent nature of argumentation restrict the capabilities of such models. In this work, we propose a novel transfer learning strategy to overcome these challenges. We utilize argumentation-rich social discussions from the *ChangeMyView* subreddit as a source of unsupervised, argumentative discourse-aware knowledge by finetuning pretrained LMs on a selectively masked language modeling task. Furthermore, we introduce a novel prompt-based strategy for inter-component relation prediction that compliments our proposed finetuning method while leveraging on the discourse context. Exhaustive experiments show the generalization capability of our method on these two tasks over within-domain as well as out-of-domain datasets, outperforming several existing and employed strong baselines.[1]

1 Introduction

Computational argument mining from texts is the fine-grained process of understanding opinion dynamics. In the most fundamental sense, argument understanding requires the identification of the opinions posed and justifications provided to support or falsify them. Generally, automated argument mining is a multi-stage pipeline identified with three general steps (Lippi and Torroni, 2015; Stab and Gurevych, 2017) – separating argumentative spans from non-argumentative ones, classifying argument components, and inducing a structure among them (support, attack, etc.). While different argumentation models define different taxonomies for argument components, popular approaches broadly categorize them as 'claims' and 'premises' (Stab and Gurevych, 2017; Egawa et al., 2019; Mayer et al., 2020). As these components are not necessarily aligned to sentence-level segments and can be reflected within clausal levels, the task of argument component identification requires a *token-level boundary detection* of components and *component type classification*.

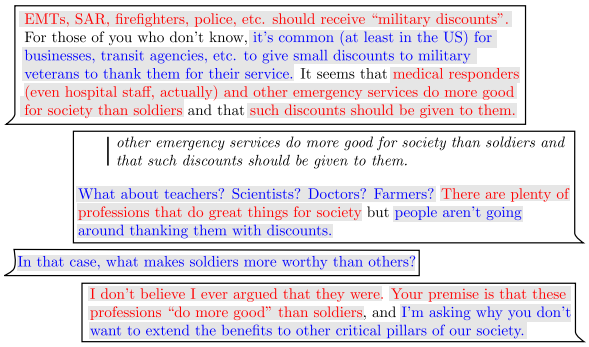


Figure 1: Token-level claim (red) and premise (blue) annotation of a discussion thread formed by consecutive posts from two users. Second post quotes a span from the first (shown in italics). Highlighted regions signify component boundaries (to demarcate consecutive components of the same kind as in the fourth post).

Context of argumentation in online discussions. Online discussions originating from back-and-forth posts from users reflect a rich interaction of opinion dynamics on large scale. In Figure 1, we show a sample argument component annotation of consecutive posts from two users. The token-level granularity of components ensures that a single sentence may contain multiple components of the same (in 1st post) or different kinds (in 2nd and 4th posts). Moreover, two adjacent spans of texts, even with the same argumentative role, can be defined as two separate components (see the 4th post for example). It is trivial to say that the meaning of any post (as well as its argumentative role) is de-

*Equal contribution

[1]We release all code, models and data used at https://github.com/Jeevesh8/arg_mining

pendent on the context. To be specific, the third post can be identified as argumentative (a premise in this case) only when its predecessor post and its components are taken as the context. Similarly, a certain span of the first post is quoted in the second one signaling a concrete manifestation of dialogic continuity. One may even observe the user-specific argumentation styles: 1st user (author of the first and third posts) usually keeps claims and premises in separate sentences, while the 2nd user prefers to use multi-component, complex sentences. Existing studies on argumentation formalism recognize such continuity and define inter-post component relations (Ghosh et al., 2014; Hidey et al., 2017). However, the previous approaches for automated extraction, classification and relating argumentative components work on individual posts only and define the inter-post discourse in the later stages of relation prediction.

This is trivially counter-intuitive for two major reasons: (i) if we consider two text spans from separate comments to be linked by some argumentative relation, then there exists a continuity of discourse between these spans and a model is likely to benefit if it decides the boundaries and types of these two components conditioned on that continuous information; (ii) users carry their style of argumentation (simple consecutive sentences vs. long complex ones, usage of particular markers like '*I think that*' etc.), and if the model is informed about these while observing the complete conversation with back-and-forth posts, it is more likely to extract correct components easily.

Scarcity of labeled data. Irrespective of the domain, argument annotation is a resource-intensive process. A few previous studies (Habernal and Gurevych, 2015; Al-Khatib et al., 2016) attempted to exploit a large amount of unlabeled data in a semi-supervised fashion. However, such methods require the components to be defined at sentence-level (and thereby adding redundant spans into the predictions) as they perform some sentence similarity matching to generate pseudo-labels. Pretrained language models like BERT (Devlin et al., 2019) provide a workaround to handle the scarcity of task-specific annotated data. A parameter-intensive model is initially trained in a self-supervised manner on a large bulk of text; this pretraining enables the model to learn general language representation, which is then finetuned on task-specific labeled data. However, the amount of the latter still determines the expressive power of such models (Wang et al., 2020).

Present work. Considering these challenges, we formulate a novel transfer learning method using Transformer-based language models. We use large amount of unlabelled discussion threads from Reddit's *r/ChangeMyView* (CMV) community as the source of argumentative knowledge. Pretrained, Transformer-based language models are finetuned on this dataset using a Masked Language Modelling task. Instead of randomly masking tokens to predict, we select several markers in the text that are shown to signal argumentative discourse in previous works (Chakrabarty et al., 2019; Eckle-Kohler et al., 2015). The language models are then made to predict these markers in the MLM task, thereby learning to relate different components of text according to their role in the argumentation presented. We call this novel finetuning method Selective Masked Language Modeling (**sMLM**). Furthermore, to explore the role of context in argument mining, we use **sMLM** to finetune a post-level language model based on BERT (Devlin et al., 2019) and RoBERTa (Liu et al., 2019) and a thread-level language model based on Longformer (Beltagy et al., 2020). We present efficient incorporation of several Reddit-specific structural cues into the Longformer architecture. These finetuned language models are then used for two fundamental components of argument mining: token-level argument component identification (**ACI**) and inter-component relation type prediction (**RTP**). To further utilize the **sMLM**-based training of the language models, we propose a novel prompt-based approach to predict relations among argument components. We perform exhaustive experiments to explore the efficacy of our proposed methods for argument mining in both *in-domain* and *out-of-domain* benchmark datasets: manually annotated Reddit discussions and scientific papers. Our experiments show clear improvements achieved by our methods (0.59 and 0.69 F1 for ACI and RTP, respectively) over several state-of-the-art baselines.[2]

2 Related Work

A general overview of argument mining can be found in the survey articles by Lytos et al. (2019) and Lawrence and Reed (2019). In the current scope, we look into three major areas of research

[2]The source codes and datasets have been submitted separately.

in argument mining.

Argument component detection and classification. Previous studies have sought to address argument boundary detection and component type prediction either as separate, successive tasks in the pipeline (Stab and Gurevych, 2017) or jointly in a single computational pass (Eger et al., 2017). Studies also explored classical machine learning frameworks like SVM-HMM (Habernal and Gurevych, 2017), CRF (Stab and Gurevych, 2017), etc. with rich manual feature engineering. With the development of neural network-based algorithms, BiLSTM-CNN-CRF models emerged as a popular choice (Schulz et al., 2018; Eger et al., 2017; Chernodub et al., 2019). Very recently, large pretrained language models like BERT have also been utilized (Mayer et al., 2020; Chakrabarty et al., 2019).

Discourse markers for learning language representation. Similar to our **sMLM** finetuning strategy, Nie et al. (2019) proposed an unsupervised sentence representation learning strategy where a neural model is trained to predict the appropriate discourse marker connecting two input sentences. Using a set of 15 markers, they showed that such a finetuning can help models in downstream NLI tasks. Chakrabarty et al. (2019) used a distant supervision approach using a single marker *In my honest opinion* to finetune BERT on a large collection of *ChangeMyView* threads and then performed argument component classification. However, they did not deal with the component identification task and performed classification of already identified components at sentence-level. Opitz and Frank (2019) suggested that while identifying the relation between two components, these models often rely more on the context and not the content of the components; discourse markers present within the context provide strong signals for the relation prediction task.

Argument mining over Reddit. A few recent studies explored argumentation over Reddit. Hidey et al. (2017) proposed a two-tier annotation scheme of claim-premise components and their relations, defining five different semantic roles of premises, using *ChangeMyView* discussion data. Egawa et al. (2019) also analyzed semantic roles of argument components over *ChangeMyView* threads; however, their primary focus remained on the dynamics of persuasion, similar to Dutta et al. (2020).

3 Selective MLM finetuning of Pretrained Language Models

Though pretrained language models are developed to overcome the problem of small annotated data on different language processing tasks, they still require task-specific finetuning for better results (Wang et al., 2020). In the specific domain of argument mining, annotated data is scarce, and attempting to finetune a massive language model with very small training data comes with the risk of overfitting. Moreover, different datasets follow different strategies for annotation. We seek to devise a novel transfer learning strategy where a given Transformer-based pretrained language model is directed to focus on argumentative discourse using large-scale, unlabelled data. We choose the *ChangeMyView* (CMV) community as the source of this transfer for two specific reasons: (i) it provides us with a large, readily available resource of interactions strictly focused on debates around versatile topics, and (ii) discussions in CMV contain a mixture of dialogic continuity over successive turns along with elaborate argumentation presented in a single turn. We hypothesize that such a versatile combination of discourse can make the language model more generalizable over dialogic as well as monologic argument mining tasks.

3.1 Discourse structure of CMV

Discussion forums like Reddit facilitate users to begin a discussion with an initial post (*submissions*, in the case of Reddit) and then comments under that post to instantiate a discussion. Users may post a comment in reply to the submission as well as the already posted comments. A typical discussion over Reddit forms a tree-like structure rooted at the submission. Any path from the root to a leaf comment can be perceived as an independent dialogic discourse among two or multiple users; henceforth, we will call such paths as *threads*. Formally, a thread T is an ordered sequence $\{(u_i, P_j)|i, j \in \mathbb{N}, u_i \in U_T\}$, where P_j is a text object (a submission when $j = 1$ and a comment, otherwise), u_i is the author of P_j, and U_T is the set of all unique users engaged in the thread T. For brevity, we indicate P_j as a post in general.

The dialogic nature of discussions naturally assumes this context to be the whole thread T. However, if we consider any two successive posts P_j and P_{j+1} in T, they manifest the interests and styles of two separate participants along with the

discourse continuity of the overall thread, which must be distinguished within the definition of the context. To take into account the complete dialogic context of the thread, we represent a thread as a single contiguous sequence of tokens with each post P_j from user u_i being preceded by a special token $[\textbf{USER}\text{-}i]$ with $i \in \{0, \cdots, |U_T| - 1\}$, to encode which post is written by which user.

Reddit also offers users a *quoting* facility: users can quote a segment from the previous post (one to which they are replying) within their posts and emphasize that their opinions are specifically focused on that segment. We delimit such quoted segments with special tokens **[STARTQ]** and **[ENDQ]** in the quoting post to demarcate the dialogic discourse. Chakrabarty et al. (2019) also used quoting as signals for following premises. Additionally, we replace URLs with the special token **[URL]** to inform the presence of external references that often act as justifications of subjective opinions.

3.2 Selective MLM finetuning

Masked Language Modeling is a common strategy of training large language models; a certain fraction of the input tokens are masked and the model is trained to predict them, consequently learning a generalized language representation. Instead of randomly selecting tokens to mask, we select specific markers that might signal argumentative discourse. While the model is trained to predict these markers, it learns the roles and relationships of the text spans preceding and following them. Following the work by Eckle-Kohler et al. (2015), we select multiple markers signaling *Opinion*, *Causation*, *Rebuttal*, *Fact presentation*, *Assumption*, *Summary*, and some additional words, which serve multiple purposes depending on the context.

As shown in Figure 2, to predict the marker *I think* in the first post, the model needs to learn that the following text span "*that most Jewish people* $\cdots$" expresses the user's opinion on the topic. Similarly, in the second post, for the input segment "$\langle span_0 \rangle$ *So* $\langle span_1 \rangle$ *if* $\langle span_2 \rangle$", to correctly predict the masked markers as *So* and *if*, a language model needs to learn the fact that the truth value of the statement expressed in $\langle span_1 \rangle$ is conditioned upon $\langle span_2 \rangle$, and this dependence is inferred from $\langle span_0 \rangle$.

Effect of context sizes. CMV threads provide a natural segmentation of the discourse context into comment/post-level vs. thread-level. We seek to explore the effect of the context size at different modules of argument mining (i.e., argument component detection and relation type prediction). For this, we use our proposed selective MLM approach to finetune a pretrained RoBERTa/BERT-base model in the comment/post-level regime, and train Longformer models in the thread-level regime. Longformer uses sparse, global attention (i.e., some tokens attend to all the tokens in the input sequence) to capture the long-range dependencies. We use the special tokens indicating the users (c.f. Section 3.1) as the globally attending tokens for Longformer.

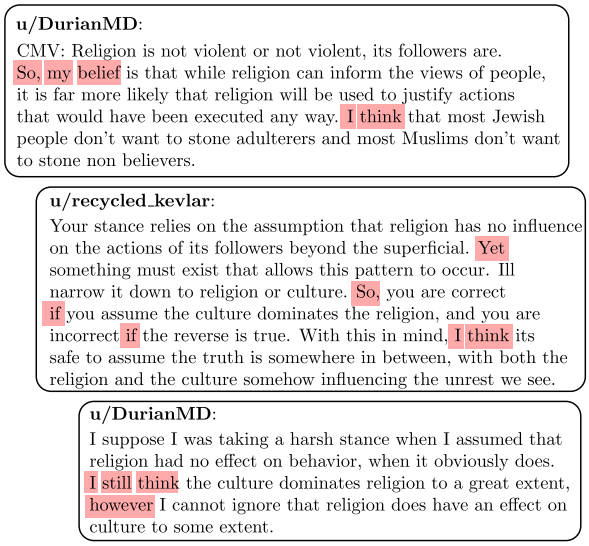


Figure 2: Example of selective masking in a sample CMV thread; **sMLM** finetuning requires a pretrained language model to predict the masked (highlighted in red) tokens (or all the subwords constituting them) based on the context.

3.3 Argument component identification

After finetuning the language model on the selective MLM task, we proceed to our first task of identifying argument components in threads. Since the detection is done in token-level, we use the standard **BIO** tagging scheme: for a component class $\langle type \rangle$, the beginning and the continuation of that component are marked as $\mathbf{B}\text{-}\langle type \rangle$ and $\mathbf{I}\text{-}\langle type \rangle$, respectively, while any non-component token is labeled as $\mathbf{O}$. Therefore, if one uses the usual claim-premise model of argumentation, the label set becomes $\{\mathbf{B}\text{-claim}, \mathbf{I}\text{-claim}, \mathbf{B}\text{-premise}, \mathbf{I}\text{-premise}, \mathbf{O}\}$.

3.4 Inter-component relation prediction

While identifying the relation between two given related argument components, it is important to understand the role of those text segments within the context of the discourse. Furthermore, we seek

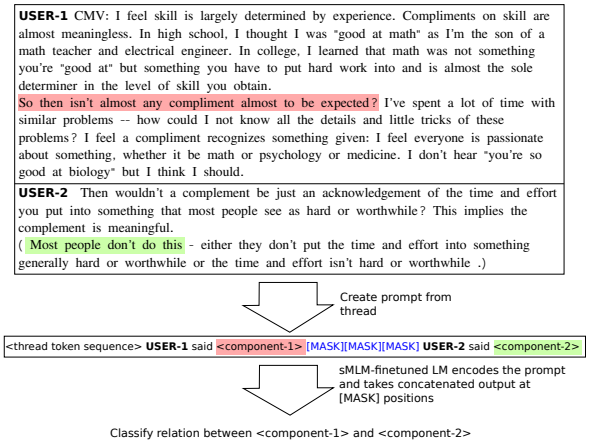


Figure 3: Example outline of prompt-based relation prediction where we seek to classify the relation between the claims posed by **USER-1** and **USER-2**, highlighted in red and green, respectively; the thread is converted to the prompt input by appending the prompt template. The language model the converts this prompt token sequence into fixed dimensional vectors from which the vector corresponding to the position of the masking token is used for relation classification.

to utilize the knowledge acquired by a language model in the **sMLM** finetuning step as well. Keeping these two factors in mind, we propose a novel, prompt-based identification of argument components. This approach is inspired by recent popularity of prompt-based fine-tuning methods in the community (Liu et al., 2021). At its core, these methods involve directly prompting the model for the required knowledge, rather than fine-tuning **[CLS]** or mean-pooled embeddings. For example, to directly use a model to summarise a text, we can append "**TL;DR:**" to the text (Radford et al., 2019), and let the model generate tokens following it; we expect the next few tokens to constitute a summary of all the previous text.

Since the underlying Transformer LMs have been trained using some Cloze task (i.e., filling the blanks from the context) previously, it is more natural for it to predict a token given a context. However, there are two challenges: (i) one needs to design a suitable prompt, and (ii) in case of classification tasks like **RTP**, it is challenging to perform **Answer Mapping**, i.e., to map all the possible tokens to some particular relation class. To tackle these challenges, we design our proposed relation prediction method in the following manner (see Figure 3)

For each pair of related components, say, component-1 and component-2, said by user-i and user-j, respectively, where component-2 refers to component-1, we append to the thread, a prompt with the template: "[USER-i] said <component1> [MASK] [MASK] [MASK] [USER-j] said <component2>" (we used three mask tokens since that is the upper bound of the marker size used for **sMLM**). We expect that the words predicted at the masked position such as "because", "in spite of what" etc. would be indicative of the relation of the two components. For the example thread shown in Figure 3, in a zero-shot prediction, **sMLM**-finetuned Longformer predicts "I", "disagree", "I" at the three masked positions. This "disagree" clearly corresponds to the **undercutter** relation between the two components. In fact, the base Longformer without **sMLM** finetuning predicts a space, a full stop and another space at the three masked positions. This additionally proves the efficacy of the **sMLM** finetuning.

Instead of engineering a token-to-relation type mapping, the predicted token embeddings at the masked positions are concatenated and fed into a linear layer to predict probabilities over the set of relation types. This way, we allow the model to learn and map from the token space to the relation type space.

4 Experiment Setup

4.1 Dataset

For the **sMLM** finetuning, we use the subset of *Winning Args (ChangeMyView)* (CMV) dataset (Tan et al., 2016) provided in ConvoKit (Chang et al., 2020). We use 99% of this data for training, and reserve 1% for checking accuracy on the **sMLM** task. The entire data consists of $3,051$ submissions and $293,297$ comments posted in the *ChangeMyView* subreddit by $34,911$ unique users. We extract the threads from these posts following the reply structure and end up with $120,031$ threads in total.

To train and evaluate all the models for **ACI** and **RTP**, we use the manually annotated Reddit discussion threads provided by Hidey et al. (2017) and further extended by Chakrabarty et al. (2019) for training and evaluation. The extended version of this dataset contains 113 CMV discussion threads manually annotated with argument components following the standard claim-premise model.

Additionally, we use the argument annotated **Dr. Inventor Corpus** (Lauscher et al., 2018) which consists of 40 scientific publications from the field of computer graphics. There are three types of argumentative components here: Background Claims

(**BC**), consisting of claims from previous works in the paper, Own Claim (**OC**) consisting of the new claims made by the authors of the paper, and **Data**. The Data class mainly consists of citations, references to figures, etc. This dataset has three relation types, viz., **support**, **contradicts** and **semantically same**. Additional dataset details are provided in Appendix A.

4.2 Baseline methods

For **ACI**, we consider two state-of-the-art token-level argument identification models: ▷ **LSTM-MTL.** Eger et al. (2017) proposed an end-to-end argument mining architecture which uses a BiLSTM-CNN-CRF sequence tagger to jointly learn component detection, classification, and relation parsing tasks. ▷ **LSTM-MData.** Schulz et al. (2018) proposed a BiLSTM-CNN-CRF based model which aims to generalize argument mining using multi-domain training data in an MTL setting. We augment our data with their original set of 6 datasets.

For **RTP**, as no prior work exists to the best of our knowledge, we consider our own baselines. First, we consider ▷ **Context-less RoBERTa**, a pretrained RoBERTa model, which takes the two components with a [SEP] token between them and predicts the relation using [CLS] token's embedding. It is context-less as only two components without the surrounding context are used to predict the label. Second, we consider ▷ **Contextless QR-Bert.** This uses the same fine-tuning methodology as **Contextless RoBERTa** and is initialized from the pre-trained Quote-Response relation prediction model of Chakrabarty et al. (2019).

For **RTP**, we try another traditional strategy, instead of prompting, for our models: ▷ **Mean Pooling**. The mean pooling approach first finds an embedding of each of the two related components by averaging the Transformer embeddings at all token positions within a component. These embeddings are concatenated and passed into a linear layer for predicting the type of relation between the two related components.

To further evaluate the efficacy of our **sMLM** training strategy, we finetune a pretrained Longformer on the Winning Args Corpus, with the usual MLM, i.e., masking 15% of tokens at random, instead of selective masking. We call this the domain-adapted Longformer, **DA-LF**.

Model	Claim			Premise			F1	Acc
	P	R	F1	P	R	F1		
sMLM-LF	0.49	0.57	**0.53**	**0.61**	**0.67**	**0.64**	**0.59**	**0.74**
Base-LF	**0.50**	0.50	0.50	0.58	0.64	0.61	0.56	**0.74**
sMLM-RoBERTa	0.49	**0.60**	**0.53**	0.55	0.57	0.55	0.55	0.72
RoBERTa	0.49	0.55	0.51	0.56	0.62	0.59	0.56	0.73
BERT	0.21	0.25	0.23	0.19	0.26	0.22	0.22	0.62
LSTM-MData	0.19	0.18	0.18	0.26	0.23	0.24	0.22	0.54
LSTM-MTL	0.19	0.18	0.18	0.24	0.25	0.24	0.21	–

Table 1: Performance of different models on **ACI**-task on CMV Modes dataset (P: Precision, R: Recall, F1: F1 score). The **F1** and **Acc.** in the rightmost columns denote the micro-averaged F1 score over claims and premises and the token level accuracy of predicting argumentative tags, respectively.

Model	BC			OC			Data		
	P	R	F1	P	R	F1	P	R	F1
sMLM-LF	0.45	**0.52**	0.48	**0.39**	0.45	0.42	**0.50**	**0.48**	**0.48**
Base-LF	**0.49**	0.51	**0.50**	0.38	**0.50**	**0.43**	0.44	0.44	0.44

Table 2: Results on Dr. Inventor dataset for argument component identification using **sMLM**-finetuned and base Longformer models.

4.3 Implementation details

We use the pretrained base version of Longformer (12 layers, 768 model size). The size of the local attention window was set to the default 512. The maximum sequence length was fixed at 4096.

Following the suggestions in Reimers and Gurevych (2017), we repeat our experiments on the 5 different data splits. The scores reported in the tables for various models correspond to the average value of the mean of 5 runs, over the last 5 epochs for that particular metric. We provide additional implementation details in Appendix B.

5 Evaluation

We evaluate the models based on precision, recall, and F1 scores for predicting claims and premises. For a more rigorous setting, we use exact match of the whole span between gold and predicted labels, i.e., if the gold label is [**O**, **B**-claim, **I**-claim, **I**-claim, **I**-claim, **O**] then only the predictions [**O**, **B**-claim, **I**-claim, **I**-claim, **I**-claim, **O**], or [**O**, **I**-claim, **I**-claim, **I**-claim, **I**-claim, **O**] can be considered as true positives. We use the popular SeqEval (Nakayama, 2018) framework.

5.1 Argument component identification

Table 1 shows the results for argument component identification on the CMV Modes dataset. We compare models based on their micro-averaged F1 over the two component types (claims, premises), and token level accuracy. Firstly, we observe huge difference in token-level accuracy scores as we move from the existing best performing LSTM based methods with accuracy of 0.54 to BERT,

Model	Support			Agreement			Direct Attack			Undercutter			Partial			Overall
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	F1
80-20 split																
sMLM-LF-prompt	**0.88**	**0.93**	**0.91**	**0.51**	**0.46**	**0.48**	**0.32**	**0.35**	**0.33**	**0.43**	**0.51**	**0.46**	0.28	0.12	0.16	**0.67**
DA-LF-prompt	0.78	0.84	0.81	0.44	0.45	0.43	0.22	0.19	0.19	0.30	0.32	0.30	0.27	0.11	0.15	0.61
sMLM-LF-mp	0.73	0.87	0.79	0.49	0.36	0.38	0.32	0.24	0.26	0.32	0.33	0.41	0.35	**0.21**	**0.25**	0.59
Base-LF-prompt	0.79	0.88	0.84	0.48	0.44	0.46	0.30	0.23	0.25	0.31	0.39	0.34	**0.37**	0.12	0.17	0.62
Base-LF-mp	0.71	0.87	0.78	0.47	0.33	0.37	0.24	0.17	0.18	0.27	0.26	0.26	0.35	0.20	0.24	0.56
RoBERTa	0.78	0.83	0.80	0.46	0.34	0.37	0.29	0.29	0.28	0.15	0.24	0.18	0.36	0.15	0.20	0.60
QR-Bert	0.76	0.85	0.80	0.46	0.27	0.34	0.21	0.13	0.16	0.19	0.25	0.20	0.32	0.16	0.20	0.59
50-50 split																
sMLM-LF-prompt	**0.87**	**0.92**	**0.89**	**0.53**	**0.47**	**0.49**	0.30	**0.28**	**0.28**	**0.45**	**0.58**	**0.50**	0.35	0.09	0.14	**0.69**
DA-LF-prompt	0.85	0.89	0.87	0.47	**0.47**	0.44	**0.32**	0.20	0.24	0.39	0.55	0.44	0.32	0.13	0.16	0.66
sMLM-LF-mp	0.70	0.90	0.79	0.426	0.22	0.28	0.28	0.20	0.22	0.32	0.26	0.28	0.38	**0.18**	0.24	0.56
Base-LF-prompt	0.78	0.87	0.82	0.49	0.44	0.46	0.30	0.19	0.22	0.32	0.40	0.35	0.32	0.13	0.18	0.62
Base-LF-mp	0.73	0.86	0.79	0.36	0.21	0.26	0.25	0.18	0.21	0.23	0.28	0.25	**0.43**	**0.18**	**0.25**	0.56
RoBERTa	0.72	0.83	0.77	0.47	0.25	0.31	0.22	0.21	0.21	0.13	0.16	0.14	0.17	0.08	0.10	0.55
QR-Bert	0.72	0.84	0.77	0.47	0.28	0.34	0.19	0.13	0.14	0.13	0.18	0.15	0.22	0.07	0.09	0.54

Table 3: Relation type wise Precision (P), Recall (R) and F1 score on the CMV Modes dataset for various models. The highest scores in every column are in **bold**. The suffix "mp" and "prompt" indicate that the model was trained using **Mean Pooling** and **Prompting** strategies, respectively. The **F1** in last column is the Micro/weighted-F1 over all the prediction classes.

Relation types	Base-LF-prompt			sMLM-LF-prompt		
	P	R	F1	P	R	F1
Support	**0.91**	0.90	**0.91**	0.89	**0.92**	**0.91**
Contradict	0.60	**0.60**	**0.60**	**0.65**	0.55	**0.60**
Semantically same	0.74	**0.77**	0.75	**0.77**	0.75	**0.77**

Table 4: Relation Type wise Precision (P), Recall (R) and F1 score on Dr. Inventor Corpus for prompt-based relation prediction using **sMLM** and base Longformer models.

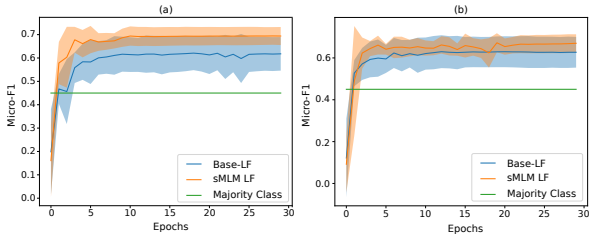


Figure 4: Micro-F1 scores for predicting relation types among argument components by Base and sMLM-finetuned Longformer models over the course of training using (a) 50-50 split and (b) 80-20 split. We use 5 different runs on random splits for each model to report the mean (solid lines) and variance.

having an accuracy of 0.62. Such a difference is expected since pretrained language models like BERT provide a head-start in case of small datasets like CMV Modes. Though the token-level accuracy increases, the micro-averaged F1 for exact component match does not increase much till we start using RoBERTa. Since pretrained Longformer was trained originally from the RoBERTa checkpoint (Beltagy et al., 2020), we can conclude that RoBERTa provides significant performance gain compared to BERT, owing to its larger training data and protocol. Longformer trained with our proposed **sMLM** finetuning clearly outperforms the rest of the models in terms of overall F1 score for component identification. However, the effects of selective MLM is more prominant in case of thread-level context (i.e, Longformer) compared to comment-level context (i.e, RoBERTa).

We can observe that context plays different roles for different component types: while **sMLM**-finetuned Longformer and RoBERTa perform comparably for claim detection, in case of premises, the access to the complete context helps the Longformer to perform better. We can observe a similar trend in **ACI**-task on Dr. Inventor dataset (see Table 2). While Base Longformer performs comparable to its **sMLM** counterpart to detect Background and Own Claims, **sMLM** provides a 4 point improvement in F1 score for the Data class which plays a similar role of premises towards the claims. Intuitively, textual segments expressing claims contain independent signals of opinion that is less dependent on the context; pretrained language models might be able to decipher their roles without additional information either from the thread-level context (in case of CMV Modes, specifically) or enhanced relation-awareness induced by the **sMLM** finetuning. However, identifying segments that serve the role of premises to a claim intrinsically depends on the claims as well as the discourse expressed in a larger context.

5.2 Relation type prediction

In Table 3, we present the results for relation type identification on the CMV Modes dataset. We again compare models based on their micro-averaged F1 over all relation types. Firstly, we consider the traditional mean pooling approach. Within this approach, we observe a 3 point improvement for the **sMLM** pre-trained Longformer on the 80-20 split, while maintaining same performance on the 50-50 split. Furthermore, the prompt based methods consistently outperform the mean pooling one, irrespective of whether we use base Longformer or **sMLM** pretrained one.

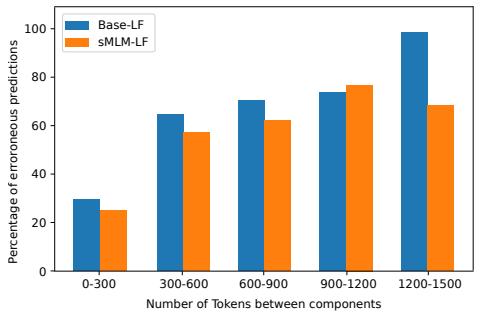


Figure 5: Percentage of erroneous classifications for **RTP** for Base-LF-prompt and LF-sMLM-prompt on component-pairs at different distances.

Within the prompting approach, we also observe increased and consistent improvement in performance due to **sMLM** pretraining on both 80-20 and 50-50 splits. The gap in micro-F1 scores between **sMLM** and base Longformer for 80-20 split increases from 3 points in mean pooling to 5 points in prompting (0 to 7 points improvements for 50-50 split). As we can observe in Figure 4, **sMLM**-finetuned Longformer admits a very narrow margin of variation on random splits, compared to the base Longformer. Furthermore, **sMLM** finetuning consistently outperforms domain-adapted finetuning (**DA-LF**), indicating the unique knowledge transfer achieved by the former.

We hypothesise that this approach works better as this regime models our final **RTP** task, as a task that is *more natural* (in a sense similar to the $(\tau, B)-$natural tasks of Saunshi et al. (2021)) for a Longformer model pre-trained with **sMLM**. Intuitively, the model learns to predict discourse markers at masked positions during **sMLM** pre-training and during fine-tuning on downstream tasks too, the model will naturally try to predict discourse markers at the masked positions. The discourse markers occurring at the masked positions are directly related to the relation between the two components. For instance, when there is a "but" between two components, we know that the two components present opposing views more or less. Here again, we observe that **sMLM** does not hurt the base performance under domain shift (Table 4).

We observe that the RoBERTa model performs worse than Base-LF-prompt, which incorporates the entire context of the thread. Also the effect worsens with reduced training set size, and RoBERTa model performs worse by 7 points in terms of micro-F1 for the 50-50 split. Furthermore, we observe that the mean pooling strategy, even though it uses context, performs worse (by 4 points on 80-20 split) than the context-less RoBERTa. Though, our **sMLM** pretrained model, manages to perform at par with the context-less RoBERTa with the mean pooling strategy. *This means, that the using the right fine-tuning method is essential. Extra context can be utilised fully in longformer, only when pre-training and fine-tuning tasks are nicely aligned.*

Model	Claim			Premise			F1
	P	R	F1	P	R	F1	
base-LF-near	0.39	0.59	0.47	0.63	0.52	0.57	0.52
base-LF-far	0.42	0.57	0.48	0.63	0.55	0.59	0.54
sMLM-LF-near	0.36	0.48	0.40	0.68	0.65	0.66	0.57
sMLM-LF-far	0.46	0.57	0.51	0.63	0.63	0.63	0.58

Table 5: Performance of base Longformer and **sMLM** Longformer for predicting segments having some markers in "**near**" (5 tokens on either side of its) boundaries, and the rest of segments ("**far**").

5.3 Dependence on the presence of markers

Following the analyses presented by Opitz and Frank (2019), we investigate whether the presence/absence of the markers used in the **sMLM** step within the vicinity of the components play any role in the **ACI** or **RTP** performances. Since the relation type among component-pairs that reside distant from each other are less likely to be inferred by the presence of markers in the context, we analyse the percentage of wrong predictions as we vary the distance between two related components, in Figure 5. While error rate does vary proportionally to the distance, we observe that **sMLM-LF** consistently yields lower percentage of wrong predictions as we vary the distance between the related components compared to base Longformer. This clearly indicates the superior capability induced by the **sMLM** finetuning to decipher the relationship among components not linked by direct context (i.e., not within a sentence or a single comment).

For the **ACI** task, however, we observe that the absence of markers in the vicinity of the components actually enables better identification, both in case of **sMLM** finetuned and pretrained Longformer (see Table 5).

6 Conclusion

We presented the results for two important tasks in the argument mining pipeline, viz., **ACI** and **RTP**. The experiments clearly elucidated the importance of alignment between the downstream and pre-trainig tasks, and the effect of various ways of modelling the tasks. The importance of entire thread's context in discussion forums, as well

as how to incorporate that into transformer-based models fruitfully has also been made clear.

Acknowledgements

The authors would like to thank Chris Hidey and Smaranda Muresan, for clarifications providing regarding their work. T. Chakraborty would like to acknowledge the support of Ramanujan Fellowship, CAI, IIIT-Delhi and ihub-Anubhuti-iiitd Foundation set up under the NM-ICPS scheme of the Department of Science and Technology, India.

References

Khalid Al-Khatib, Henning Wachsmuth, Matthias Hagen, Jonas Köhler, and Benno Stein. 2016. Cross-domain mining of argumentative text through distant supervision. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1395–1404, San Diego, California. Association for Computational Linguistics.

Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. Longformer: The long-document transformer. *CoRR*, abs/2004.05150v2.

Tuhin Chakrabarty, Christopher Hidey, Smaranda Muresan, Kathy McKeown, and Alyssa Hwang. 2019. AMPERSAND: Argument mining for PERSuAsive oNline discussions. 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)*, pages 2933–2943, Hong Kong, China. Association for Computational Linguistics.

Jonathan P. Chang, Caleb Chiam, Liye Fu, Andrew Z. Wang, Justine Zhang, and Cristian Danescu-Niculescu-Mizil. 2020. Convokit: A toolkit for the analysis of conversations.

Artem Chernodub, Oleksiy Oliynyk, Philipp Heidenreich, Alexander Bondarenko, Matthias Hagen, Chris Biemann, and Alexander Panchenko. 2019. TARGER: Neural argument mining at your fingertips. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 195–200, Florence, Italy. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. 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)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Subhabrata Dutta, Dipankar Das, and Tanmoy Chakraborty. 2020. Changing views: Persuasion modeling and argument extraction from online discussions. *Information Processing & Management*, 57(2):102085.

Judith Eckle-Kohler, Roland Kluge, and Iryna Gurevych. 2015. On the role of discourse markers for discriminating claims and premises in argumentative discourse. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2236–2242, Lisbon, Portugal. Association for Computational Linguistics.

Ryo Egawa, Gaku Morio, and Katsuhide Fujita. 2019. Annotating and analyzing semantic role of elementary units and relations in online persuasive arguments. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 422–428, Florence, Italy. Association for Computational Linguistics.

Steffen Eger, Johannes Daxenberger, and Iryna Gurevych. 2017. Neural end-to-end learning for computational argumentation mining. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11–22, Vancouver, Canada. Association for Computational Linguistics.

Matt Gardner, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson Liu, Matthew Peters, Michael Schmitz, and Luke Zettlemoyer. 2018. Allennlp: A deep semantic natural language processing platform.

Debanjan Ghosh, Smaranda Muresan, Nina Wacholder, Mark Aakhus, and Matthew Mitsui. 2014. Analyzing argumentative discourse units in online interactions. In *Proceedings of the First Workshop on Argumentation Mining*, pages 39–48, Baltimore, Maryland. Association for Computational Linguistics.

Ivan Habernal and Iryna Gurevych. 2015. Exploiting debate portals for semi-supervised argumentation mining in user-generated web discourse. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2127–2137, Lisbon, Portugal. Association for Computational Linguistics.

Ivan Habernal and Iryna Gurevych. 2017. Argumentation Mining in User-Generated Web Discourse. *Computational Linguistics*, 43(1):125–179.

Christopher Hidey, Elena Musi, Alyssa Hwang, Smaranda Muresan, and Kathy McKeown. 2017. Analyzing the semantic types of claims and premises in an online persuasive forum. In *Proceedings of the 4th Workshop on Argument Mining*, pages 11–21, Copenhagen, Denmark. Association for Computational Linguistics.

Anne Lauscher, Goran Glavaš, and Simone Paolo Ponzetto. 2018. An argument-annotated corpus of scientific publications. In *Proceedings of the 5th Workshop on Argument Mining*, pages 40–46, Brussels, Belgium. Association for Computational Linguistics.

John Lawrence and Chris Reed. 2019. Argument mining: A survey. *Computational Linguistics*, 45(4):765–818.

Marco Lippi and Paolo Torroni. 2015. Argument mining: A machine learning perspective. In *Theory and Applications of Formal Argumentation*, pages 163–176, Cham. Springer International Publishing.

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2021. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach.

Anastasios Lytos, Thomas Lagkas, Panagiotis Sarigiannidis, and Kalina Bontcheva. 2019. The evolution of argumentation mining: From models to social media and emerging tools. *Information Processing & Management*, 56(6):102055.

Tobias Mayer, Elena Cabrio, and Serena Villata. 2020. Transformer-based argument mining for healthcare applications. In *ECAI 2020 - 24th European Conference on Artificial Intelligence, 29 August-8 September 2020, Santiago de Compostela, Spain, August 29 - September 8, 2020 - Including 10th Conference on Prestigious Applications of Artificial Intelligence (PAIS 2020)*, volume 325 of *Frontiers in Artificial Intelligence and Applications*, pages 2108–2115. IOS Press.

Marius Mosbach, Maksym Andriushchenko, and Dietrich Klakow. 2021. On the stability of fine-tuning bert: Misconceptions, explanations, and strong baselines.

Hiroki Nakayama. 2018. seqeval: A python framework for sequence labeling evaluation. Software available from https://github.com/chakki-works/seqeval.

Allen Nie, Erin Bennett, and Noah Goodman. 2019. DisSent: Learning sentence representations from explicit discourse relations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4497–4510, Florence, Italy. Association for Computational Linguistics.

Juri Opitz and Anette Frank. 2019. Dissecting content and context in argumentative relation analysis. In *Proceedings of the 6th Workshop on Argument Mining*, pages 25–34, Florence, Italy. Association for Computational Linguistics.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.

Nils Reimers and Iryna Gurevych. 2017. Reporting score distributions makes a difference: Performance study of lstm-networks for sequence tagging.

Nikunj Saunshi, Sadhika Malladi, and Sanjeev Arora. 2021. A mathematical exploration of why language models help solve downstream tasks.

Claudia Schulz, Steffen Eger, Johannes Daxenberger, Tobias Kahse, and Iryna Gurevych. 2018. Multi-task learning for argumentation mining in low-resource settings. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 35–41, New Orleans, Louisiana. Association for Computational Linguistics.

Christian Stab and Iryna Gurevych. 2017. Parsing Argumentation Structures in Persuasive Essays. *Computational Linguistics*, 43(3):619–659.

Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, page 613–624, Republic and Canton of Geneva, CHE. International World Wide Web Conferences Steering Committee.

Yaqing Wang, Subhabrata Mukherjee, Haoda Chu, Yuancheng Tu, Ming Wu, Jing Gao, and Ahmed Hassan Awadallah. 2020. Adaptive self-training for few-shot neural sequence labeling. *CoRR*, abs/2010.03680v2.

A Dataset Details

Stats for the CMV Modes dataset are provided in Table 6. These stats are obtained after truncation of threads to 4096 token sequence length. During data analysis, we observed that several threads share the same initial post(*submission*). Hence, we make sure that all threads with the same initial post entirely lie in either the train split, or the test.

For both CMV Modes, and Dr. Inventor Corpus, we only consider contiguous spans of texts as single components, as opposed to the labelling in the dataset. Discontiguous spans are re-labelled as separate components and the model is trained and tested with these new labels, instead.

For CMV Modes dataset, we add an extra "continue" class of relations to denote relation between two dis-contiguous spans of same argumentative component annotated in the data. We group together various relation types annotated in the CMV modes data into the 5 broad classes as follows: **support**("continue" class and "support" class), **agreement**("agreement", "understand" classes), **direct attack**("attack", "rebuttal attack", "rebuttal", "disagreement" classes), **undercutter attack**("undercutter", "undercutter attack" classes), **partial**("partial agreement", "partial attack", "partial disagreement" classes). These groupings are based on the broad annotation guidelines provided for the annotations of CMV Modes data.

For Dr. Inventor Corpus, due to the low number of **semantically same** relations(44) compared to **support**(4535) and **contradicts**(564) in the original dataset, we add the label("**parts-of-same**") which indicates that two dis-contiguous spans belong to the same argumentative component to the **semantically same** category. We also, merge together sections of papers to efficiently utilise 4096 token length of Longformer model. The detailed statistics after truncation to 4096 sequence length are presented in Table 7.

Component Type	**# Tokens**
O	28186
B-C	1650
I-C	26529
B-P	1980
I-P	36552
Relation Types	**# of relations**
support	1859
agreement	421
direct attack	283
undercutter attack	330
partial	215

Table 6: Statistics for the CMV-Modes dataset.

Component Type	**# Tokens**
O	153429
B-BC	3215
I-BC	39574
B-OC	5300
I-OC	74239
B-D	3994
I-D	19058
Relation Types	**# of relations**
support	4535
Contradicts	564
Semantically Same	1049

Table 7: Statistics for the Dr. Inventor dataset.

B Implementation Details

We use the pretrained base version of Longformer (12 layers, 768 model size). The size of the local attention window was set to the default 512. The maximum sequence length was fixed at 4096. We added the special tokens that we used, to the pretrained Longformer tokenizer. For **ACI** our models use a CRF layer[3]. **sMLM** training for Longformer based models was done on thread level and for BERT and RoBERTa based models on comment-level. We used mini-batch learning; approximately similar length input threads were batched together keeping the total number of tokens per batch fixed to $8,194$ for Longformer and 1024 for BERT and RoBERTa models, and accumulated gradients over 3 batches.

We trained our models for a total of 10 epochs on sMLM task, while saving checkpoints after each epoch. We used Adam optimizer with a learning rate of 10^{-6}. For all downstream tasks, we train our models for 30 epochs, again, with Adam optimizer with learning rate of $2e-5$ as suggested by Mosbach et al. (2021). We use same batch sizes as **sMLM** training and accumulate gradients over 4 batches. We observe that for prompting **RTP** on CMV-Modes, not making **[USER-i]** tokens global, leads to better performance, hence we report results for same.

We find that **sMLM** training for 4 epochs is most beneficial, for performance on downstream task. Hence, we report results for the same checkpoint. Following the suggestions in Reimers and Gurevych (2017), we repeat our experiments on 5 different data splits and present the distributions in the Appendix. For the results at any epoch, the score plotted corresponds to mean over the 5

[3]We use the implementation of AllenNLP (Gardner et al., 2018)

Type	Markers
Opinion	*i agree, i disagree, i think, in my opinion, imo, imho*
Causation	*because, since, as, therefore, if, so, according to, hence, thus, consequently*
Rebuttal	*in contrast, yet, though, in spite of, but regardless of, however, on the contrary*
Factual	*moreover, in addition, further to this, in fact, also, firstly, secondly, lastly*
Assumption	*in the event of, as long as, so long as, provided that, assuming that, given that*
Summary	*tldr*
Misc.	*why, where, what, how, when, while*

Table 8: Types and examples of different discourse markers used for selective MLM finetuning.

runs, and error regions correspond to the Bessel corrected standard deviation. The scores reported in the tables for various models correspond to the average value of the mean of 5 runs, over the last 5 epochs for that particular metric. Table 8 provides examples of markers of various kinds, that are masked during the **sMLM** training.

C Additional results

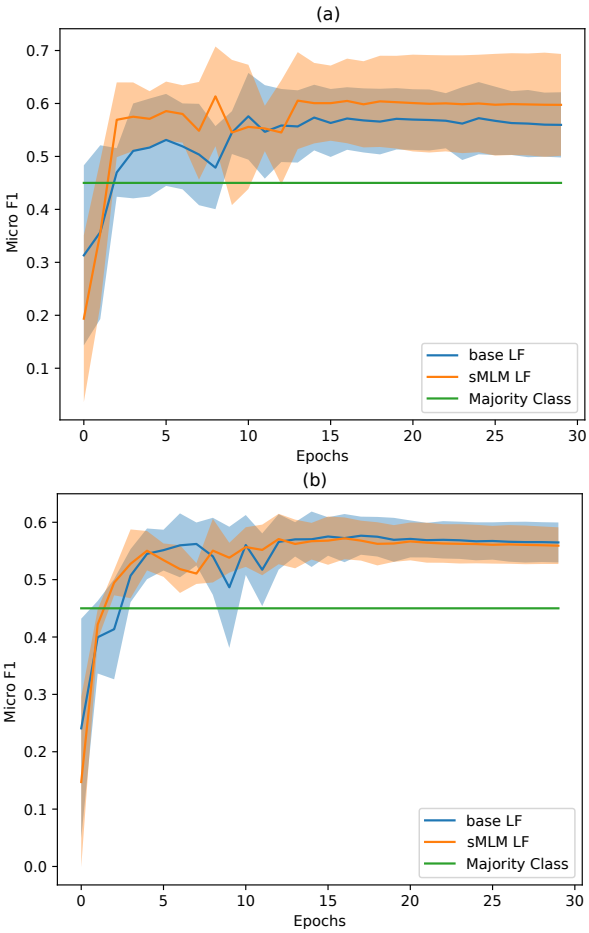


Figure 6: On CMV Modes data, sMLM-LF-mp's mean F1 converges to 0.59 compared to 0.56 for Base-LF-mp in 80-20 split (a) and 0.56 in 50-50 split (b).

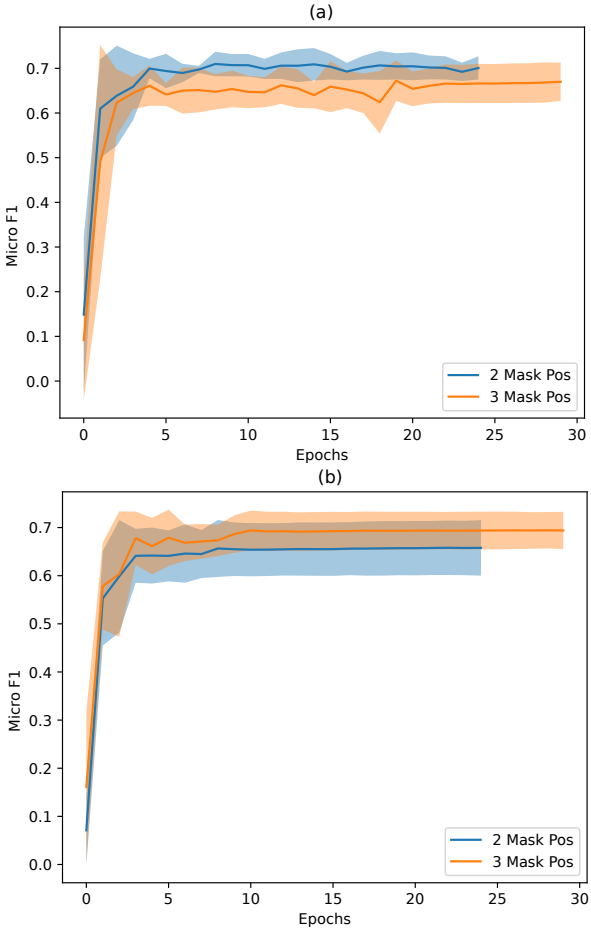


Figure 7: Change in sMLM-LF performance on CMV Modes **RTP** (a) 80-20 and (b) 50-50 split when number of mask tokens in the prompt is changed from 3 to 2. The model with 2 masked token converges to 0.70 (0.66) and the mean for 3 masked tokens converges to 0.67 (0.69).

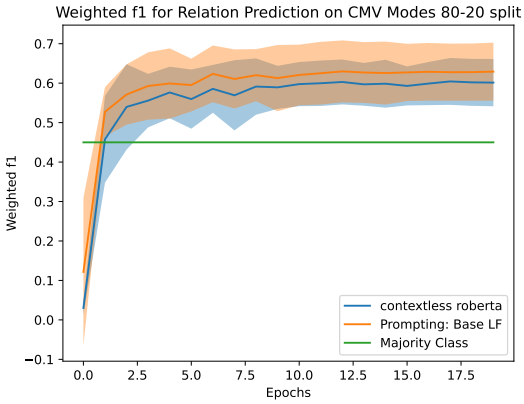


Figure 8: Contextless Roberta's mean f1 converges to around 0.599, compared to 0.62 of Base Longformer on **RTP**.

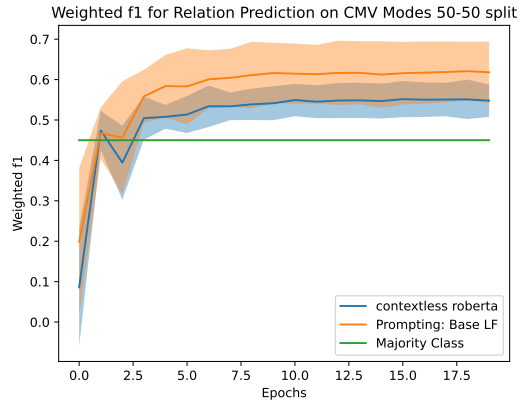


Figure 9: Contextless Roberta's mean f1 converges to around 0.55, compared to 0.617 of Base Longformer on **RTP**.

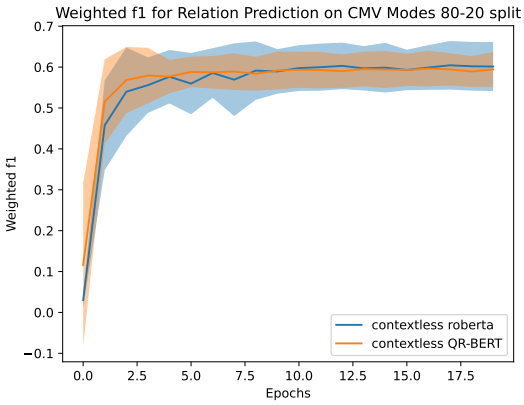


Figure 10: QR-BERT converges to an f1 score 0.59 compared to 0.60 for RoBERTa on the 80-20 split of CMV-Modes for **RTP**.

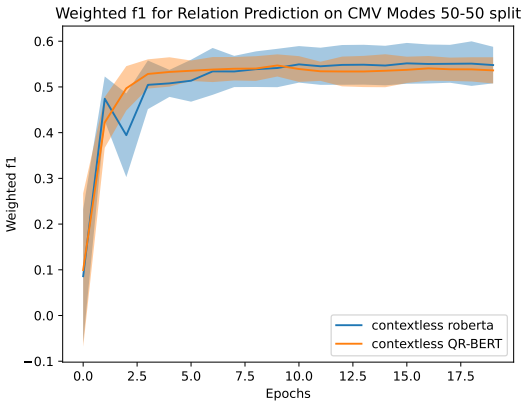


Figure 11: QR-BERT converges to an f1 score 0.54 compared to 0.55 for RoBERTa on the 50-50 split of CMV-Modes for **RTP**.

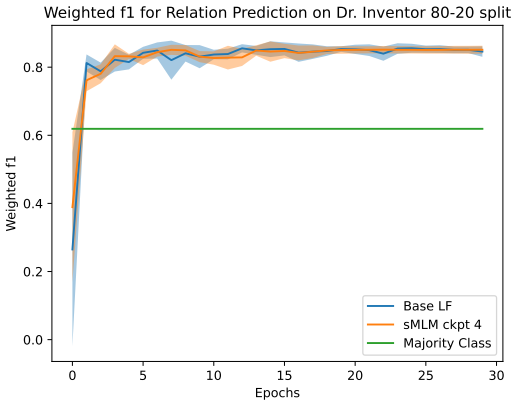


Figure 12: Both Base LF and our sMLM pretrained Longformer converge to an f1 of 0.85 with prompt-based **RTP** on Dr. Inventor corpus.

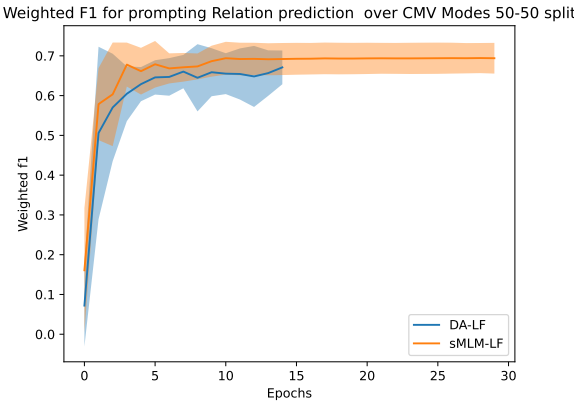


Figure 13: The Domain Adapted LF converges to around 0.66 compared to 0.69 for sMLM-LF, on the 50-50 split on CMV-Modes

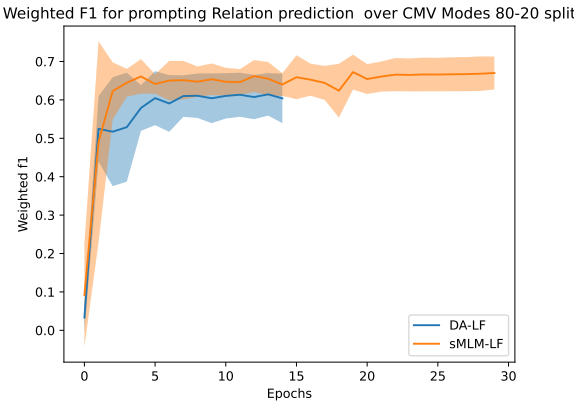


Figure 14: The Domain Adapted LF converges to around 0.61 compared to 0.67 for sMLM-LF, on the 80-20 split on CMV-Modes

Computers and Electrical Engineering 100 (2022) 107850

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Compatibility of a Truncated Fin-FinFET as a k-modulated Biosensor with Optimum parameters for Pre-emptive Diagnosis of Diseases

Mridul Prakash Kashyap, Harshal Gudaghe, Rishu Chaujar *

Engineering Physics Department, Delhi Technological University, New Delhi, India

ARTICLE INFO

Keywords:
Biosensor
TF-FinFET
Nano-cavity
Dielectric
Biomolecules
Immobilization

ABSTRACT

This work provides a comprehensive study demonstrating the dielectric modulated biosensors by considering Truncated Fin (TF) FinFET as the cardinal component. After perceiving the improvements in results, a subsequent investigation is performed to know the best possible architecture of TF-FinFET based biosensor. Therefore, it is competed from itself by varying its cavity length and the unwanted charge density that immobilizes inside the cavity on a real-time basis. TF-FinFET based biosensors with larger cavity-length occupied a higher spot with switching ratio and sensitivity (i.e., at 100x and 2x enhancement, respectively). Additionally, various figure-of-merits are also compared by varying the charge density associated with biomolecules. Finally, the label-free characterization of TF-FinFET based biosensor is founded with better results in terms of the capability to diagnose the diseases (like ovarian cancer, breast cancer, Marek's diseases etc.) by detecting their relevant biomolecules, thus making it a better choice among its counterparts.

1. Introduction

The twenty-first century brought many revolutionary technological innovations providing potential solutions to the inevitable challenges faced globally. The major obstacle in curing the discomforts is not always the severity of the disease, the delay in diagnosing that malady also plays a crucial role. For this reason, in most cases, symptoms are not recognized by the pathetic person in an early stage of the disease [1–4]. The ailments could be trapped if diagnosed timely. However, if not detected, they can make a person severely ill or even untreatable, leading to demise. These include coronary artery disease, stroke, Alzheimer's disease, diabetes, dengue fever, breast cancer. One of the reliable and fast techniques to diagnose these diseases is to detect their presence by sensing specific biomarkers associated with them. These biomolecules can be distinct proteins whose changed levels (below or above the standard range) in the body confirm the occurrence of a particular disease at an early basis. The nano-devices that can detect such biomolecules are biosensors, which detect the specific biomolecule in a given target sample such as a human biological sample (i.e., urine, blood, saliva) [5–8].

The induct of field-effect transistors (FETs) for biosensing was initiated in 1970 with ISFET (Ion sensitive field-effect transistors)

This paper was submitted for regular issues of CAEE, but it is for special section VSI-mnds. Reviews were processed by Guest Editor Dr. Angsuman Sarkar and recommended for publication.

* Corresponding author.
E-mail address: chaujar.rishu@dtu.ac.in (R. Chaujar).

https://doi.org/10.1016/j.compeleceng.2022.107850
Received 15 July 2021; Received in revised form 18 February 2022; Accepted 21 February 2022
Available online 4 March 2022
0045-7906/© 2022 Published by Elsevier Ltd.

[9], which was further became compatible with protein and DNA sensing by incorporating some structural changes. However, its drawback of finding in dry surroundings encouraged the development of DM biosensor, in 2007, with better ability to do the same. After substantial advancements, Double Gate-MOSFET [10] came into existence having improved characteristics with better durability. This evolution did not hold sufficiently. Hence, the inclusion of enhanced current and threshold voltage sensitivity came up with FinFET based biosensor. Now a member, called TF-FinFET, also joined the hierarchy with improved optimization on electrical properties if compared with the predecessors [11].

The main target of this investigation is to provide an ameliorative device in the field of detection. The detection approach by an unconventional device is applied to overcome the efficiency and sustainability issues faced in previous technologies. Subsequently, a thorough investigation of the Truncated Fin (TF) FinFET, as the more efficient and reliable option for the early diagnosis of critical ailments, is performed through modeling and simulation. First and foremost, one must know insights regarding the biomolecules to be detected using FinFETs. The biomolecules are propagated by the solution and immobilized in the pre-defined nanocavities, altering the overall transfer properties of the transducer and allowing one to detect the presence of a required biomolecule in a given solution [12]. The label-free technique of biomolecule detection is proposed to reduce the complexity and evaluation cost, which analyzes the target biomolecule based on intrinsic physical characteristics such as dielectric permittivity, mass, volume and conductivity. Compared to an air-filled cavity, a biomolecule contained cavity alters the threshold voltage characteristic of the biosensing device by {ΔVth = Vth (biomolecule filled) - Vth (air-filled)}. As far as the benefits to the medical community are concerned, it can firmly stand with lower power consumption and corresponding dissipation, hence has better efficiency. Moreover, since it possesses a considerably high switching ratio, it provides better detection accuracy in the presence of biomolecules at real-time analysis [13].

The prospective readers will be able to get a comprehensive view of the nanoscale biosensing mechanism from the perspective of the novel 'Truncated Fin-FinFET'. From section-2, they can grasp the simulation techniques applied to verify the presence of the disease, with the help of sub-sections (i.e., structural and doping variation in the device, detailed information on the biomolecules to be detected, and the mathematical formulation – used to detect their presence). Section-3 focuses on the validation of applied mathematical models along with section-4, which shows the detailed analysis of the simulated results in real-time. Finally, with section-5, the conclusion of the entire study has been presented at the end.

2. Technique of Simulation

While adapting the Truncated Fin FinFET as a biosensor (shown in Fig. 1), some modifications have been made in conventional

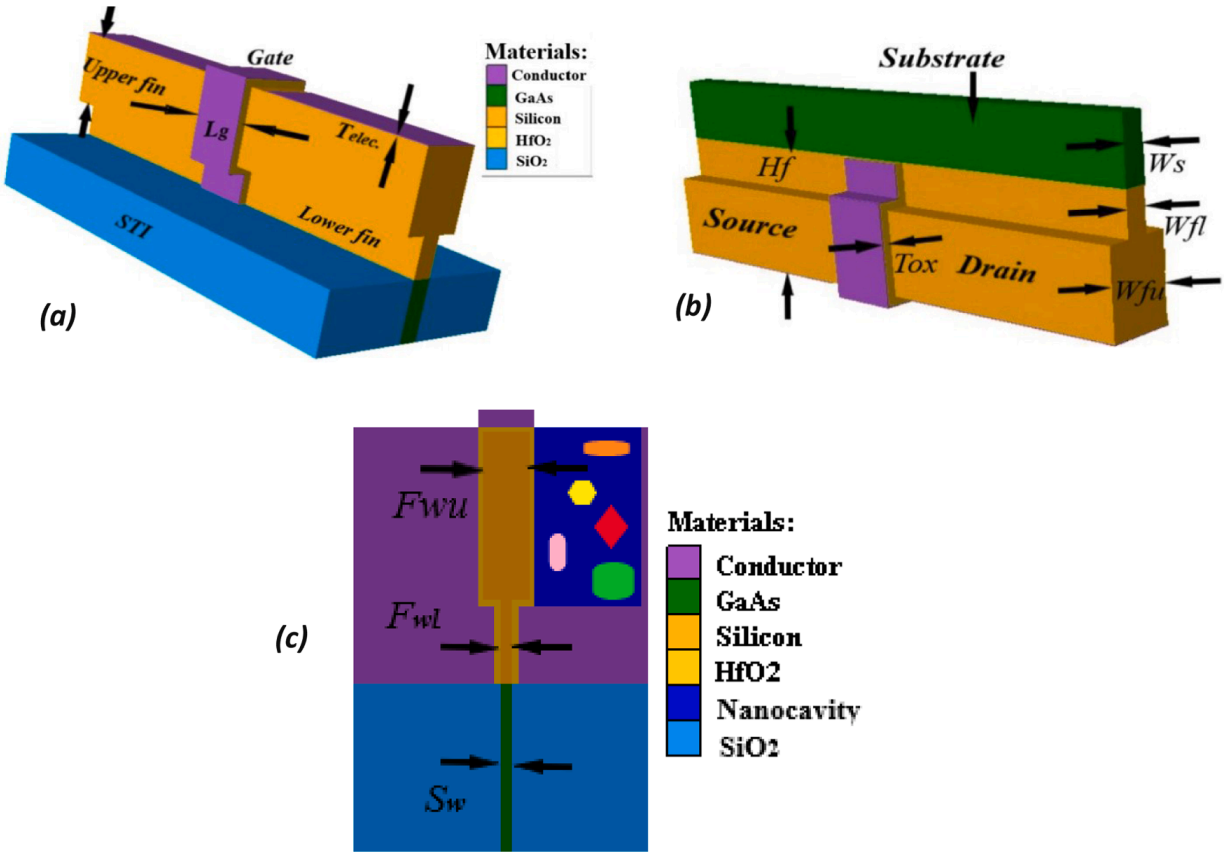


Fig. 1. (a) 3-Dimensional representation of a Truncated Fin (TF) FinFET structure without including its nano-cavity Fig. 1.(b) A more detailed 3-D inverted view of Truncated Fin FinFET without demonstrating shallow trench isolation Fig. 1.(c) 2-D representation of TF-FinFET based biosensor from latitudinal view with carrying prospective biomolecules to be immobilized at the nano-cavity.

FinFET like lower fin modulation, replacement of substrate material and gate oxide. For all the simulations performed, both types of devices and derivatives of Truncated Fin FinFET, n-type doping has been taken into consideration as follows: $10^{20}/cm^3$ for source region, $10^{19}/cm^3$ for drain region, $10^{18}/cm^3$ for the channel, and $10^{15}/cm^3$ for the GaAs substrate region. There is no usage of high-profile Gaussian style of doping. This methodology could be beneficial if seen from the perspective of fabrication. A doping profile has been chosen [1,14], in a way, to provide an effective transition of charges from source to drain region in the device. The amount of doping in the channel region has been designated relatively lower because the prospective biosensing device should have been dedicated to minimum OFF-state current (i.e., leakage current). While, the doping for the source region is maximum to render conduction through diffusion between the terminals, which improved the ON-state characteristics of the device as well.

The TF-FinFET resembles GAA (Gate All Around) FinFET and 3-gated FinFETs due to the truncation and replaced material (i.e., GaAs) of substrate used. The priority of Truncated Fin FinFET based biosensor is much inclined towards the lower power consumption and to give minimal power dissipation when detecting a biomolecule. Hence, a reliable setup was required, which provides better gate controllability over the channel region to sense more effectively. Thus, a design called 3-gated FinFET has been adopted through the investigation. However, sensing of biomolecule was conducted at a lateral side to prevent the inclusion of unwanted charge density (from targeted biomolecule) on the surface of detection and to provide implementational feasibility for future aspects.

2.1. Description of Biosensing Btructure

Since conventional FinFET was considered as a reference throughout the beginning of the result section, the same design parameters and version of simulations were adopted. Apart from this, a gate metal with a 4.55 eV work function was chosen for all the simulations.

For the parameters generated in the Table 1, the subsequent efforts have been applied to extract out the optimum dimensions before implementing them for analysis to the biosensing device. The extraction of parameters was driven by closely considering the guidelines established by the International Technology Roadmap for Semiconductors (ITRS) [15].

2.2. Insight on Targeted Biomolecules

The size of the targeted biomolecules also played a pivotal role in assigning dimensional constraints to the parameters. For sensing analysis, three types of biomolecules with their respective dielectric constant have been inducted subsequently into the nanogap cavity of both biosensors. Streptavidin ($K = 2.1$), 3-aminopropyltriethoxysilane (APTES $K = 3.57$) and His61 protein ($K = 8$). Therefore, when it comes to their respective sizes, Streptavidin is a 53 KDa tetrameric protein with a 5 nm structural diameter [16]. Apart from this, the individual (non-integrated) size of the APTES biomolecule was found as 0.5 nm [17]. His61 has also been observed as an amino-acid residue (histidine) at the 61st position in the AztC protein [18]. The molecular weight and average density of AztC protein were used to calculate its approximate spherical diameter, which was found out to be 4.17 nm. As a result, His61′s size should be much smaller than 4.17 nm.

Streptavidin is a tiny bacterial protein that binds to the vitamin biotin with a high affinity. Radioisotopes and monoclonal antibodies can be linked together using this streptavidin-biotin combination. These bound products have the property of being drawn to cancer cells and binding to them rather than normal cells. The radio-labelled products are more easily eliminated from the body, resulting in a reduced amount of radioactivity and toxicity in the body. Hence, Streptavidin-biotin could be implemented in the diagnosis as well as in the treatment of various types of cancer such as ovarian cancer, breast cancer, etc. [19]. Additionally, it aids in the diagnosis of Marek's disease virus (MDV) using the enzyme-linked immunosorbent assay (ELISA) method [20]. In the process of diagnosis of dengue virus, APTES functions as a facilitator for immobilizing biomolecules on the sensing surface. His61 (histidine) derivates as 1-methylhistidine and 3-methylhistidine, and hence identified as the biomarkers for the health and related diseases [21].

2.3. Simulation Methodology

Several physical simulation models were employed to anticipate & analyze the electrical behavior of the concerned devices just before their practical implementation on the silicon wafer [22]. The ***Bohm Quantum Potential model***, demonstrated by Eq. (1) evaluates the effects of quantum confinement on electron and hole concentrations. Hence, the position-dependent quantum potential

Table 1
Structural Parameters of the biosensor.

No.	Structural Dimensions of simulated devices (nm)		
	Parameters	*Truncated Fin FinFET*	*Conventional FinFET*
1	Length of Gate	20, 30	20
2	Width of Substrate (S_W)	3	8
3	Thickness of gate oxide	1	1
4	Height of Nano-cavity	30	30
5	Width of Nano-cavity (Wnc)	15	15
6	Width of Upper fin (F_{WU})	8	8
7	Width of lower fin (F_W)	3	8

(Q) was investigated using the BQP model, which ultimately determines the charge carrier potential energy at the concerned region. ***Band-to-Band Tunnelling*** (BTBT) is a non-local phenomenon that causes some of the most common leakage mechanisms (GBBT-Tunnelling Generation Rate) in transistors, including Gate Induced Drain Leakage. To effectively capture the leakage current, BTBT in Technology Computer-Aided Design simulations has been incorporated.

Moreover, the distortion caused by the undesirable entry of the electrons towards the substrate and gate oxide of the biosensor was analysed using the BTBT model, shown by Eq. (2). The statistics of recombination and formation of charges in semiconductors, which occur through the trapping mechanism, is described using the doping dependent ***Shockley-Read-Hall model*** with Eq. (3). Hence, to address the distortion of the simulated device properties, by the rate of combination (R_{SRH}) of the electrons and ionized atoms created by the development of a sufficiently large electric field at the drain-channel interface, the doping dependent model of SRH recombination was employed. The concentration of doping affects the charge carrier movement and the respective device consistency that have been analyzed using the ***Lombardi CVT mobility model*** to consider the transverse field, doping dependent and, temperature-dependent parts of the mobility. It is demonstrated by Mathiessen's rule as per Eq. (4). The correlation between the generation of electric field and the ionization rate was established using the *Impact-Ionization model* with the help of n,p (coefficient of ionization) and shown by Eq. (5).

$$Q = \frac{-h^2}{2}\frac{\gamma\nabla\left(M^{-1}\nabla(n^{\propto})\right)}{n^{\propto}} \tag{1}$$

$$G_{BBT} = D\left(4e^{14}\right)\left(E^{2.5}\right)exp\left(-\frac{1.9e^{7}}{E}\right) \tag{2}$$

$$R_{SRH} = \frac{pn - n_{ieff}^2}{\tau_p[n+n_1]+\tau_n[p+p_1]} \tag{3}$$

Where, $\left\{n_1 = n_{ieff}.expexp\left[\frac{E_{trap}}{KT_L}\right]\right\}$ and $\left\{p_1 = n_{ieff}.expexp\left[-\frac{E_{trap}}{KT_L}\right]\right\}$

$$\frac{1}{\mu_T} = \frac{1}{\mu_{AC}} + \frac{1}{\mu_b} + \frac{1}{\mu_{sr}} \tag{4}$$

$$\alpha_{n,p} = \frac{1}{\lambda_{n,p}} exp\left[C_o(r) + C_1(r)x + C_2(r)x^2\right] \tag{5}$$

3. Calibration and validation of models

As mentioned below, Fig. 2 depicts the calibration of the simulated results of this work with the results of previously published experimental research [23], which is annotated by respective keywords for our simulated results (i.e., "Simulated") and for the results of previously published research (i.e., "Reference") which is being referenced by us. The minor deviations can be grasped from this Figure. By taking the results of previously published research as a reference [23], approximately 4.65% and 1% deviations have been

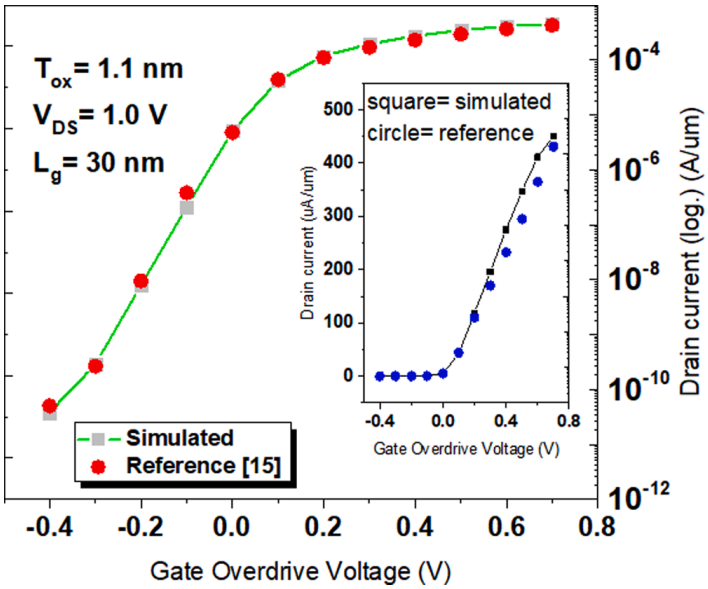


Fig. 2. Calibration of the simulated points to the reference data points, extracted from [23], simulated under the similar model condition as mentioned in Sachid et al., thus end up with the validation of used models during simulation.

deduced, if seen from the point of view of I_{ON} and I_{OFF} respectively.

The simulations were conducted in the specified gate voltage range (from V_{GS}= 0 V to V_{GS}= 1 V). Moreover, the deviation was also discovered when V_{th} was taken into consideration. As its detected value by simulation was 69.98 mV/dec, associated deviation has been deduced to be 6%. While taking into consideration the above calibration depiction and respective deviations, the validity of the models employed in the simulations of this research work can be clearly deduced.

4. Discussion on simulated results

Both the transistors were designed in a way that they can reflect the changes, in input and output characteristics, when bio-molecules with different dielectric properties, were immobilized in their nano-cavities. The main idea is to analyze the sample material in real-time when biomolecules from the sample immobilize in the designated nano-cavity of FinFET. Thus, altering the overall analogue characteristics of the detecting devices. Since the changes that occurred through this process can easily be identified, one can detect the presence of biomolecules effectively as conventional real-time approaches do [24,25].

4.1. Comparison of TF-FinFET based biosensor with conventional FinFET based biosensor

As expressed in this section, the main benefit lies with Truncated Fin FinFET is its leakage power optimization capability, which could not be underscored if seen from system-on-chip aspect.

The comparative investigation has been taken out, especially a transistor-based biosensing, to evident the TF-FinFET based biosensor ahead than the conventional FinFET based biosensor (as shown in Fig. 3. Mentioned below, all the simulations have performed on a real time basis and in the presence of biomolecules at device nano-cavity.

Here, as interpreted from the Fig. 4, irregular nature of the electric field (ranges from 0 to 3 × 10^6 V/cm) within the gate oxide (SiO_2) was extracted much higher than that of TF-FinFET with HfO_2 and in fact, higher electric field accounted for the dielectric degradation of gate-dielectric. Therefore, one could depict that the deterioration would be least in TF-FinFET based biosensor which evident its robustness over its conventional FinFET based biosensor.

One can clearly grasp the features of both the simulated FinFETs from Fig. 5 (a, b & c). Sensing ability, in Fig. 5(c), of Truncated Fin FinFET was observed a little bit lower than its standard counterpart when analysed according to increased k-values of the biomolecule, however, it was depicted superior when viewed from the aspects of I_{OFF} and I_{ON}. The OFF-state current with TF-FinFET found within the range of 10^{-8} to 10^{-9} A/um whereas it is in the span of 10^{-7} to 10^{-8} A/um for another FinFET, indicating a significant reduction in terms of power leakage for TF-FinFET. In addition, the I_{ON}/I_{OFF} (i.e., switching ratio) of TF-FinFET was detected out as 100x than that of conventional one and the same is being shown from Fig. 5(b).

In Fig. 5(a), the transconductance (g_m) for both the devices has been shown to grasp the idea of their orientation towards the signal applied at input terminal. Form the concerned figure, one can estimate the gate voltage (i.e., V_{gs}= 0.3 V) through which the g_m of TF-FinFET begins to overtake that of conventional FinFET. Hence, The TF-FinFET could be assumed more input driven in the gate voltage range of 0.3 V to 0.7 V. The analysis was done by taking Eq. (6) into consideration.

The ON-state leakage current plays a crucial role if seen from the power dissipation perspective at real-time analysis. Therefore, the variation of substrate current in Fig. 6(a) has been taken into account to predict the on-time feasibility of the biosensing device. This analysis demonstrates the advantage of Truncated fin FinFET over the conventional FinFET based biosensing device. While, substrate current attained by the conventional FinFET marked in the range of 10^{-9} to 10^{-8} A/um, whereas it stood in the span of 10^{-12} to 10^{-9} A/um, showing diminished leakage at TF-FinFET. The enhancement in TF-FinFET is evident due to the indulgence of truncation at lower fin and supplanted substrate material (i.e., from Si to GaAs).

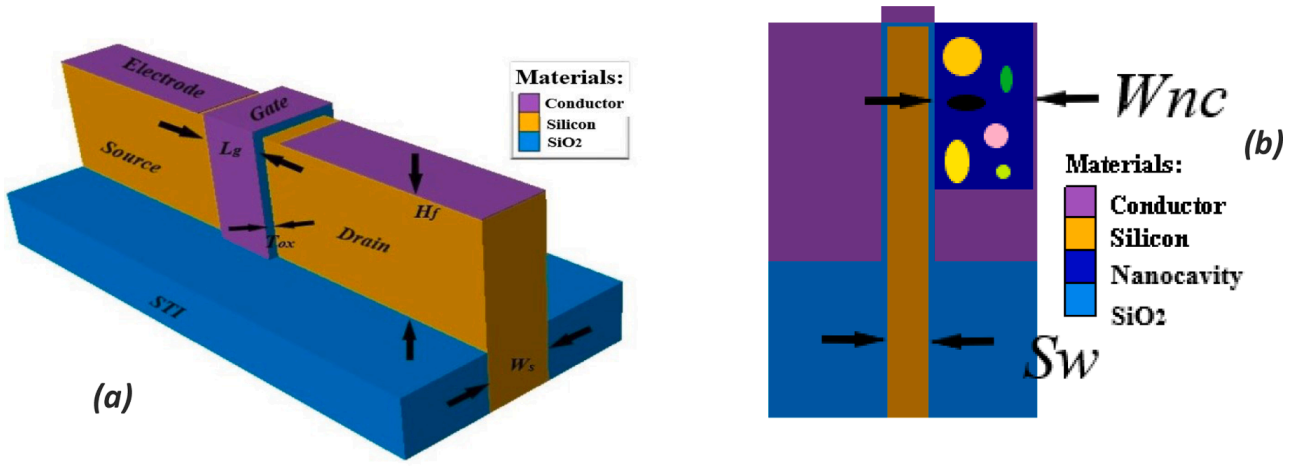


Fig. 3. (a) 3-D representation of conventional FinFET without showing its attached nano-cavity Fig. 3.(b) 2-D representation of conventional FinFET with a nano-cavity attached to one of its lateral sides.

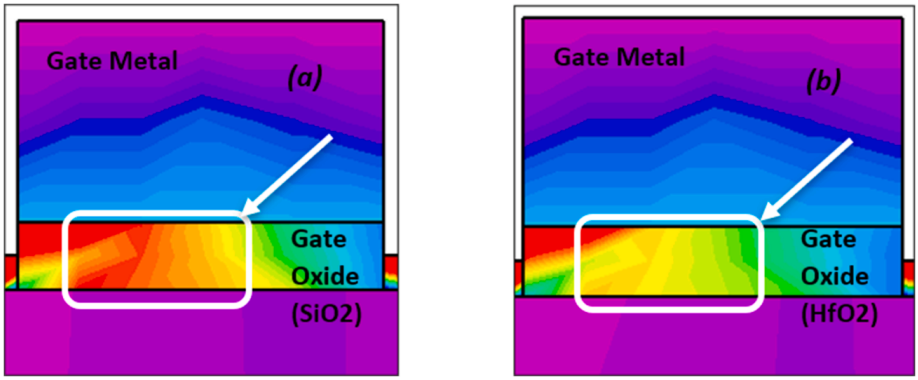


Fig. 4. Comparison on variance of electric spread in the gate oxide of conventional FinFET (left) [1][24] and Truncated Fin FinFET (right), from their respective magnified view.

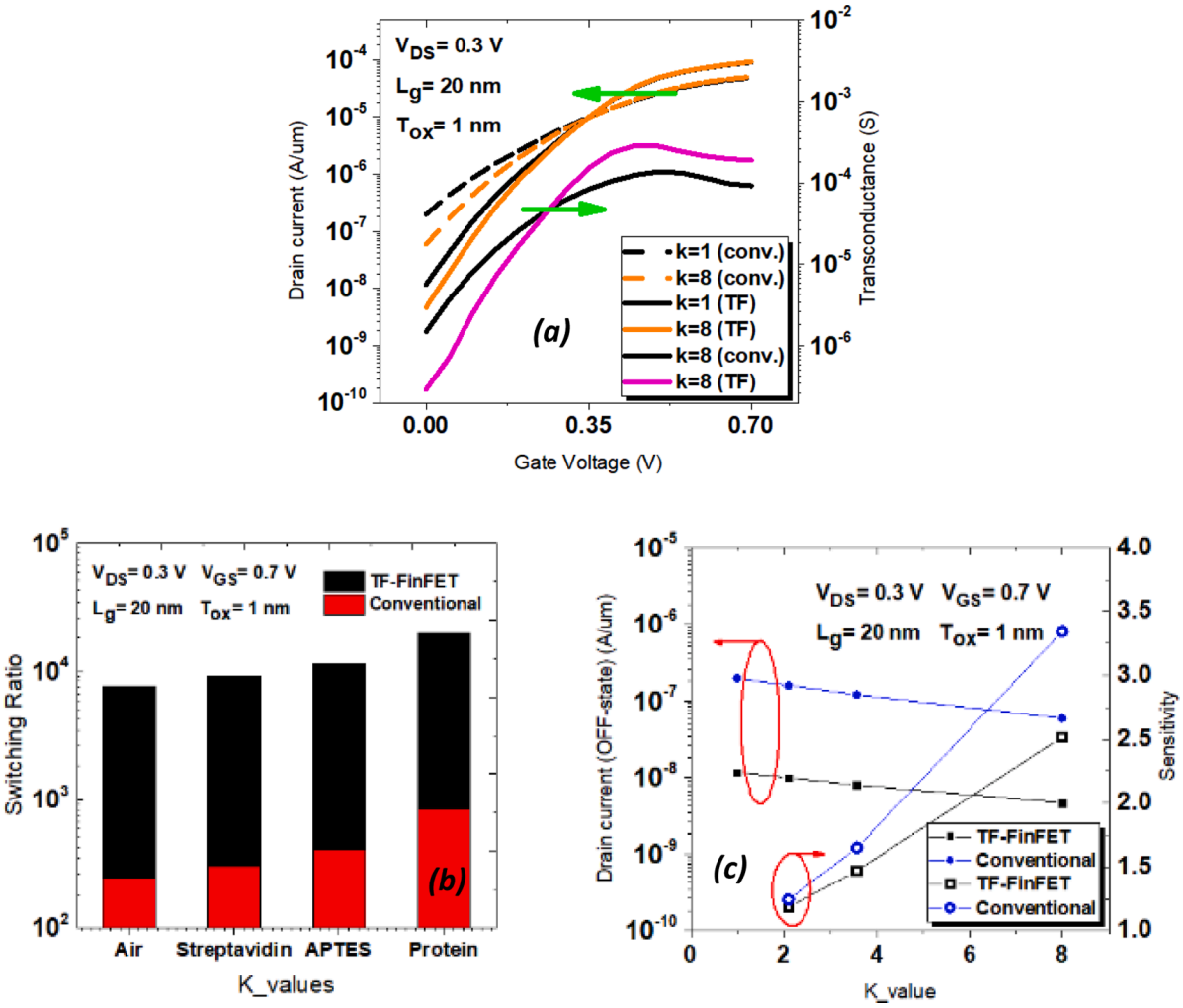


Fig. 5. (a) Comparison of drain current between conventional and Truncated Fin FinFET [1,14,24] additionally, dependency of transconductance on gate voltage for TF-FinFET. Fig. 5 (b) Comparison of Switching ratio for both the simulated devices. Fig. 5 (c) Comparison on the basis of OFF-state drain current at no gate voltage and sensitivity for both devices w.r.t applied dielectric values.

As far as the reliability of the biosensing device is concerned, it has to withstand the lower subthreshold slope (i.e., around 60 mV/dec) and a considerable amount of threshold voltage as shown in Fig. 6(b). These properties denote the overall capacity of the device to prevent it from getting collapsed at extreme input applied voltages. Through simulation, the minimum subthreshold slope value for TF-FinFET at $K = 8$ (i.e., for protein) has been recorded as 81.24 mV/dec and 114 mV/dec for conventional FinFET with the same biomarker. Apart from this, the maximum value of threshold voltage for TF-FinFET, at $K = 8$, was observed as 191.2 mV and 159.6 mV that for conventional FinFET at the same K-value. Consequently, one can choose the TF-FinFET based biosensor for a better reliability

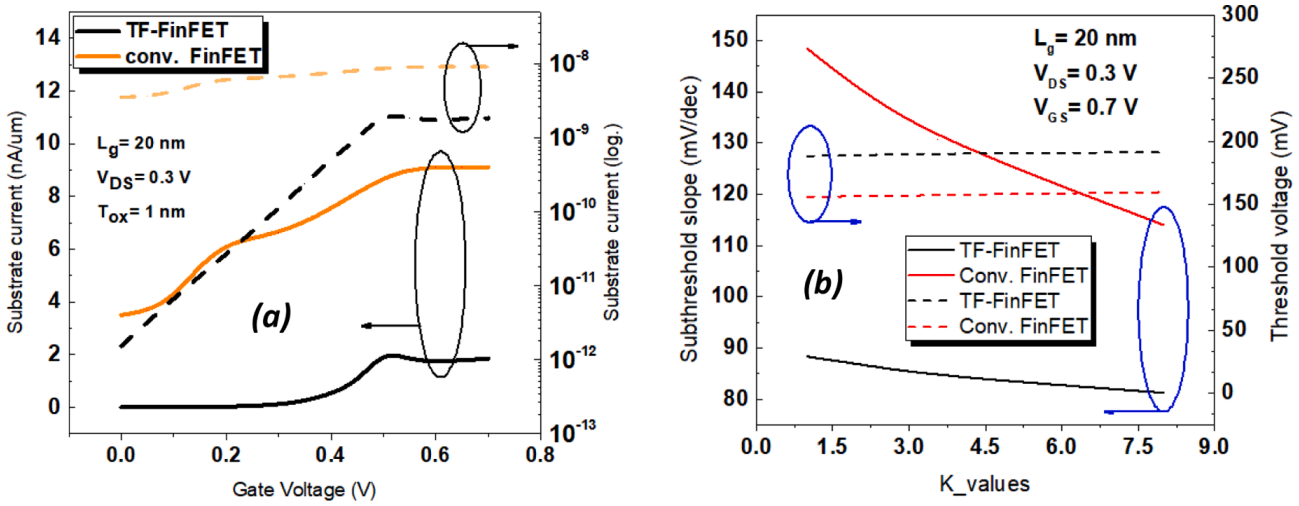


Fig. 6. (a) Dependency of substrate current on the input gate voltage Fig. 6.(b) Variation of subthreshold slope and threshold voltage w.r.t bio-molecules possessing different dielectric constant values.

experience.

To have a clear comparison between TF-FinFET based transducer and its counterpart, several figure-of-merits (FOMs) have been considered. The analysis was conducted at extreme K-values (i.e., $K = 1$ and $K = 8$) for both FinFET devices. Since TF-FinFET provides minimum output conductance (gds) and high transconductance at $K = 1$, it possesses the highest intrinsic gain, in Fig. 7(a), among rest. Hence, the depiction could be made that the TF-FinFET has significantly higher signal processing strength (of about 3 times) in comparison to the conventional one. On the other hand, Conventional FinFET, at $K = 1$, has the lowest intrinsic gain (of 11.968 dB) hence signifying it as the least powerful among the rest, as per the Eq. (7).

Another FOM, transconductance generator factor (TGF), has been investigated in Fig. 7(a), which denotes the efficiency of the operated transducer throughout the active state. Since TF-FinFET provides the highest transconductance, at $K = 1$, among the counterparts, it holds the highest slot in terms of efficiency (i.e., 64.9 V^{-1}) from initial to gate voltage of 0.5 V. The deduction has been performed as per Eq. (8). Similarly, the conventional FinFET at $K = 1$ stood at the lowest spot of intrinsic gain (i.e., 23.42 V^{-1}) as it possesses small transconductance.

Since its as important to consider the output characteristics as much as it is to investigate the input behavior of the transducer, early voltage (Vea) has also been analyzed in Fig. 7(b). By considering Eq. (9) and Eq. (10), the simulations were performed to have insights into the output behavior of the device against the applied input signal. As this FOM signifies the amplifying capacity of the transducer, it should be appreciably strong to convert even a diminished signal into observable quantity. Due to inverse proportionality with gds, the early voltage conveys the idea of device orientation towards either side (input or output).

Output conductance, in Fig. 7(b), stood minimum for TF-FinFET (i.e., 3.06 uS) at $K = 1$ and maximum for conventional FinFET (i.e., 7.26 uS) at $K = 1$, indicating the lowest orientation of TF-FinFET at least K-value towards the output side. At the same time, the early voltage of TF-FinFET, at $K = 1$, was marked maximum (30.156 V) and minimum (6.55 V) for conventional FinFET based biosensing

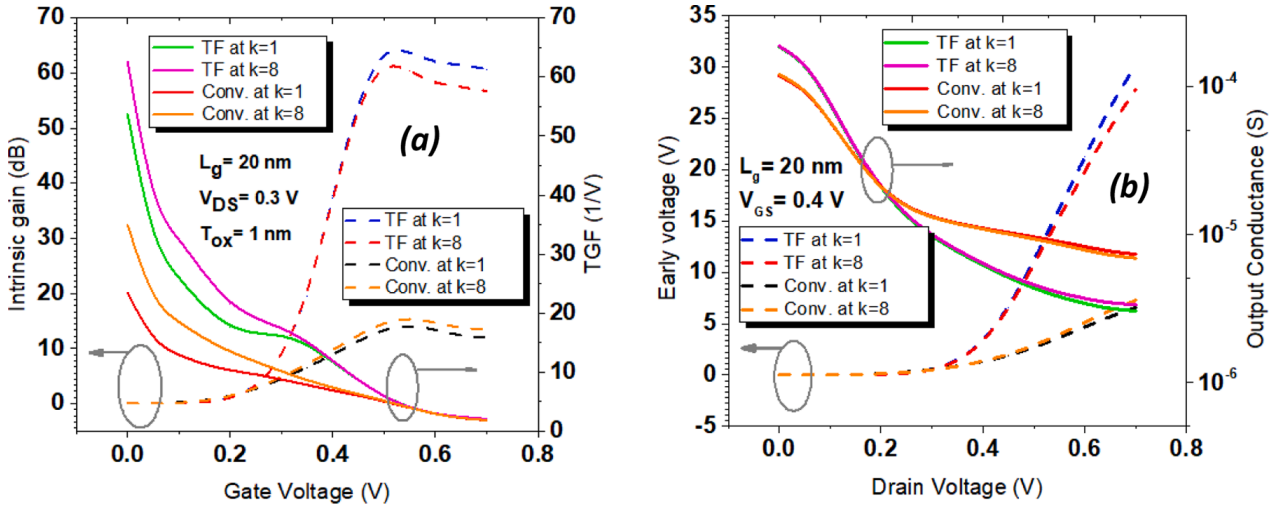


Fig. 7. (a) Dependence of intrinsic gain and transconductance generator factor (TGF) on the input gate voltage Fig. 7.(b) Variation of early voltage and output conductance w.r.t applied drain voltage for extreme K-values.

devices. This analysis clearly states that the highest input-driven signal amplification occurred at $K = 1$ for TF-FinFET based biosensor. Hence, this evident the robustness of TF-FinFET over the conventional device-controlled biosensor.

These devices are studied through the analog FOMs and evident high-performance characteristics. These characteristics are described and analyzed as shown below.

$$g_m = \partial I_D / \partial V_G \tag{6}$$

$$A_v = g_m / g_d \tag{7}$$

$$TGF = g_m / I_D \tag{8}$$

$$g_{ds} = \partial I_D / \partial V_D \tag{9}$$

$$V_{EA} = I_D / g_{ds} \tag{10}$$

After perceiving the compatibility of Truncated Fin FinFET in the field of biosensing, some more deep analysis has been conducted from 2 major aspects. First one is the variation of its cavity length (i.e., 20 nm and 30 nm). Second one, usage of commonly viable surface charge densities at the appropriate volume (i.e., the designated nano-cavity).

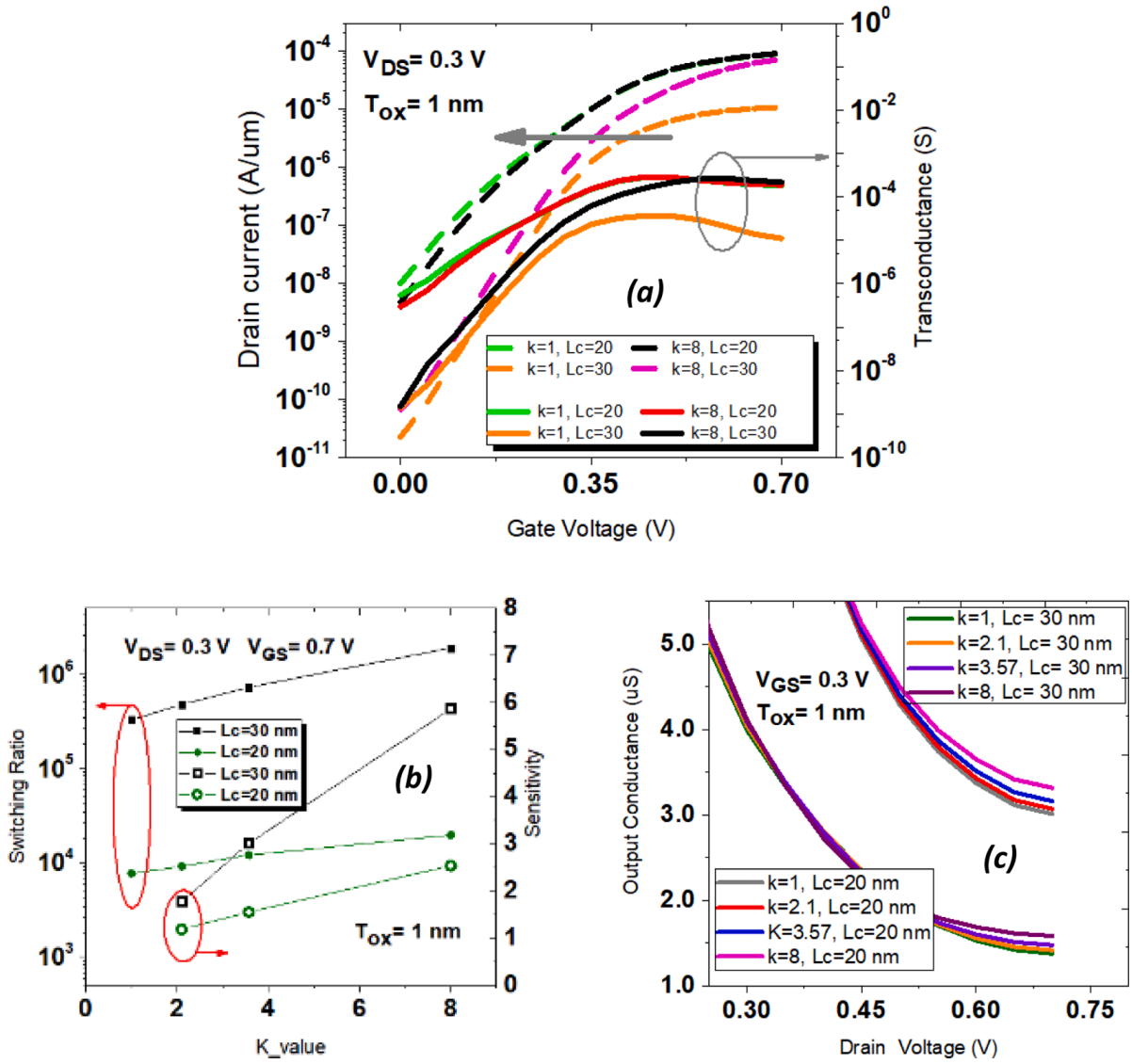


Fig. 8. (a) Dependency of drain current and transconductance on gate voltage at 2 different cavity lengths Fig. 8.(b) Switching ratio and sensitivity both the device with both cavities Fig. 8.(c) Output conductance w.r.t drain voltage for both cavity lengths.

4.2. Effect of cavity length on TF-FinFET based biosensor

By taking Fig. 8(a) into account, the overall optimization could be estimated if seen to larger cavity (Lc = 30 nm). The maximum achievable ON-state current has been attained by the device with a smaller cavity (Lc = 20 nm) at $k = 8$ (i.e., 9.19×10^{-5} A/um) but, the device with a longer cavity proved out to be more profitable when it comes to OFF-state current at same k-value.

There is a noticeable difference (of about 10 times) between I_{ON} of the device with different cavity lengths at $k = 8$ and a significantly large difference (of around 230 times) between their respective OFF-state currents. This improvement is due to minimal leakage at ON-state and higher stability at OFF-state. So, biosensor with Lc =30 nm marked its dominance over its competing cavity length of Lc = 20 nm.

With the characteristics that signify the device orientation towards the input port, the claim of enhancement at Lc = 30 nm was found more valid. The transconductance, in Fig. 8(a), for Lc = 30 nm at $K = 8$ is greater than that of Lc = 20 nm in input voltage span of 0.45 V to 0.6 V. The dependency, having a longer cavity and the smaller one, of drain current on the input applied signal was observed to favouring the longer cavity length. So, one can depict that on moving up with the cavity length, the orientation of the biosensing device to the input side increases.

The ratio of ON-state current to OFF-state current (i.e., switching ratio) has been taken into account to determine their sensing productivity and precision. In Fig. 8(b), the device with Lc = 30 nm was found at switching ratio of 1.9×10^6 and at 1.97×10^4 for Lc = 20 nm, which ultimately made the larger device more than 100 times productive than the smaller counterpart at $k = 8$. The definite reason behind this is the large separation between the I_{ON} and I_{OFF} of the device with Lc = 30 nm. The larger cavity stood better when it came to the sensitivity as well. It was detected at 5.87 sensitivity mark in comparison with that of Lc = 20 nm (i.e., 2.54), which ended up as more than 2 times enhancement for the longer cavity. The reason hiding is the more capturing volume has been deployed for observing this phenomenon in case of Lc = 30 nm.

As far as the output conductance is concerned, one can depict the behavior of the biosensing device at both mentioned cavity

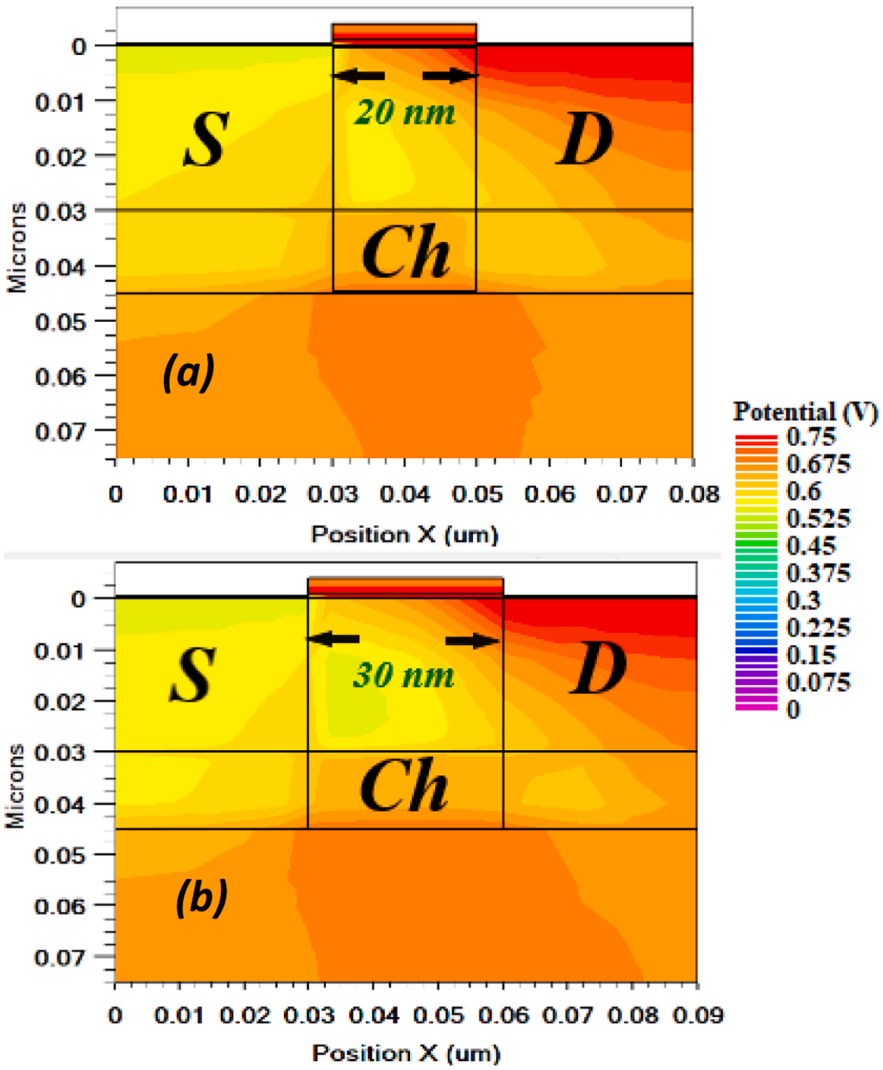


Fig. 9. (a) Potential spread inside the biosensing device with Lc = 20 nm at V_{GS}= 0.7 V and V_{DS}= 0.3 V Fig. 9.(b) with Lc= 30 nm at V_{GS}= 0.7 V and V_{DS}= 0.3 V.

lengths in Fig. 8(c). The higher the output conductance (g_{ds}), the greater is the device orientation towards signal at output port. This characteristic, with Lc = 20 nm at $k = 1$, stood about 2.19 times (i.e., 3×10^{-6}) to that of larger cavity length (i.e., 1.37×10^{-6}). From the perspective of analog terminology, lower output orientation is desirable. As a result, the device with a longer cavity has been proved more inclined towards input signal.

The potential spread inside the biosensor has been studied to have better insights of biosensing operation governed by it, in Fig. 9. Viewing the channel region, one can analyze the operational variables of the biosensor at both cavity lengths. In the case of larger length cavity, 14.28% more decrement of potential has been identified at the channel center, hence developing more potential difference inside the channel. This increment as a result improved the conduction across the channel and made it more productive than its smaller competing length.

For Figure of merits (FOMs), the above-mentioned characteristics have been used to analyze the overall intrinsic behavior of the cavity length varying biosensing device. TGF has been observed as per the biomolecules detected inside the nano-cavity while ramping gate voltage, in Fig. 10(a). Here, the device with Lc = 30 nm has been investigated with superior variations. If seen from a k-value point of view, the device with Lc = 30 nm at $k = 8$ has been found 2.7 times greater than Lc = 20 nm. Consecutively, device with longer cavity has more operating efficiency due to more detectable volume for immobilization and higher stability. Similarly for intrinsic gain in Fig. 10(b), both cavity lengths have also been studied to predict the signal processing strength of the biosensing device. The device at Lc = 30 nm unlike the device at Lc = 20 nm, has been found as a more gain providing device with the peak value of 170.44 dB for $k = 1$, whereas device with shorter cavity length found with 63.78 dB gain. This 2.67 times improvement has been developed by waxing up the cavity volume 50%.

4.3. *Effect of biomolecular charge density on TF-FinFET based biosensor*

For anticipating the more practical behavior of our modelled biosensing device, the charges contained with the targeted biomolecules have also been taken into account. These charges, along with the biomolecules immobilize into the nano-cavity, alter the desirable properties of the gate dielectric and hence sabotage the overall transfer characteristics.

The distortion could be observed by investigating the positive V_{th} shift for negative charge density and vice-versa. The higher the amount of charge density, the greater would be the V_{th} shift. The same could easily be depicted from Fig. 11(a), the variation in the OFF-state currents from the concerned plots demonstrates the variation of V_{th}. A more than 10 times difference has been observed, if seen particularly for $K = 8$, at two different applied surface charge densities Fig. 11.(b) could be considered; where switching ratio (I_{ON}/I_{OFF}) was not varied considerably if observed at each applied charge density inside the cavity with a constant cavity length of 30 nm. A cavity length of 30 nm has been taken because of its superiority analysed in the previous section. The I_{ON}/I_{OFF} in case of lower applied charge density was detected (i.e., 9.58×10^5) about 10 times the case having higher applied charge density (i.e., 8.85×10^4).

Apart from this, with lower charge density, the device was proved extremely reliable from the perspective of sensitivity. It stood maximum for $K = 8$ in both the cases with the sensitivity value of 1.62 and 3.36 at charge densities of 10^{12} and 10^{12} respectively. Therefore, the lower the impact of charges at nano-cavity, the higher is the sensitivity of the biosensing device.

Now to signify the orientation of the device towards the dielectric material applied at the nano-cavity, change in transconductance (in Fig. 12) due to the material having unwanted biomolecular charge density w.r.t the material having not at $k = 1$, has been analysed subsequently. The conclusion can be drawn that the dielectric characteristics dominate in the linear as well as in the subthreshold region, and drain current is inclined as per its sensitivity to the drain-source contact resistance. If seen comparatively, transconductance change at higher applied charge density has been found around 7 times, thus making it more influenced by immobilized material [24,25].

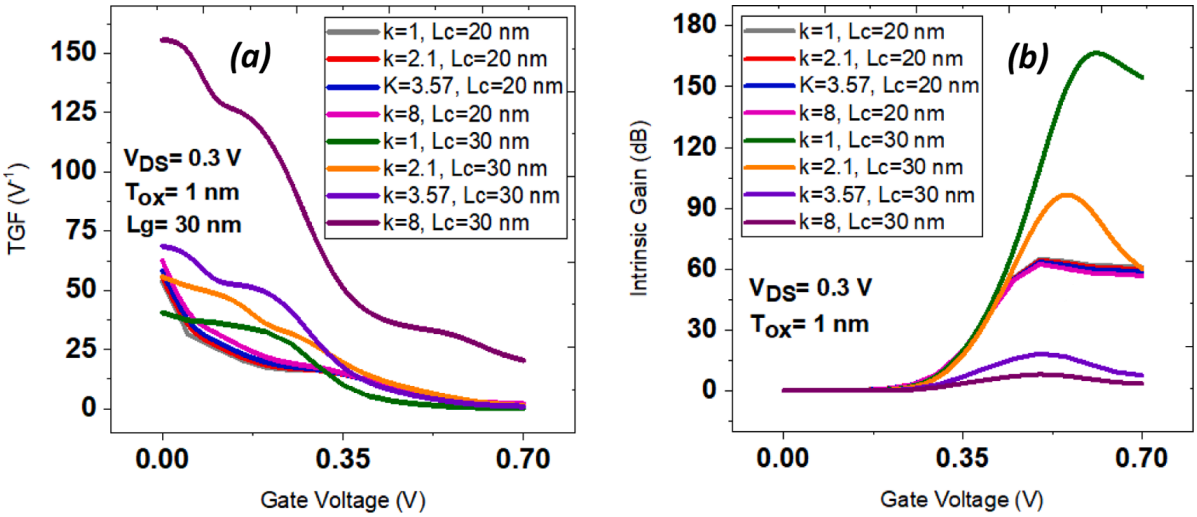


Fig. 10. (a) Variation of TGF (transconductance generator factor) w.r.t applied gate voltage Fig. 10.(b) Variation of Intrinsic gain w.r.t input gate voltage.

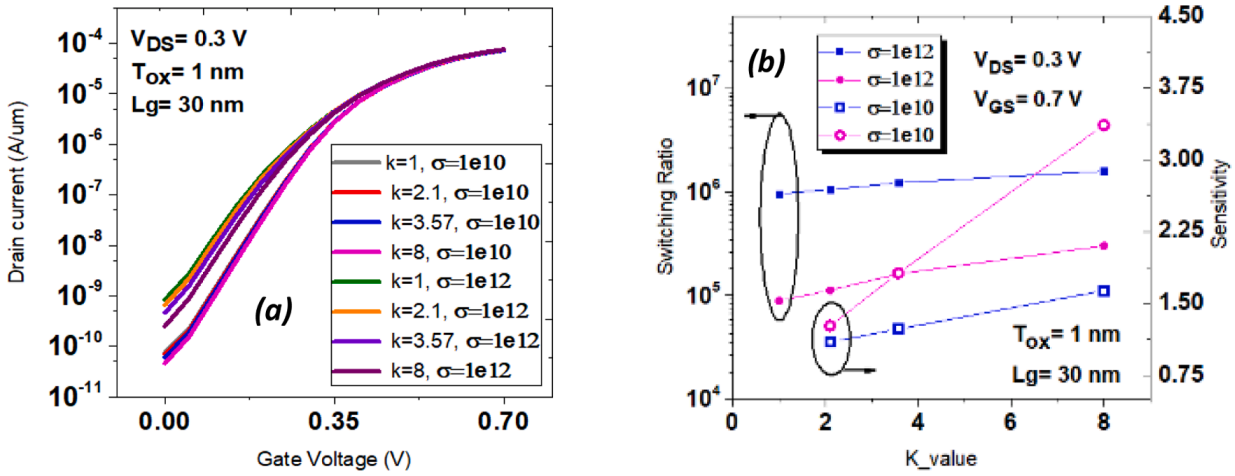


Fig. 11. (a) Dependency of drain current on ramping up input voltage at two differently applied surface charge densities Fig. 11.(b) Variation of switching ratio and sensitivity w.r.t k-value contained by biomolecules at same criteria.

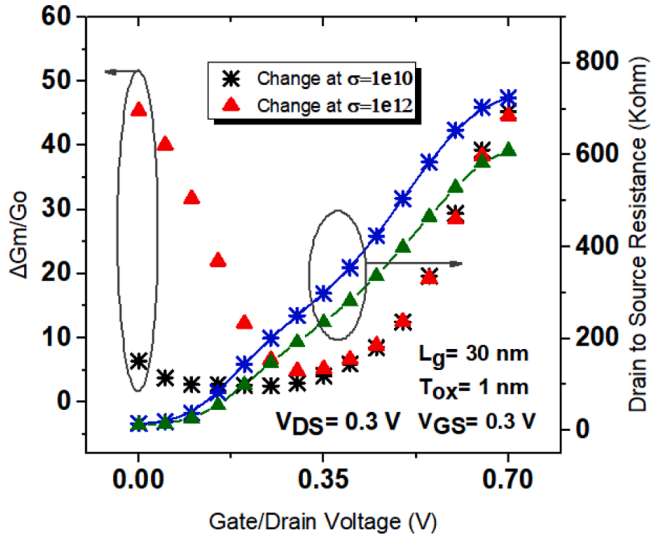


Fig. 12. Dependency of change in transconductance on the gate voltage (ramped voltage from 0 to 0.7 V) at V_{DS}= 0.3 V, and variation of drain to source resistance w.r.t drain voltage (ramped voltage from 0 to 0.7 V) at V_{GS}= 0.3 V.

Another parameter which decides the extend of conduction occurring inside the biosensing device is the drain to source resistance, in Fig. 12. Device with lower impurity charge density has been found with greater drain to source resistance, which infers the robustness of device with lesser charge density. Since larger amount of positive charge not only recombines with the conducting electrons but also attracts them towards the gate terminal and thus ended up with less resistance as well as higher gate leakage current.

As shown in Fig. 13(a), the TGF of the device (with lesser surface charge density at Si-SiO_2 interface) found with more desirable results instead of the other one. Since TGF in itself exemplify the efficiency on the device under test, device with lesser undesirable charge density has more efficiency. For the case, device having biomolecule of $k = 8$ has been investigated with 51.34 V^{-1} for $\sigma = 10^{10}$ and 47.25 for $\sigma = 10^{12}$, ultimately making 8.8% more distortion w.r.t device with lesser charge density.

As far as the lossless output generation and amplifying capacity of the device are concerned, the intrinsic gain plays a vital role Fig. 13.(b) shows the variation of signal processing strength of the device, at 2 differently applied charge densities w.r.t gate voltage, that was analysed maximum for dielectric material of $k = 1$. In contrast with the case of charge density 10^{12}, the case with lesser charge density has 27.1% higher processing strength.

5. Conclusion

A comprehensive study has been conducted, with Atlas Silvaco TCAD tools, to anticipate the adaptability of Truncated Fin FinFET on label-free biosensing operation. TF-FinFET has been predilected over the convention FinFET technology due to improvements at ON and OFF state power dissipation. Subsequently, the analog behavior of Truncated Fin FinFET was taken into account while varying the dimensional and undesirable parameters to predict its performance in near actual circumstances. Initially, comparison analysis was

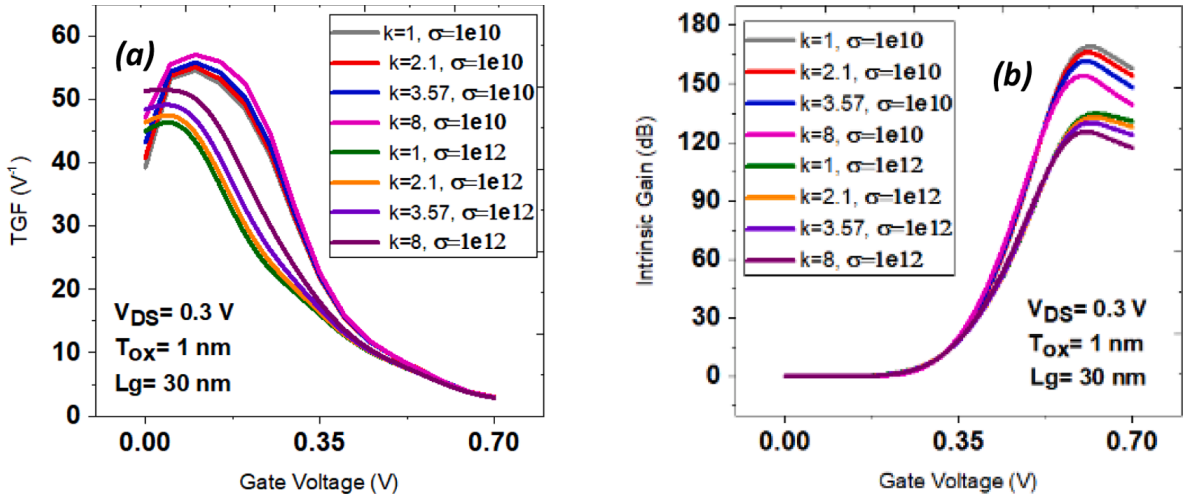


Fig. 13. (a) Variation in TGF with respect to input gate voltage Fig. 13.(b) Variation in Intrinsic gain with respect to input gate voltage at differently applied surface charge densities.

performed between the TF-FinFET and its conventional counterpart to draw the superior candidate between them. Various properties (such as OFF-state current, substrate current, switching ratio and input-output FOMs) were analyzed for each type of biosensing device. The improvements have been observed in the case of TF-FinFET based biosensors when compared with the conventional FinFET based biosensor. The improvements marked more than 5 times and 10 times for I_{ON} and I_{OFF} respectively and ultimately rendered about 100 times enhancement in case of switching ratio. For FOMs, TF-FinFET achieved 3 times greater intrinsic gain and 4.8 times enhanced early voltage, etc. After perceiving TF-FinFET as a better option, the two different cavity lengths (20 nm and 30 nm) have been taken to study the impact on biosensing possessed by Truncated Fin FinFET. This investigation was ended up by analysing 230 times and 2 times enhancement at I_{OFF} and sensitivity respectively, and 2.67 times improvement of intrinsic gain in case of the longer cavity. Finally, 2 distinct charge densities have been employed which impacted the analog characteristics such as 10 times and 2.1 times refinement at switching ratio and sensitivity respectively, and an 8.8% increase in efficiency in case of lower applied charge density. All these achievements made Truncated Fin FinFET a better choice in the biosensing world and left a great future scope in similar operations implemented at radio frequencies.

Biographies

Mridul Prakash Kashyap received the B. Tech degree in Engineering Physics (Major in Electronics and Minor in material science and Technology) from the department of Applied Physics, Delhi Technological University (2021), INDIA. He is currently working as Student Researcher at Microelectronics Research Laboratory at Delhi Technological University. His current interests include Nano-electronics, modeling and simulation od nanodevices and bio-electronics.

Harshal Gudaghe received B.Tech. degree in Engineering Physics with majors in electronics from Delhi Technological University (2021), INDIA. He currently works as a research project intern at the Physical Research Laboratory (PRL), Department of Space, Government of India, Ahmedabad. His research interests include Nanoelectronics, FinFET based biosensing technology, and astrophysics.

Rishu Chaujar (SM'16) is currently working as a Professor with the Department of Applied Physics, Delhi Technological University (formerly the Delhi College of Engineering), New Delhi, INDIA. She has co-authored over 200 papers in various reputed international and national journals and conferences. Her research interest includes FETs, modeling of semiconductor devices, molecular bio-physics, nano-electronics, etc.

CRediT authorship contribution statement

Mridul Prakash Kashyap: Methodology, Software, Investigation, Writing – original draft, Writing – review & editing. **Harshal Gudaghe:** Data curation, Conceptualization, Visualization. **Rishu Chaujar:** Supervision, Writing – original draft.

Declaration of Competing Interest

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

ACKNOWLEDGEMENT

The authors are dedicatedly thanking the authorities of Engineering Physics department of Delhi Technological University for all its

financial support provided while conducting this research work.

References

[1] H. Gudaghe, M. Kashyap and R. Chaujar, "Investigation of TF-FinFET based biosensor for early diagnosis of protein carrying diseases," *2021 Devices for integrated circuit (DevIC),*2021, pp. 45–49, doi:10.1109/DevIC50843.2021.9455938.

[2] Kaisti M. Detection principles of biological and chemical FET sensors. Biosens Bioelectron 2017;98:437–48. https://doi.org/10.1016/j.bios.2017.07.010. Dec.

[3] Sani MH, Khosroabadi S. A novel design and analysis of high-sensitivity biosensor based on nano-cavity for detection of blood component, diabetes, cancer and glucose concentration. IEEE Sens J 2020;20(13):7161–8. https://doi.org/10.1109/JSEN.2020.2964114. 1 July1,.

[4] Mukhopadhyay S, Sen D, Goswami B, Sarkar SK. Performance evaluation of dielectrically modulated extended gate single cavity InGaAs/Si HTFET based label-free biosensor considering non-ideal issues. IEEE Sens J 2021;21(4):4739–46. https://doi.org/10.1109/JSEN.2020.3033576. 15 Feb.15,.

[5] Avvakumova, Svetlana & Colombo, Miriam & Mazzucchelli, Serena & Rotem, Rany & Prosperi, Davide. (2018). "Bioengineered approaches for site orientation of peptide-based ligands of nanomaterials". doi:10.1016/B978-0-323-50878-0.00006-9.

[6] Basu J, RoyChaudhuri C. Graphene nanoporous FET biosensor: influence of pore dimension on sensing performance in complex analyte. IEEE Sens J 2018;18 (14):5627–34. https://doi.org/10.1109/JSEN.2018.2841060. 15 July15,.

[7] Massey R, Bebe S, Prakash R. Aptamer-enhanced organic electrolyte-gated FET biosensor for high-specificity detection of cortisol. IEEE Sensors Lett 2020;4(7): 1–4. https://doi.org/10.1109/LSENS.2020.3002446. JulyArt no. 4500604.

[8] Gonzalez JL, Jo PK, Abbaspour R, Bakir MS. A disposable and self-aligned 3-D integrated bio-sensing interface module for CMOS cell-based biosensor applications. IEEE Electr Device Lett 2018;39(8):1215–8. https://doi.org/10.1109/LED.2018.2851969. Aug.

[9] Bergveld. Development of an ion-sensitive solid-state device for neurophysiological measurements. IEEE Trans Biomed Eng 1970;(1):70–1. https://doi.org/10.1109/TBME.1970.4502688. vol. BME-17, no.Jan.

[10] Buvaneswari B, Balamurugan NB. 2D analytical modeling and simulation of dual material DG MOSFET for biosensing application. AEU - Int J Electr Commun 2019;99:193–200. https://doi.org/10.1016/j.aeue.2018.11.039. Jan.

[11] Kim S, et al. Multiplexed silicon nanowire tunnel FET-based biosensors with optimized multi-sensing currents. IEEE Sens J 2021;21(7):8839–46. https://doi.org/10.1109/JSEN.2021.3054052. 1 April1.

[12] Kuo P-Y, Chen Y-Y. A novel low unity-gain frequency and low power consumption instrumentation amplifier design for RuO_2 uric acid biosensor measurement. IEEE Trans Instrum Meas 2021;70:1–9. https://doi.org/10.1109/TIM.2021.3060571.

[13] Anand S, Singh A, Amin SI, Thool AS. Design and performance analysis of dielectrically modulated doping-less tunnel FET-based label free biosensor. IEEE Sens J 2019;19(12):4369–74. https://doi.org/10.1109/JSEN.2019.2900092. 15 June15.

[14] Kashyap MP, Chaujar R. Gate oxide variability analysis of a novel 3nm truncated Fin–FinFET for high circuitry performance. Silicon 2021;13:3249–56. https://doi.org/10.1007/s12633-020-00734-5.

[15] ITRS 2.0 Publication, "International technology roadmap for semiconductors, 2015," Available online at http://www.itrs2.net/.

[16] David G. Castner, Buddy D. Ratner, Chapter 31 – "Proteins controlled with precision at organic, polymeric, and biopolymer interfaces for tissue engineering and regenerative medicine", *Principles of regenerative medicine (Third edition), academic press*, 2019, pp. 523–34, doi:10.1016/B978-0-12-809880-6.00031-X.

[17] Munguía-Cortés Laura, Hermosillo Pérez, Jhonnatan Isaac, López Ojeda-. Reyna & Esparza-Schulz, Juan & Felipe, Carlos & Cervantes- Uribe, Adrián & Domínguez, Armando, "APTES-Functionalization of SBA-15 Using Ethanol or Toluene: textural Characterization and Sorption Performance of Carbon Dioxide. J Mex Chem Soc 2018. https://doi.org/10.29356/jmcs.v61i4.457.

[18] Neupane DP, Avalos D, Fullam S, Roychowdhury H, Yukl ET. Mechanisms of zinc binding to the solute-binding protein AztC and transfer from the metallochaperone AztD. J Biol Chem 2017;292:17496–505. https://doi.org/10.1074/jbc.M117.804799.

[19] A. Kumar, M. Tripathi and R. Chaujar, "Sub-30nm In2O5Sn gate electrode recessed channel MOSFET: a biosensor for early stage diagnostics," Vacuum, vol. 164, 2019, doi:10.1016/j.vacuum.2019.02.054.

[20] Wasik Daniel, Mulchandani Ashok, Yates Marylynn V. Salivary detection of dengue virus NS1 protein with a label-free immunosensor for early dengue diagnosis. Sensors 2018;18(8):2641. https://doi.org/10.3390/s18082641.

[21] Dhama K, Latheef SK, Dadar M, Samad HA, Munjal A, Khandia R, Karthik K, Tiwari R, Yatoo MI, Bhatt P, Chakraborty S, Singh KP, Iqbal H, Chaicumpa W, Joshi SK. Biomarkers in stress related diseases/disorders: diagnostic, prognostic, and therapeutic values. Front Mol Biosci 2019;6(91). https://doi.org/10.3389/fmolb.2019.00091.

[22] P.H. Drive, S. Clara, Atlas user's manual, device structure editing and simulation software; 2017. Pp. 1–1715. Available: www.silvaco.com.

[23] Sachid AB, Chen M, Hu C. Bulk FinFET with low-K spacers for continued scaling. IEEE Trans Electr Devices 2017;64(4):1861–4. https://doi.org/10.1109/TED.2017.2664798. April.

[24] Sehgal HD, Pratap Y, Gupta M, Kabra S. Performance analysis and optimization of under-gate dielectric modulated junctionless FinFET biosensor. IEEE Sens J 2021;21(17):18897–904. https://doi.org/10.1109/JSEN.2021.3090263. Sept.1.

[25] Singh D, Patil GC. Performance analysis of feedback field-effect transistor-based biosensor. IEEE Sens J 2020;20(22):13269–76. https://doi.org/10.1109/JSEN.2020.3006986. 15 Nov.15.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/358093497

Comprehensive Analysis of Design Software Application in Solar Distillation Units

Article *in* Materials Science for Energy Technologies · January 2022

DOI: 10.1016/j.mset.2022.01.005

CITATIONS
0

READS
47

5 authors, including:



Om Prakash
Birla Institute of Technology, Mesra
42 PUBLICATIONS **1,038** CITATIONS
SEE PROFILE



Asim Ahmad
Birla Institute of Technology, Mesra
15 PUBLICATIONS **40** CITATIONS
SEE PROFILE



Anil Kumar
Delhi Technological University
256 PUBLICATIONS **4,286** CITATIONS
SEE PROFILE



Gaurav Kumar
K. K. University
8 PUBLICATIONS **4** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Solar desalination View project

modelling, simulation and environomical analysis of domestic direct type multi-shelf solar dryer View project

All content following this page was uploaded by Gaurav Kumar on 15 February 2022.

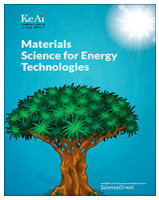
The user has requested enhancement of the downloaded file.

Materials Science for Energy Technologies 5 (2022) 171–180

Contents lists available at ScienceDirect

Materials Science for Energy Technologies

journal homepage: www.keaipublishing.com/en/journals/materials-science-for-energy-technologies



Comprehensive analysis of design software application in solar distillation units



Om Prakash [a,*], Asim Ahmad [b,*], Anil Kumar [c,*], S.M. Mozammil Hasnain [d], Gaurav Kumar [d]

[a] Department of Mechanical Engineering, Birla Institute of Technology, Mesra, Ranchi 835215, India
[b] Faculty of Engineering and Applied Sciences, Usha Martin University, Ranchi 835103, India
[c] Department of Mechanical Engineering, Delhi Technological University, Delhi 110042, India
[d] Department of Mechanical Engineering, K. K. University, Nalanda, 803115 Bihar, India

ARTICLE INFO

Article history:
Received 25 October 2021
Revised 17 January 2022
Accepted 19 January 2022
Available online 24 January 2022

Keywords:
FORTRAN
CFD
Simulation
Solar Still
MATLAB
FLUENT
COMSOL

ABSTRACT

The use of solar stills in rural regions are becoming increasingly popular as it is an economical solution for drinking water from saline water sources. Many researchers have worked for the improvement of conventional solar still to enhance productivity. Costly and time-consuming processes of operation in solar stills encourage many scholars to analyze mathematical simulation. This paper presents comprehensive reviews of the application of different design software to solar still systems. Design software is essential for developing and analyzing the mathematical models and predicting the most suitable performance parameters for the enhanced production rate of distilled water for still systems. Numerical modeling of solar still systems is necessary to analyze and investigate air movement, temperature variation for knowing water temperature, and air temperature through software like CFD, MATLAB, FORTRAN, TRYNSYS AutoCAD. The simulation technique's application using CFD is made with TRNSYS, FLUENT, ANSYS, FORTRAN and MATLAB which are useful tools to develop such mathematical models for the prediction of flow parameters. Engineering Equation Solver (EES) package and COMSOL Multiphysics solve the differential energy balance equation. All newly developed software employed for the utility of still solar systems is discussed. This article provides a comprehensive overview of the various software tools used in solar still to help researchers, scientists, and academicians.
© 2022 The Authors. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Freshwater is considered as the basic need for the survival of human beings and other living organisms. Earth contains 71% of total water, out of which 2.5% of water is freshwater. Only 1.04% of freshwater is available for human beings because of groundwater and ice availability. Also, 7.6 billion population in the world depends on this freshwater for domestic needs [1]. Uncontaminated freshwater is reducing day by day due to an increase in pollution. Due to the inadequate availability of water, health issues are prevailing among living beings. Hence it is essential to save freshwater for human beings. Many developing countries face an acute shortage of portable drinking water due to the rapid population increase. In rural areas, drinking water infrastructure is underdeveloped and incapable of meeting various domestic purposes, drinking purposes, and industrial purposes. Solar energy is feasible and economical for all solar-based appliances [2–3].

Solar distillation is a method of extracting drinking water from saline water using solar radiation. Solar distillation process is used to produce the solar still. It provides safe drinking water for humans. The condensation and evaporation phenomena are used in this distillation process [4]. Commercial water distillation processes are not practicable in rural regions due to the high cost of electricity or fossil fuel. As a result, solar still continues to be a viable option for providing clean water in rural areas. [5].

Natural convection solar stills, also known as passive solar stills, and forced convection solar stills, also known as active solar stills, are the two forms of solar stills [6–7].

A natural convection solar still works on the principle of natural evaporation and condensation process. Evaporation occurs due to the greenhouse effect. Water vapor gets condensed at the inner glass cover [6]. The glass slants down towards the distillate channel and then to the dilute beads roll and gathered in the compart-

* Corresponding author.
E-mail addresses: omprakash@bitmesra.ac.in (O. Prakash), asimlife91@gmail.com (A. Ahmad), anilkumar76@dtu.ac.in (A. Kumar).

https://doi.org/10.1016/j.mset.2022.01.005
2589-2991/© 2022 The Authors. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co., Ltd.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

ment [7–8]. The above-discussed system's schematic diagram is shown in Fig. 1.

Forced convection still differs from natural convection still because it forces air and vapor mixture motion inside the solar still leads to more heat transfer as compared to natural convection[9–10]. It enhances the heat and mass transfer coefficients of the convection and evaporation process. A schematic outline of an active type of solar still is shown in Fig. 2 [11].

Modelling is an essential tool that enhances the productivity of distilled water by improving prototypes. Khare et al. used "FLUENT-ANSYS" for passive solar distiller. The mass fraction of distilled water was 2.0 in 14 h and 0.9 at 8 h. Experimental and simulated maximum temperature for water liquid was 72 °C and 87 °C [8].

Tripathi and Tiwari have used MATLAB to analyze the convective coefficient. Output of distilled water was 0.856 kg/m^2 per day [4]. Prakash et al. has found the efficacy and performance of solar distillation technology. They also analyzed the research of various authors working on solar desalination technology [5].

CFD analyzes the flow pattern of air (and moisture), temperature variation, humid zone, and stress on all adjacent walls. CFD also helps in analyzing the flow pattern and behavior of condensing cover and adjacent wall. The various zones of a solar still where evaporation and condensation take place, such as the evaporation zone, condensing zone, and moisture zone, are depicted using CFD analysis. The thermal efficiency can be achieved in a solar still simulation by changing geometric and physical parameters such as cooling water velocity, film cooling thickness, and water glass cooling thermal characteristics.

The productivity of a vertical multi-effect diffusion solar still (VMEDSS) combined with a solar tracking system and a photovoltaic/thermal solar water heater system (PV/T-SWH) was investigated by Boubekri et al. [13]. The findings demonstrate the impact of the solar tracking system on the rise in still production over the summer and winter seasons. Summer production increased by 293.01%, while winter production increased by 15.10%.

Rahdan et al. [14] used CFD to simulate and optimise a hybrid solar chimney and water desalination system by focusing on the desalination. The authors analysed the impact of optimal geometric parameters on power generation and freshwater output. According to the findings, a divergent chimney may boost average air velocity from 17% to 23% and enhance electricity generation from 64% to 95%. In addition, models with divergent solar collectors might create nearly twice as much freshwater as models with parallel collectors.

Shoeibi et al. [15] use a hybrid nanofluid (Al_2O_3-TiO_2) at a concentration of 0.4 % to numerically analyse the glass cooling of a double-slope solar still. A pump circulates a hybrid nanofluid (Al_2O_3-TiO_2) glass cooling fluid over the glass to lower the cover temperature and increase freshwater generation. Hybrid nanofluid cover cooling increases the temperature difference between the evaporation and condensation zones. The use of hybrid nanofluid glass cooling improves the solar still's natural convection heat transmission. The solar's glass cover and water surfaces still have boundary conditions of convective heat transfer coefficient and constant temperature, respectively. As hybrid nanofluid glass cooling is utilised, the freshwater yield and energy efficiency of solar desalination increase by 11.09% and 28.21%, respectively, compared to solar desalination without hybrid nanofluid. According to the CFD simulation, the ideal nanoparticle concentration was 0.45%.

2. Simulation techniques of different solar stills

The simulation occurs on both the types of solar still i.e. natural convection solar and forced convection solar still by using different programming language as follows:

2.1. Simulation for natural convection solar still

2.1.1. CFD simulation

Khare et al. have developed a 3D model for CFD analysis to understand the condensation and evaporation phenomenon in sin-

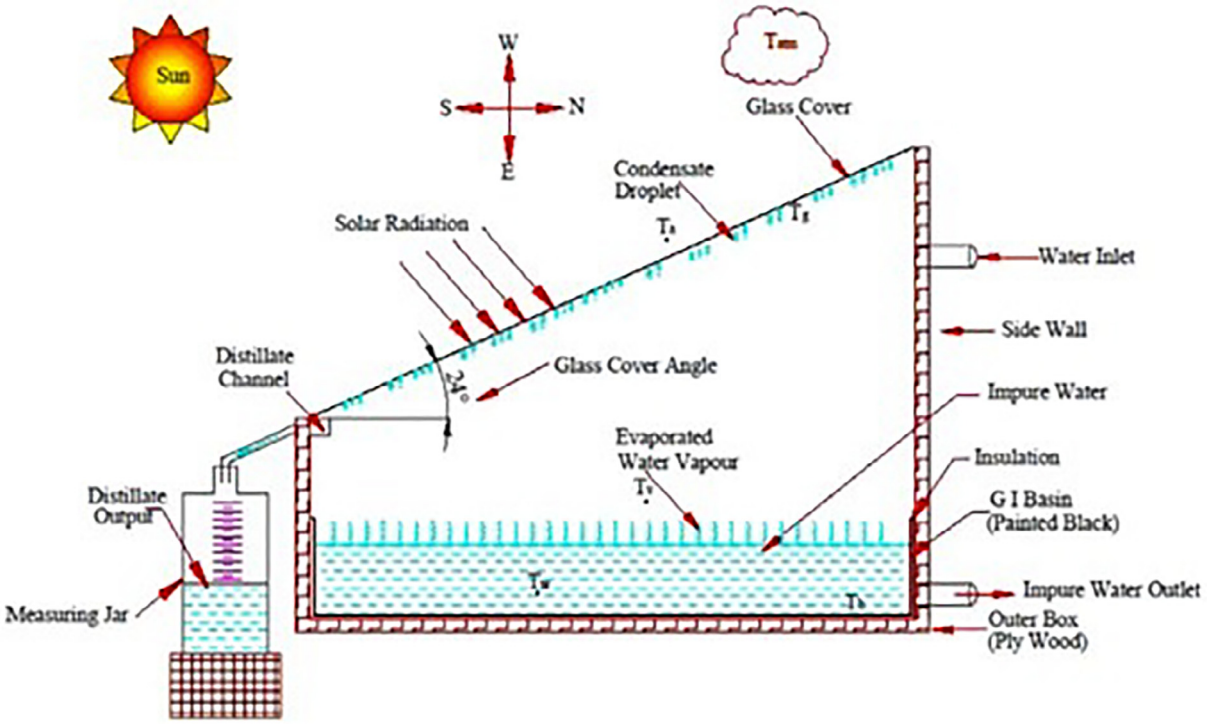


Fig. 1. Schematic outline of single basin single slope solar still [1].

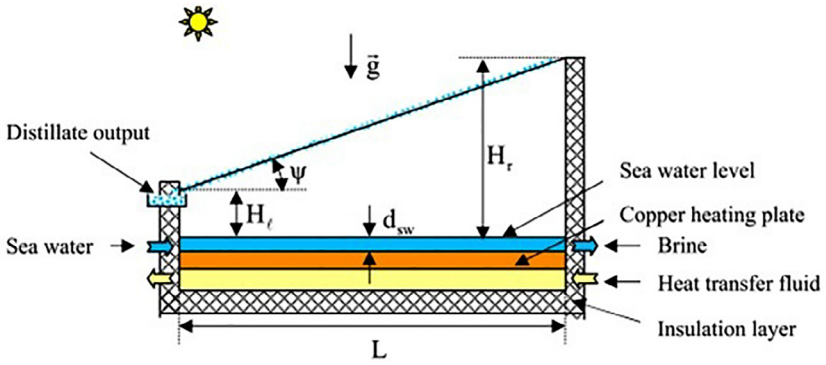


Fig. 2. Schematic outlines of Active type solar still [12].

gle slope solar still operating under passive mode[8]. ANSYS workbench was used to develop the model, and then it is simulated with the help of Fluent. The study parameters were solar radiation, basin material and water depth. The meshing geometry of model was shown in Fig. 3, and simulation results were shown in Fig. 4. The simulation result shows that the reflecting mirror used in the solar still increases production rate by 22%. The simulated result validates the experimental data.

Rahbar et al. worked on 2D CFD simulation for tubular solar still to estimate heat and mass transfer coefficients in ANSYS-FLUENT 14.0 [9]. CFD simulation indicates a recirculation region of water vapors with a clockwise heading inside the still. The simulation result is shown in Fig. 5. The analysis showed that most condensation takes place towards the inner glass cover surface. It can be concluded from the result that on increasing glass temperature, the condensation rate slows down. It also shows that as water temperature increases, it increments the performance of solar still productivity. As the difference between water and glass temperature increases, the Rayleigh number (Ra) increases flow strength. Accuracy between the experimental and computational result of the convective heat transfer coefficient and water productivity between glass and water was measuredea as 10% and 18%.

Shakib et al. used CFD FLUENT software for 3D modelling to study fluid flow behavior in natural convection solar still [10]. Parameters considered for evaluation in the simulation process were velocity, temperature profiles, circulation of shear stress, and heat transfer coefficients. The geometry portion is created and meshed in Gambit software. The fine mesh contains cells of about 1×10^5, and Ra value for simulation was 1000. Simulation results show that the velocities are lower on the plane, which is close to the wall. Analysis of heat transfer coefficient was done for both lower and higher heights for a vertical wall. It shows that heat transfer coefficient values for the higher vertical wall are high compared to lower vertical height due to more temperature gradients across it. It was also seen that the shear stress values are higher for the bottom and top center of the still system. There was a fair agreement of the simulation result with the experimental result. The simulation is shown in Figs. 5-8.

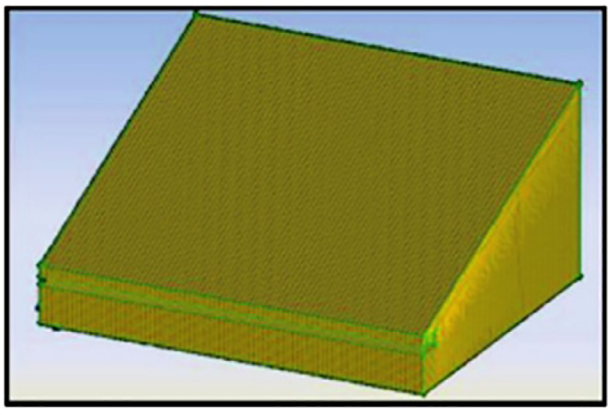


Fig. 3. Meshing geometry of model [9].

2.1.2. ANSYS CFX simulation

Singh and Mittal have simulated single slope solar still operating in passive mode to analyze the best inclination angle to enhance the productivity rate. The simulation was carried out with ANSYS CFX 13 [11]. For the simulation process in the solar still, two condensing glasses with a slope of 15° and 30° were chosen. The temperature differential in each reading was between 40 and 60 °C, with a 2 °C delay. The simulation result is shown in Fig. 9. Adhesive forces were taken for the formation of a droplet on the condensing cover. It was observed that the condensing cover at an angle 30° produces a high evaporative and convective heat transfer coefficient by 29.4% from a 15° slope angle.

Different depth of solar still was optimized by Thakur et al. using computational method. [12]. Work was accomplished for single slope solar still operating under passive mode for a water depth of 0.01 m, 0.02 m and 0.03 m. Meshing geometry has several nodes i.e., 632,088 and elements were 553048. Hex-dominant was used for meshing. The result produced by CFD analysis shows a good agreement between simulated data and experimental data. It was also found that the optimum depth was to be 0.01 m for a good agreement.

Rahbar and Esfahani did the computational work for the CFD modelling and its validation for single slope solar still [16]. The design of the model was created with ANSYS workbench in design modeller in 3-D. The model meshed with a hexahedral type of meshing. The optimum number of elements for the domain was 100229. The element skewness and aspect ratios were 0.3 and less than 2, respectively. Finite volume method was used to solve energy equations in the CFD work turbulence model of spalart - Allmaras was used [17].

2.1.3. Autocad

Thermal analysis of single slope solar still operating in passive and active mode was done consedering different solar fraction parameters by Tripathi and Tiwari [4]. AUTOCAD 2000 was used in single slope solar still for 3D model geometry. The glass cover slope was at an angle of 10.2° and the basin area of the system was 1 m × 1 m. Experimentation was done under New Delhi, India's clear sky weather condition (Longitude 77°12′E, latitude 28°35′N). MATLAB program was used to analyze the evaporative and convective heat transfer coefficient. Fig. 10 represents the

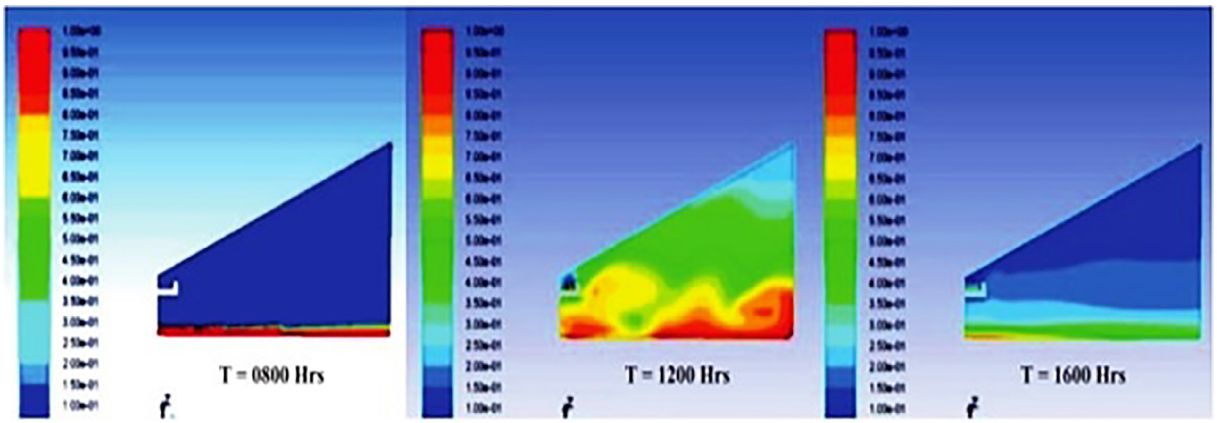


Fig. 4. General Volume Fraction of Water at a specific Time [9].

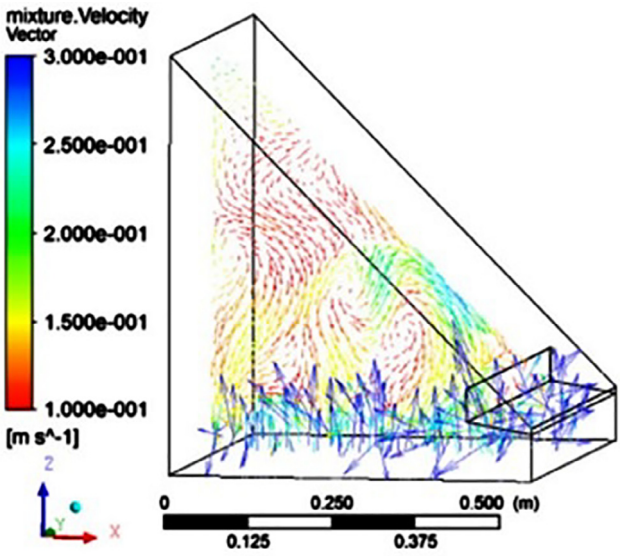


Fig. 5. Gas mixture velocity on a plane inside the solar still [7].

workflow chart of the AUTOCAD 2000 model. The results show that the theoretical and experimental approaches to heat transfer coefficient are in good agreement.

2.1.4 vol. of fluid (VOF) simulation

Rashidi et al. performed VOF simulation work for the entropy generation in passive-type single slope solar still[18]. This simulation work has been focused on the evaporation and condensation phenomenon in solar still with nano-fluids flow. For the solution of these equations, the finite volume method (FVM) opted. Staggered meshing was done, and simulation was carried out in ANSYS fluent. Contours were developed using simulation for viscous entropy generation and thermal entropy generation. It was noticed that about 95% and 25% entropy generation increase if the substantial volume fraction of Nano-fluids increases by about 5%. The VOF model results are compared with experimental data, and the VOF model is found to have good accuracy.

2.1.5. MATLAB programming

Tiwari and Tiwari have studied the effect of water depth on evaporative mass transfer coefficient in passive single slope solar still [19]. A computer program was written in MATLAB version 7.0 software to solve the thermal model. Solutions of the equations give the values of hourly convective heat transfer coefficients (h_{ow}), evaporative heat transfer coefficients(h_{ew}), Dunkle convective heat transfer coefficients (h_{cwDUNK}) and Dunkle evaporative heat transfer coefficient ($h_{ew\ DUNK}$). Due to Dunkle's constraints and the assumption made during the construction of the energy balance equation, the matrix solution findings differ significantly from experimental data.

The numerical study of a double slope solar still was done by Ali et al. [20]. This is undergoing a special improvement to make it more efficient. The brackish water, which is made up of granular activated carbon, is used. The MATLAB software is used to calculate various heat transfer calculations. The entire analysis was carried out using graphical solutions that were plotted. The Simulink toolkit is used to create a representation of the simulation model. The MATLAB language was coded using M−files of ASCII text.

2.1.6. SIMPLEC algorithm

Edalatpour et al. did research work for naturally circulated single slope solar still [21]. For the enhancement of heat transfer, vortex is generated inside the solar still. Numerical modelling was done for the still system using the energy balance equations. Numerical equations were solved with SIMPLEC (semi-implicit method for pressure linked equations consistent) algorithm. Vortex generation was build up in CFD work using a FLUENT package. Model creation for solar still was two dimensional. Steady-state, laminar, incompressible, ideal conditions were assumed for the simulation process grid size was kept 7216 elements of a non-uniform square-shaped. All the non-dimensionless numbers (Nu, Ra) were estimated for the solar still with and without blade material. The result shows that the SIMPLEC method is an excellent algorithm for the solution technique of energy equations and estimating process parameters.

2.1.7. Fortran

El-Samadony et al. used FORTRAN programming language to develop a computer simulation program for passive-type stepped solar still [22]. This research aimed to investigate the effect of the radiation shape factor on still productivity and radiation heat transfer within the solar still. The result shows that when the slant angle of the glass cover increases, the radiation form factor decreases, resulting in an improvement in productivity of up to 18.8%. It was observed that the maximum deviation of 3.16% occurred with productivity as parameters were solar insolation and glass inclination. It can also be concluded that when the value of radiation shape factor approaches 1, then the experimental and theoretical values are validated.

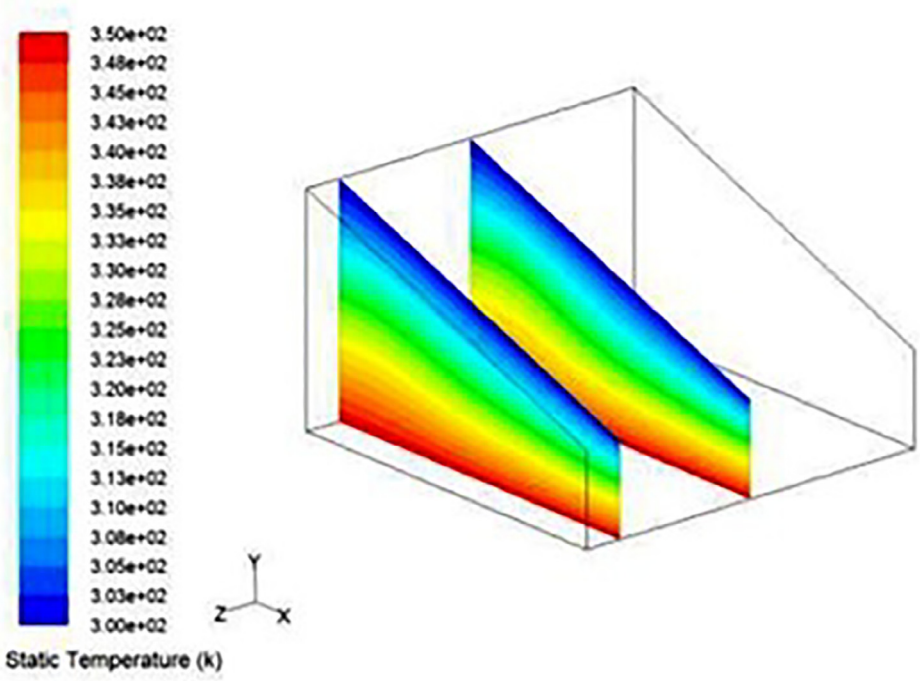


Fig. 6. Temperature profiles on two planes [8].

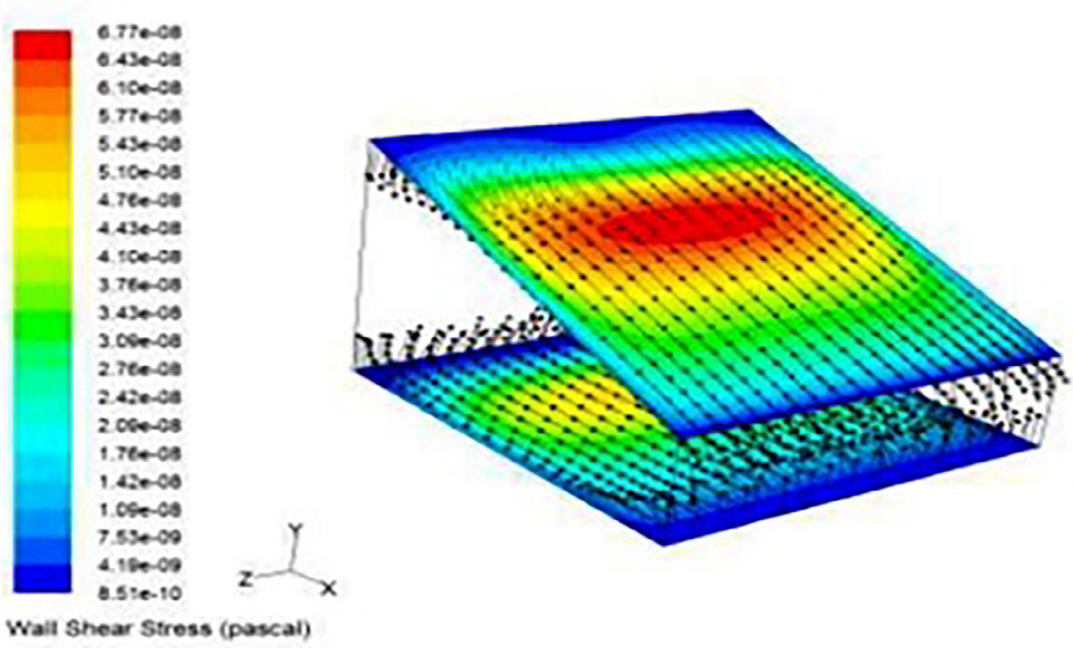


Fig. 7. Shear stress on top and bottom surfaces [8].

Ghoneyem et al. worked on the thickness of the glass cover of solar still to analyze the production rate [23]. A variation of 15% has been observed from the experimental result and numerical solutions. It was observed that glass cover thickness 3 mm has more efficiency (i.e., 26.22%), compared to the glass of thickness 5 mm and 6 mm.

The simulation was carried out by Adhikari et al. in a multi-stage stacked tray solar still. The simulation was carried out in a steady-state environment [24]. The FORTRAN-77 programming language was used to assess steady-state water temperature and produce results. The results suggest that simulation parameters can be utilised to analyse the performance of a solar still.

2.1.8. TRNSYS simulation

Chaibi developed the simulation model for the generation of distilled water and analyzed double slope solar still performance (incorporated in the rooftop of a greenhouse) operating under passive mode [25]. TRNSYS (Transient system simulation) method was used to analyse visual material properties and solar irradiation. Consequence of solar irradiation are a consideration in the system.

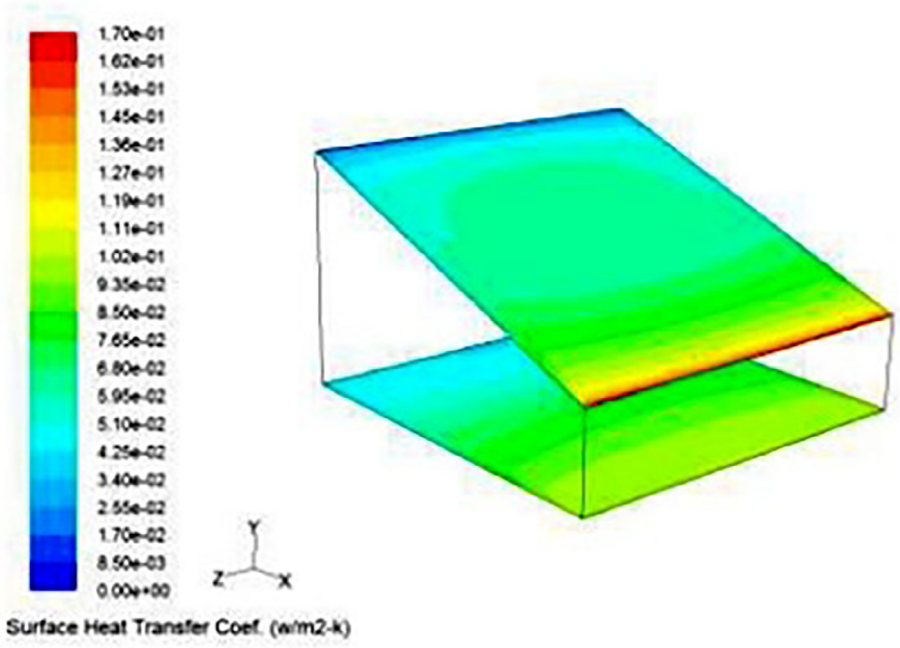


Fig. 8. Distribution of local heat transfer coefficients [8].

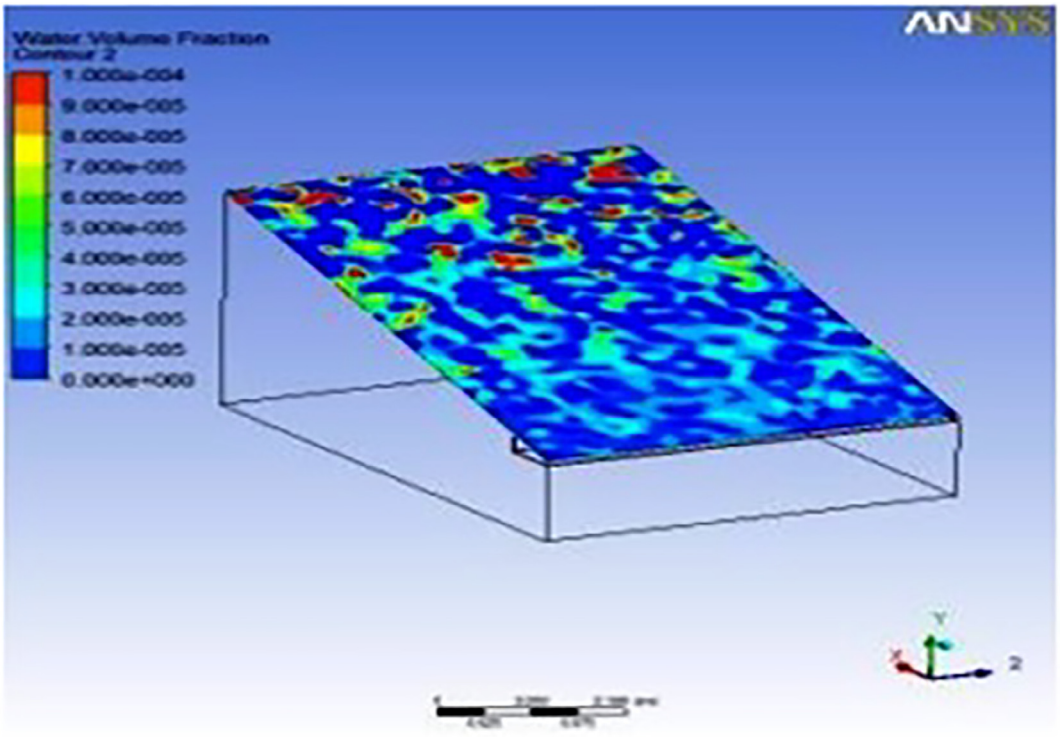


Fig. 9. Condensed water droplets on the glass as the water heated and water volume fraction on the glass for 15° inclinations [9].

Engineering equation solver (EES) was used to solve energy balance equation. It was observed that solar radiation affected the efficiency of the rooftop integrated greenhouse.

2.2. Simulation for forced convection solar still

2.2.1. MATLAB programming

Ibrahim et al. worked on the stepped type active solar still assembled with an air-cooled condenser to modify solar still with MATLAB coding [26]. Parameters considered for investigation were the water depth of the basin, wind speed, the slant of the glass cover, the absorptivity of the base, and glass cover thickness. The simulation result shows that if wind velocity increases by 1 to 5 m/s, it gives a productivity decrement of 3%. The increased inclination angle of a solar still gives higher solar radiation. Experiment was done under the climatic states of Cairo, Egypt (30° 05′N, 31° 25′E) with south-facing. Results of simulations are very close to experimental data.

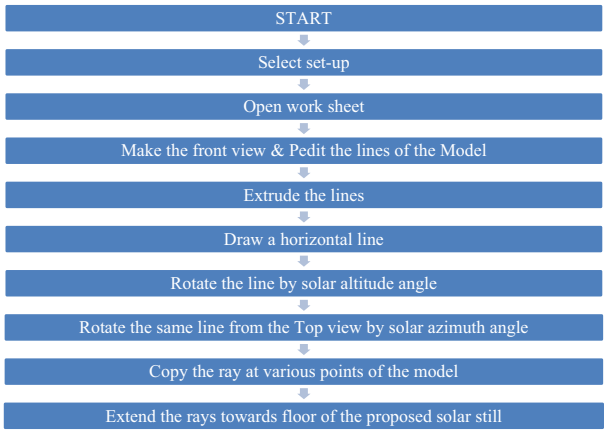


Fig. 10. Work flow chart of the AUTOCAD 2000 [11].

Halima et al. created a thermal model for a compression heat pump combined with an active type single slope solar still [27]. The mass and energy balance equations were used to create the mathematical model. According to the results, the still's profitability is 75 % higher than that of traditional passive solar still. The fourth-order Runge-Kutta(R-K) method is being utilised to solve the differential problem. The results of the model were compared to experimental data. It was discovered that they have a reasonable agreement.

The mathematical modelling of a single slope solar still with a solar air heater was done by Kumar and Kumar [28]. For each component of the solar still, such as the glass cover, basin water, basin liner, and bottom surface, an energy balancing equation was constructed. The equations were solved iteratively using the MATLAB programming language. This software was used to determine the glass temperature, water temperature, basin temperature, and heat transfer. MATLAB is a must-have programme to calculate the equations generated for heat transmission inside the solar still.

2.2.2. COMSOL MULTIPHYSICS simulation

Abed et al. worked for the desalination process in tubular type passive solar still [29]. Theoretical modelling was done to estimate glass cover temperature, water surface temperature, humid air temperature, and production yield of the solar still system. Comsol Multiphysics software-5.0 was used to solve the energy balance problem. The whole computations were done for two cases, i.e., with outside wall insulation and another one with insulation. It was beneficial to predict the different temperatures of solar still system. To verify the experimental result, the validation with the CFD results with the Rahbar theoretical result. It was noticed a fair agreement between them.

2.2.3. Simulation error/deviation from experimental results

Khare et al. and Shakib et al. .worked using CFD/Fluent simulation software in a solar still[8–9]. It was seen as a fair agreement of simulation results with the experimental data. An average error of 9.98% was observed for a water temperature between simulation and experimental data[9]. Singh & Mittal CFD work has an error of 8% for average temperature between experimental and simulation results[11]. Tripathi and Tiwari work has a coefficient of correlation between 0.91 and 0.99, which shows fair agreement between experimental and simulation work[4].

Functions of design software, applications, benefits, and limitations are listed in Table 1. Simulation error/deviation from experimental results and solar still type studied by various authors are given in Table 2. Errors were having some variation/deviation in

Table 1
Application and benefits/ limitations of different software used in solar still.

Authors	Modelling software	Application	Function	Limitation/Benefits
Singh & Mittal [11]	ANSYS CFX	ANSYS CFX was used in single slope solar still to develop a 3D two-phase model for evaporation and condensation.	ANSTYS CFX is a powerful, pliable CFD software used to analyze a large range of fluid flow issues and the issues related to fluid and heat transfer behavior inside the solar still	In CFX, the grid quality is more admissible and the solver is almost in lock mode.
Khare et al. [8]Rahbar et al.[9] Shakaib et al.[10]	CFD Fluent/Ansys	The forecast of exact size and shape can save money and be done quickly by reducing extensive exercise and rehashing fabricating.	The simulation software CFD FLUENT provides precise information on heat transport and fluid flow characteristics within the solar still.	Learning software and Geometry meshing are time-consuming.
Tiwari & Tiwari[19] Ibrahim et al.[26] Halima et al. [27]	MATLAB	The software helps develop numerical models to analyze yield, humidity and glass, and water temperature.	MATLAB is programmable and modelling software. It takes less time to provide solution of non-linear differential equations.	MATLAB mathematical modeling takes long time to create and test the model and it requires magnificent programming skills.
Tripathi & Tiwari[4]	AUTO CAD	The solar distribution factor is determined by performing a thermal analysis of active and passive solar distillation systems.	Auto-CAD is 3-D and 2-D drafting software.	It requires particular computer skill and expensive equipment.
Chaibi & Illeri [25]	TRNSYS	Its help in replacing differential equation into easy numerical calculations. This numerical calculation helps in determining heat transfer rate and humidity inside solar still.	TRNSYS is a scientific tool used for simulation universally in the field of solar energy. This software is used to portray and create solar still.	It requires a shorter time to provide accurate results.
El-Samadony et al.[22]	FORTRAN	It helps to do the performance analysis of solar still. It minimizes material usage. It reduces expensive trial, error and time.	FORTRAN helps in the modeling and simulation to figure out the partial differential equations with a computer program's help.	The FORTRAN software needs to be developed first in a prototype mode like IDL (Interactive Data Languages) and Matlab and then it can be run in FORTRAN.
Maalem et al. [17]	COMSOLMULTIPHYSICS	It anticipates the moisture air movement through glass to basin surface. It was also used to predict the measurement and shape of solar stills.	This simulation software provides the data on fluid flow and heat transfer profiles inside the solar.	The learning of this software is easy as compared to FLUENT- CFD.

Table 2
Simulation Error/Deviation from experimental results.

Simulation Software	Authors/ Researchers	Type of Solar Still	Percentage Error/ Deviation
CFD/Fluent	Khare et al. [8]	Passive SSSS	No error
	Rehbar et al.[9]	Passive Tubular solar still	18%
	Shakaib[10]	Passive SSSS	30%
ANSYS CFX	Singh and Mittal[11]	Passive SSSS	11.8%
AUTO CAD	Tripathi and Tiwari[4]	Active SSSS	0.99%
MATLAB	Tiwari and Tiwari[19]	Passive SSSS	3.66%
	Halima et al. [27]	Active SSSS	20%
Comsol Multiphysics	Maalem et al. [17]	Passive trapezoidal solar still	7%
FORTRAN	El-Samadony et al. [22]	Passive Stepped solar still	3.16%
TRNSYS	Chaibi[25]	Passive Double slope solar still	No error

productivity, a basin water temperature of solar stills. Table 3 shows the boundary and geometry conditions used by various researchers for simulation.

3. Performance analysis of solar distillation

In the present work, heat and mass transfer is being considered. For each phase volume-average continuity and time are numerically adjusted. In this analysis, the efficiency of the solar still system needs to be calculated. Efficiency of solar still is defined as the ratio between the total amount of energy used to obtain distilled water (in a specific duration) and the amount of energy supplied to the solar still during the same time duration. Efficiency equation is given in Eqn. (1) and (2) [30].

$$\text{Efficiency} = \frac{\text{Yield of 24 hours} \times \text{Latent heat of Evaporisation}}{\text{Incident energy on collector} + \text{Incident energy on solar still}} \times 100\% \quad (1)$$

$$\text{Efficiency} = \frac{\sum m_{ew}^{\cdot} \times L_v}{\sum I_g \times A_{collector} \times 6 \times 3600 + \sum I_g \times A_{glass} \times 6 \times 3600} \times 100\% \quad (2)$$

Model equations based on the momentum, continuity, mass and energy transfer conservation principles at steady state condition follow:

For Gas:

$$\nabla(r_G \rho_G V_G) + S_{LG} = 0 \quad (3)$$

For Liquid:

$$\nabla(r_L \rho_L V_L) - S_{LG} = 0 \quad (4)$$

4. ANN modelling in solar Still: A case study

This case study aims to determine the efficacy of employing artificial neural networks to model a solar still's distillate output (ANNs). By using fewer elements than (heat and mass transfer) HMT models, ANNs have a potential advantage in foreseeing the execution of solar stills [31]. The ANN model will assist in determining the impact of each input variable on the performance of the still model.

ANNs are data processing frameworks that are non-algorithmic, non-computerized, and concentrated parallel. ANNs comprise many neurons (preparing components), ordered in various layers in the system, such as output layers, hidden layers, and input layers. The neuron in the system gets input signals, sense them, and then sends an output signal. The fundamental steps for planning a neural network model are information gathering, pre-handling information, building system, preparing system, testing system, and system approval. The input parameters for the ANN model network were daily insolation, ambient temperature, wind velocity, cloud cover temperature, and wind direction [32]. A schematic outline of the ANN design model is shown in Fig. 11.

5. CFD modelling in solar Still: A case study

García-Chávez et al. 2020 [33] have presented the design, behavior and modeling of solar still using CFD. Different parameters like internal temperature, volume and cover help to analyze the desali-

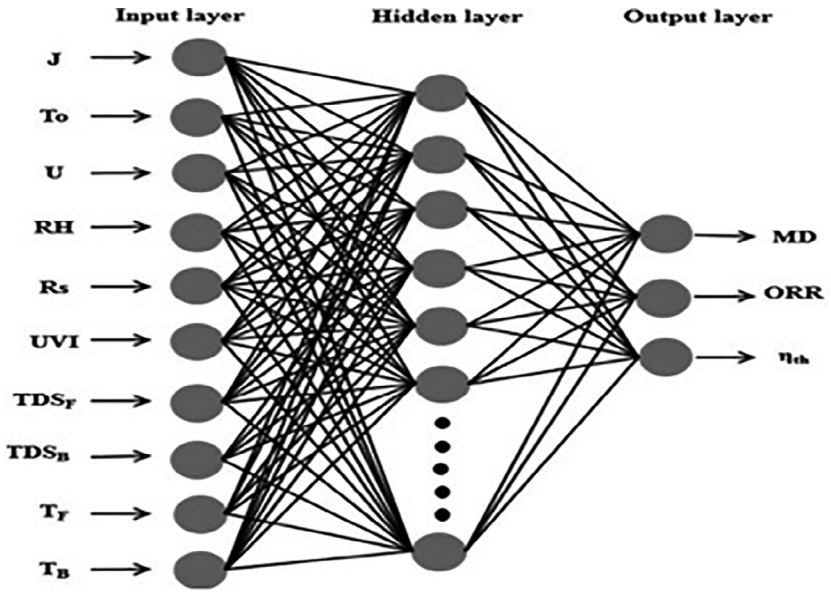


Fig. 11. Architecture based ANN design model [17].

Table 3
Boundary conditions for simulation by various researchers.

Simulation Software	Authors/ Researchers	Boundary Conditions
CFD/Fluent	Khare et al. [8]	ANSYS Design Modeler was used to model the 3D geometry of the solar still, including all geometrical limitations. The physical model of the building was then meshed in ANSYS's workbench MESHING utilising 3D hexahedral meshing, which has a total of 1.5 million cells (elements) and a growth rate of 1.2.To solve the continuity and momentum equations, appropriate boundary conditions were given at all boundaries. Because of the various time steps and computer time constraints, the CFD simulation took ten hours to complete. The received sun energy by the basin, as well as the water and glass temperatures, were assumed to be based on the solar calculator under fair weather conditions for 1 h. On the glass, the bottom, and the collecting surface, constant temperature boundary conditions were imposed.
	Rehbar et al.[9]	The flow is considered to be axisymmetric, two-dimensional, stable, and laminar. The author uses constant temperatures for the water and cover, as well as an adiabatic boundary condition for the trough wall, in this investigation. Furthermore, the fluid is modelled as humid air, an incompressible ideal gas with constant physical parameters and negligible viscous dissipation.
	Shakaib[10]	Six faces of the geometry are considered impenetrable walls. The temperatures on the bottom and top surfaces are regulated at 300 and 350 K, respectively, and the rest is adiabatic. The unit's vertical heights are 0.61 m and 0.25 m, resulting in a 20° top surface slope. The unit's width and depth are both 1 m. The geometry is constructed and meshed in Gambit. Mesh contains fine cells. It has about 100,000 cells which yield mesh-independent results. The mesh is exported in ANSYS FLUENT 12, which solves continuity, energy equations and momentum. The fluid is ideal air, with temperature-dependent thermal conductivity and viscosity. The simulation's Rayleigh number was 1000. The convective terms are discretized using a second-order upwind approach, and the pressure and velocity fields are coupled using the SIMPLE algorithm.
ANSYS CFX	Singh and Mittal[11]	The ANSYS CAD module is used to construct a geometric modal of a solar still, which is then imported into the ANSYS meshing module to generate the meshing.Solar still mesh files are then imported into the CFX Pre. The use of pre-physics and boundary conditions on the domain is explained in CFX in order to solve the continuity and momentum equation. The (VOF) framework creates a two-phase domain for liquid water and a mixture of water vapour and air. The evaporation process is simulated as laminar at quasi-steady, with thermal energy heat transfer and buoyancy effects taken into account. Both the vapour and liquid phases have a distinct interface, therefore both are continuous. Two resistance models are used to transmit heat: a zero resistance model for the gas phase and a heat transfer coefficient for the water phase. The bottom temperature was expected to be the same as the water bath temperature.
Comsol Multiphysics	Maalem et al. [17]	Non-adiabatic walls have temperatures of 294.2 (K) for aluminium, 322.6 (K) for water and 308.6 (K) for glass. These temperatures are used as non-adiabatic wall boundary conditions.

nation process and heat transfer coefficient. The results show that the maximum average temperature was obtained in May, i.e., 44 °C while the maximum production of resultant distilled water was 1000 ml per day. The average global solar radiation was found to be 5.8 kWh/m^2 on the month of experimentation. For evaluating the thermal behaviour of the solar still, the author employed the Kumar and Tiwari model. The incident radiation increases when the cover and water internal temperatures rise. The glass temperature affects heat transmission coefficients (internal).

The quantity of distillate water is proportional to the above variables, implying that more distillate is produced. Different distillation equipment requires a lot of heat transmission. The CFX simulation depicts the behaviour of the distillation machine in extracting water as well as the temperature of the glass and water. The differences between the CFX simulation and the experimental technique were investigated. The difference in distiller volume was 8.2 %, whereas the difference in water temperature was 5.8 %. The equipment's efficiency was within the acceptable range. It was found that the behavior of the solar distiller can be obtained be affected by the weather conditions. Radiation is a critical factor for high efficiency. However, ambient temperature and wind speed can also affect the system yield. The results obtained in CFD were different compared to that of the experimental results. It provides an opportunity to improve the equipment's efficiency by adding heat, enhancing construction materials, or adding additional components to perform forced convection. At this moment, the production of distilled water can be increased in the water-scarce regions. The water quality is being tested to analyze the drinkability of distilled water. The water quality standard meets the NOM-127-SSA1-1994.

6. Conclusion

This paper reviewed the research on various design software and programming languages in still solar applications for the design parameters. Essential areas of their application are towards the optimization of design parameters. Application of software not only reduces the consumption time for experimental work but also saves the cost. CFD ANSYS/FLUENT/CFX for visualization of flow parameters like the circulation of air-vapor mixture, temperature contours inside the solar still, condensing and evaporating zones, velocity distribution, shear stress flow analysis and flow patterns near-vertical walls inside the solar still system. MATLAB can develop a simulation for equivalent mathematical models of solar still systems for different parameter tools. Comsol Multiphysics programming, TRNSYS, FORTRAN can be used for the energy balance equation and solution of the partial differential. These programming languages aid in predicting heat transfer coefficients, wall temperatures, and water temperatures within solar stills. These coefficients determine the convective, evaporative, and radiative heat transfer rates of a solar still. As a result, scientists' simulation models in the field of solar still distillation systems will be useful in future improvements in this field.

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] P. Kalita, A. Dewan, S. Borah, A review on recent developments in solar distillation units, Sadhana. 41 (2) (2016) 203–223.
[2] A. Ahmad, O.m. Prakash, A. Kumar, Drying kinetics and economic analysis of bitter gourd flakes drying inside hybrid greenhouse dryer, Environ. Sci. Pollut. Res. (2021) 1–15.
[3] A. Ahmad, O.m. Prakash, Thermal analysis of north wall insulated greenhouse dryer at different bed conditions operating under natural convection mode, Environ. Prog. Sustainable Energy 38 (6) (2019) 1–12.

[4] R. Tripathi, G.N. Tiwari, Thermal modeling of passive and active solar stills for different depths of water by using the concept of solar fraction, Sol. Energy 80 (8) (2006) 956–967.
[5] O.m. Prakash, B. Bhushan, A. Kumar, A. Ahmed, Thermal analysis of domestic type single Slope–Basin solar still under two different water depths, Mater. Today:. Proc. 46 (2021) 5482–5489.
[6] A. Ahmad, O.m. Prakash, Performance Evaluation of a Solar Greenhouse Dryer at Different Bed Conditions Under Passive Mode, J. Sol. Energy Eng. 142 (1) (2020) 1–10.
[7] S. Vaithilingam, G.S. Esakkimuthu, Energy and exergy analysis of single slope passive solar still: an experimental investigation, Desalin. Water Treat. 55 (6) (2015) 1433–1444.
[8] V.R. Khare, A.P. Singh, H. Kumar, R. Khatri, Modelling and performance enhancement of single slope solar still using CFD, Energy Procedia 109 (2017) 447–455.
[9] N. Rahbar, J.A. Esfahani, E. Fotouhi-Bafghi, Estimation of convective heat transfer coefficient and water-productivity in a tubular solar still–CFD simulation and theoretical analysis, Sol. Energy 113 (2015) 313–323.
[10] M. Shakaib, M.A. Khan, Modeling of Fluid Flow and Temperature Profiles in Solar Stills Using CFD, In Proceedings of international conference on chemical, metallurgy and environmental engineering, 2015.
[11] A. Singh, M.K. Mittal, Simulation of single slope solar still at different inclinations using CFD, In Int. Conf. Adv. Res. Innov. (2014) 512–519.
[12] T.A. Kumar, A. Thakur, P.S. Kumar, Single basin solar still with varying depth of water: optimization by computational method. Iranian Journal of, Energy Environ. (2017) 216–223.
[13] Boubekri, Mahmoud, Abla Chaker, and Souad Saadi. "Modeling and simulation of a vertical multi-effect diffusion solar still coupled with a tracking solar system and a PV/T-SWH system." J. Therm. Sci. Eng. Appl. 13, no. **5**: 051003 (2021).
[14] P. Rahdan, A. Kasaeian, W.-M. Yan, Simulation and geometric optimization of a hybrid system of solar chimney and water desalination, Energy Convers. Manage. 243 (2021) 114291, https://doi.org/10.1016/j.enconman.2021.114291.
[15] S. Shoeibi, H. Kargarsharifabad, N. Rahbar, G. Ahmadi, M.R. Safaei, Performance evaluation of a solar still using hybrid nanofluid glass cooling-CFD simulation and environmental analysis, Sustainable Energy Technol. Assess. 49 (2022) 101728, https://doi.org/10.1016/j.seta.2021.101728.
[16] N. Rahbar, J.A. Esfahani, Productivity estimation of a single-slope solar still: Theoretical and numerical analysis, Energy 49 (2013) 289–297.
[17] M.S. Maalem, A. Benzaoui, A. Bouhenna, Modeling of simultaneous transfers of heat and mass in a trapezoidal solar distiller, Desalination 344 (2014) 371–382.
[18] S. Rashidi, S. Akar, M. Bovand, R. Ellahi, Volume of fluid model to simulate the nanofluid flow and entropy generation in a single slope solar still, Renewable Energy 115 (2018) 400–410.
[19] A.K. Tiwari, G.N. Tiwari, Effect of water depths on heat and mass transfer in a passive solar still: in summer climatic condition, Desalination 195 (1-3) (2006) 78–94.
[20] M.I. Ali, R. Senthilkumar, R. Mahendren, Modelling of solar still using granular activated carbon in matlab. Bonfring International Journal of Power Systems and Integrated Circuits. 1(Inaugural Special Issue):05-10 (2011).
[21] M. Edalatpour, A. Kianifar, S. Ghiami, Effect of blade installation on heat transfer and fluid flow within a single slope solar still, Int. Commun. Heat Mass Transf. 66 (2015) 63–70.
[22] Y.A.F. El-Samadony, W.M. El-Maghlany, A.E. Kabeel, Influence of glass cover inclination angle on radiation heat transfer rate within stepped solar still, Desalination 384 (2016) 68–77.
[23] A. Ghoneyem, A. Ileri, Software to analyse solar stills and an experimental study on the effects of the cover, Desalination 114 (1997) 37–44.
[24] R.S. Adhikari, A. Kumar, G.D. Sootha, Simulation studies on a multi-stage stacked tray solar still, Sol. Energy 54 (5) (1995) 317–325.
[25] M.T. Chaibi, Analysis by simulation of a solar still integrated in a greenhouse roof, Desalination 128 (2) (2000) 123–138.
[26] A.G.M. Ibrahim, S.E. Elshamarka, Performance study of a modified basin type solar still, Sol. Energy 118 (2015) 397–409.
[27] H. Ben Halima, N. Frikha, R. Ben Slama, Numerical investigation of a simple solar still coupled to a compression heat pump, Desalination 337 (2014) 60–66.
[28] D. Kumar, P. Kumar, Mathematical modelling of conventional solar still coupled with solar air heater, Int. J. Innovat. Sci., Eng. Technol. 1 (9) (2014) 379–383.
[29] F.M. Abed, M.S. Kassim, M.R. Rahi, Performance improvement of a passive solar still in a water desalination, Int. J. Environ. Sci. Technol. 14 (6) (2017) 1277–1284.
[30] D.B. Singh, A.K. Singh, V.K. Kumar Navneet, J.K.Yadav Dwivedi, Singh Gajendra, Performance analysis of special design single basin passive solar distillation systems: a comprehensive review, in: Advances in Engineering Design, Springer, Singapore, 2019, pp. 301–310.
[31] M.Y. Wong, B. Traipattanakul, C.Y. Tso, C.Y.H. Chao, H. Qiu, Experimental and theoretical study of a water-vapor chamber thermal diode, Int. J. Heat Mass Transf. 138 (2019) 173–183.
[32] A.H. Elsheikh, S.W. Sharshir, M. Abd Elaziz, A.E. Kabeel, W. Guilan, Z. Haiou, Modeling of solar energy systems using artificial neural network: A comprehensive review, Sol. Energy 180 (2019) 622–639.
[33] R.J. García-Chávez, A.U. Chávez-Ramirez, H.I. Villafán-Vidales, J.B. Velázquez-Fernández, I.P. Hernández, Rosales., Thermal study of a solar distiller using computational fluid dynamics (CFD), Revista Mexicana De Ingeniería Química 19 (2) (2020) 677–689.

View publication stats

Proceedings of the Fourth International Conference on Smart Systems and Inventive Technology (ICSSIT-2022)
IEEE Xplore Part Number: CFP22P17-ART; ISBN: 978-1-6654-0118-0

Crime Based Evaluation of GPS Network Using Machine Learning Techniques

Rahul
Department Of Software Engineering
Delhi Technological University
Delhi, India
rahul@dtu.ac.in

Monika
Department of Computer Science
Shaheed Rajguru College of
Applied Sciences for Women,
University of Delhi,
Delhi-110096
Monika.siwaliya@gmail.com

Vartika Hari Durgapal
Department Of Software Engineering
Delhi Technological University
Delhi, India
vartikaharidurgapal_2k19se141@dtu.ac.in

2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT) | 978-1-6654-0118-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICSSIT53264.2022.9716447

Abstract— **With the increased usage of data transmission, data leakage and privacy protection are becoming increasingly critical. Data comes in a variety of forms, and the amount of protection required for each one differs. With the rapid development and demand of Global Positioning System (GPS) networks in every of big data technology, the vulnerability of the transportation system to unintentional and intentional GPS disruption is also increasing. The GPS signal is degraded and lost as a result of opposing interests' attacks. Potential assaults vary from GPS signal jamming and spoofing to GPS ground stations and satellite disruption. Security issues with respect to data increases during data transmission. Data travelling across an organization's network is frequently sensitive and internet being the storehouse of every kind of data with methods for accessing the precise information serves as an auspicious medium for the hackers [19]. Network security solutions must deal with complicated data in this setting in order to detect and minimize any threats in real time. There are several techniques to the problem of data security such as cryptographic algorithms, classification-based models, and block-chain methods [1]-[6]. Through this research, crime rates are evaluated in a GPS network data with different other parameters using machine learning (ML) algorithms. Moreover, the inferred features effecting the crimes during data transmission are extracted. Some of the techniques used includes techniques such as 'Decision Tree (DT)', 'K Nearest Neighbors (KNN)', 'Random Forest (RF)', 'Multivariable Regression (MR)' and 'Naïve Bayes (NB)'. It was seen that the best accuracy was given by RF - of approximately 84.25%.**

Keywords— ***Data Security, Crime Evaluation, Machine Learning, GPS Network***

I. INTRODUCTION

With the rapid development of social information and network, the data on the network appeared explosive growth making secure and safe data transmission with data tracking facilities at every location as increasing demand of every organization. GPS is one such utility that has gained remarkable significance due to its numerous applications in navigation, tracking, mapping, and other areas. GPS, formerly known as Navstar GPS, is a US government-owned and managed satellite orbiting the earth. In a day, each satellite completes two full rotations around the earth at a distance of about 12,000 miles (19,300 kilometers). Meanwhile, researchers have discovered that GPS is susceptible to spoofing and jamming. In order to benefit from the advantages of GPS Networks and to avoid uncertain security risks, users and service providers of GPS Networks always want to evaluate the security status of the current network, and a Security Measurement Method for GPS Networks is required.

As mentioned in [14], the GPS signal is susceptible to degradation and loss as a result of hostile interference. Potential assaults vary from GPS signal jamming and spoofing to GPS ground stations and satellite disruption. Natural occurrences, unintentional (for example, some TV broadcast frequencies, mobile phone emissions), and purposeful (for example, spoofing and jamming) interference, as well as vulnerabilities to signal deterioration, have all raised the risks of using GPS networks. The common use of these vulnerabilities is made by attackers to manipulate information in the signals and time delays. Such spoofing techniques are known as GPS meaconing, emphasizing the necessity of scrutinizing aspects that effect data delivery.

Currently, many data security frame works to enhance data security such as encryption methods and cryptographic algorithms have been proposed [1]-[6] and techniques for achieving the task of classification of GPS networks. [7]-[9] includes classification of GPS network through neural networks. In this method the impact of roadway features is analyzed using convolutional neural network (CNN) algorithm after labeling large-scale GPS data through strategic algorithms. Some more common machine learning algorithms such as 'Random Forest' to classify the GPS data [13] and gradient boosting decision tree-based GPS signal reception classification algorithm [11] are also mentioned in various research**.** However, some drawbacks of using neural networks (NN) are slow model learning, high compute power, and environmental sensitivity, which is a major issue in classification-based algorithms.

However, still evaluation of crimes rate in GPS networking dataset on the basis of various parameters based on location is yet to be studied. A deep study of crime analysis identifying patterns, disorders and trends using big data is needed [20]. A machine learning-based approach is used to analyze crime rates for a GPS network in this study. Several ML algorithms were trained using features that came out to be the most significant from the data using their significance values. The machine learning techniques used for classification are Decision Tree, K Nearest Neighbors, Random Forest, and Naive Bayes to classify the data. This technique has been used in many other network classifications as well such as 4G networks on the basis of internal protocol (IP) traffic [21]. For this research, the information regarding the dataset used and methods used to analyze and evaluate it is in [18]. The dataset used for training and testing purposes is a GPS trajectory dataset containing information of different routes. Each route

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 05,2022 at 05:45:53 UTC from IEEE Xplore. Restrictions apply.

includes information on the duration of the trip, the length of the route, traffic, criminality, accidents, nature, and adjacent tourist sites. After building different models and observing results RF algorithm gave the best accuracy of 84.255%.

II. RELATED WORKS

Several studies have been conducted for the classification of GPS in various field such as satellites, vehicles. Some of these research projects are described in the subsections below:

Detection of spoofing in GPS networks was done in [10]. An artificial NN-based approach for identifying GPS spoofing signals is proposed in this research. The data used for this purpose is of unmanned ariel systems such as satellite systems, broadcast receivers, etc. Five features which are used for this evaluation purposes are five features: 'satellite number', 'carrier phase', 'pseudo range', 'doppler shift', and 'signal-to-noise ratio (SNR)'. The efficiency of different models with varying hidden layers are compared and the efficiency was evaluated in the basis of detection, missed detection and false detections and the accuracy. To collect and analyse spoofing messages unsupervised learning is used.

GPS geometry dilution of precision (GDOP) categorization using a NN has been discussed in [7]. For this purpose, two types of neural networks are used and compared with the traditional ones. These are termed as back propagation neural network (BPNN) and generalizes regression neural network (GRNN). A detailed study of the results given by the two NN approximators varying parameters such as no of hidden layers, no of iterations, training time, etc. is done in this paper. It is found out that these algorithms perform much better that the commonly used networks in terms of accuracy and the training time taken. The best results are given by BPNN. However, this technique has its drawbacks, such as slow learning also mentioned in section 1. The accuracy and good performance involve major tradeoff between user's requirements.

A GPS signal reception categorization system based on gradient boosting decision tree (GBDT) is mentioned in [11]. The GBDT categorization technique is used to find the most optimal algorithm iteration number to ensure highest classification accuracy without overfitting the GBDT algorithm. Its purpose is to reduce the GPS network deviation from its transmission path due to reflection or blockage by buildings or any other obstacles in urban environments. It is also used to improve GPS positioning accuracy by taking various factors such as satellite elevation angle, pseudo range residuals, etc. known as carrier to noise ratio. But these factors do have their own cons such as consumption of more time for training and testing due to its high computation power. For the detection of the location, environmental sensitivity of the area is the major issue as the accuracy of algorithm varies from place to place.

Other related techniques include a graph neural network approach mentioned in [12], employs a graph convolutional neural network (GCNN) model. Since it deals with taxi GPS trajectory data various geo-semantic embedding features are examined using deep learning techniques. GCNN model is created using three level NN architecture and parameters such as cross-entropy loss and accuracy are evaluated. Model with most optimal number of iterations along with some other ML techniques such as linear regression, support vector machines, KNN, RF. However, while in such studies no of geo-semantic features vary from location to location and different components play different significance role in different locations. But due to the limitation of this study the categorization is done in only three main parts (commercial, public, and traffic). Another such method is done using RF of GPS sensors for speed bump detection in [13] and some others [8]-[9]. However, not much research has been done to analyze the contributors of crime rates in data transmission. Thus, this analysis might prove to be helpful in further utilizing this in future models to create a safer medium for transferring data.

III. METHODOLOGY

Following steps were followed in the given manner in this research:

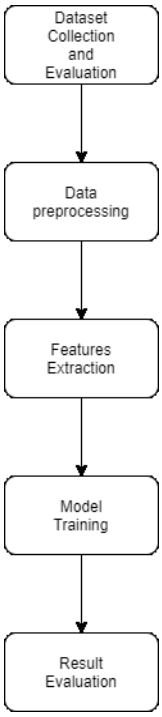


Fig 1. Steps of implementation

The steps were initiated with the proper dataset collection to evaluate the GPS networks properties and then evaluated by preprocessing and analyzing the features and their respective roles in crime rates. The features extraction was a major step to get the idea of overpowering attributes in model's performance. For creating the model, the common machine learning algorithms such as DT, KNN, RF, NB were used and results of each algorithm was evaluated separately and at the end was a comparative study of all the algorithms with their respective results.

A. Dataset

Dataset plays a major role in training machine learning models and improving their accuracy. For evaluating the crime rates, the dataset used is a GPS network dataset

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 05,2022 at 05:45:53 UTC from IEEE Xplore. Restrictions apply.

containing information about 600 sets of routes from different origin/destination pairs from London. The route information includes trip duration, route length, traffic, crime, accidents, nature, and adjacent tourist sites. The data of duration estimations, criminality, and accidents, were obtained from the UK Government open data website for each route [22]. More information about the attributes of dataset is in table 1:

TABLE I DATASET FEATURES DESCRIPTION

S.no.	Attributes	Format	Description
1.	Crime	Numeric	Represents severity of the crime s from National Statistics Crime Severity Score
2.	Accidents	Numeric	Represents weighted accident severity
3.	Tourist Attraction	Numeric	Represents no of tourists in a particular area
4.	Traffic	Numeric	Ratio of no-traffic route duration and the rou te duration with traffic
5.	Duration	Numeric	Represents route duration
6.	Length	Numeric	Length of the route
7.	Nature	Numeric	Represents intersecting the natural area polygons

These routes have length and duration already. A visible area is estimated taking into account the route segments and a margin of 100 m for the left and right side for some of the feature's calculations. The nature feature was calculated intersecting the natural area polygons given by OpenStreetMap and the visible area. The attractions feature was calculated as the number of tourist attractions given by OpenStreetMap inside the route's visible area. The criminality was calculated as the sum of the severity of the crimes, provided by the UK police crime dataset, found inside the visible area. The severity of the crimes was weighted by the UK Office for National Statistics Crime Severity Score. The accidents feature was calculated as the sum of accidents severity and accidents casualties that happened in the route. The data was provided by the UK Government Open Data Website. The traffic feature was calculated as the ratio between the no-traffic route duration and the route duration considering traffic.

B. Features Selection

After the collection of dataset, various methods of feature selection were used such as 'Select-K-Best' and 'Extra Tree Classifier'. "Extra Tree Classifier" works similar to random forest classifier by using normalized total reduction method to split features to create the tree and give Gini index or importance to every attribute. Using "Select-K-Best" top most attributes with best Gini scores were selected for further usage in model creation. The top 6 features for the dataset are:

TABLE II FEATURES IMPORTANCE SCORES

Features	Feature Importance
Traffic	0.0782
Accidents	0.0767
ID	0.0767
Length	0.0765
Duration	0.0759
Nature	0.0751

C. Data Preprocessing

After the extraction of top features of a dataset, the dataset was evaluated removing any null or missing values. It was then scaled using feature scaling methods to normalize the data such that no feature would dominate over others. A train-test ratio of 7:3 was used to split for training and testing the models using model_selection train_test_split function of sklearn.

D. Model Training

After preprocessing the data and understanding it, various models were built for different algorithms [15]-[17] and trained using the training dataset. The DT, KNN, RF, NB algorithms were employed in this study. A short description of algorithms used are:

- Decision Tree is a supervised learning method having structure like a tree uses each node as features and branches as decision rules and each leaf node as the outcome.
- Random forest is used for classification and regression contains individual decision trees which gives a class prediction and the class with maximum prediction is the output.
- The K nearest neighbor uses feature similarity to predict the values. It uses data points in training set to get weighted average of the distance of K nearest neighbors and uses it to evaluate test data.
- Naïve Bayes is a probabilistic model which uses bayes theorem as its fundamental. The features are considered independent while using bayes theorem.

The trained models were then sorely tested on the testing dataset to determine the quantities such as 'R2 scores', 'mean absolute errors', 'mean squared errors', 'accuracy'. Also, their training time is compared.

E. Performance Measures

The measures which have been used to evaluate the performance of different models are accuracy, model training time, mean squared error, r2 scores. The following is a brief breakdown of various performance measures [15], [16], and [17]:

1) Accuracy: Accuracy is one of the parameter for evaluating models as the percentage of correct predictions this model made. Formally, accuracy has the following definition:

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

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 05,2022 at 05:45:53 UTC from IEEE Xplore. Restrictions apply.

TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

2) R^2 - Score: R-squared (R2) is a relative statistical measure that represents the variablity explained by the regression model divided by the total variability of the dataset. In other words, R2 measures the goodness of fit of the model. The value of R-squared ranges from 0 to 1.

$$R^2 = 1 - \frac{\text{Unexplained Variation}}{\text{Total Variation}} \quad (2)$$

3) Mean-Squared-Error: The Mean Squared Error (MSE) or Mean Squared Deviation (MSD) of a model measures the mean of squares of the diffrence between actual and predicted values. Lower the error better the model is. It provides a relative measure of how close the model predicts the values to the real values.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \quad (3)$$

N = No. of data points
Y_i = Observed values
$\hat{Y}_i$ = Predicted values

4) Mean absolute error(MAE): It is the mean of absolute values of the diffrence between predicted and actual values of a parameter. It provides a measure of how accurate the model is in detecting values as compared to the actual values.

$$\text{MAE} = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \quad (4)$$

y_i = Prediction
x_i = True value
n = Total no. of data points

5) Training Time: Time required to fit a model with training dataset is the training time.

IV. RESULTS

After creating the models and testing it with the data following results were evaluated.

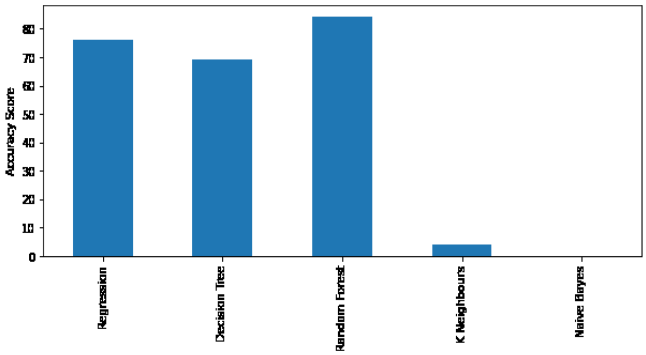


Fig 2. Accuracy Scores of the used models

The accuracy scores of different machines learning techniques is shown in Fig 2. From the figure, Random Forest regression model gave the best accuracy and naïve bayes model accuracy is negligible as compared to other given. Multivariable regression proved to be second most effective

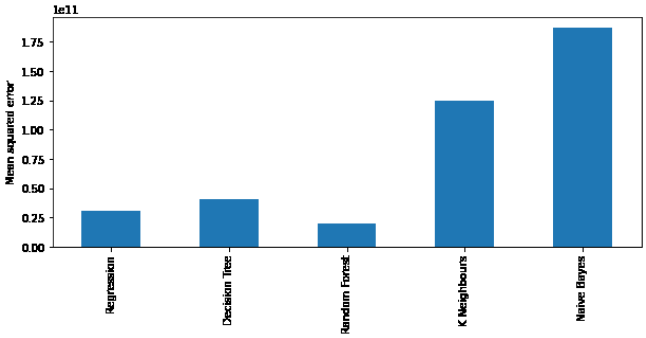


Fig 3. Mean Squared Error of all the models

Mean squared error defines the difference of estimated and actual values, so lower the error better is the model. From the fig 3. random forest algorithm has the lowest value which further makes it the better among the others. Naïve bayes again have the highest error thus having the most deviation of results of model from actual ones.

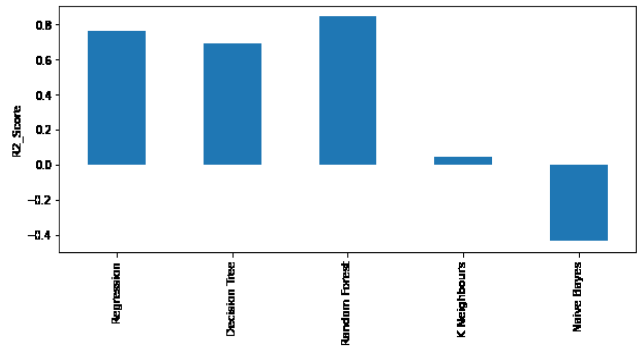


Fig 4. R2_Score of all the models

R2 score is the estimation of how model fits the data. Thus, a higher r-squared indicates a better fit for the model. From fig 4. random forest has the highest r2 score with regression and decision tree next in the line.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 05,2022 at 05:45:53 UTC from IEEE Xplore. Restrictions apply.

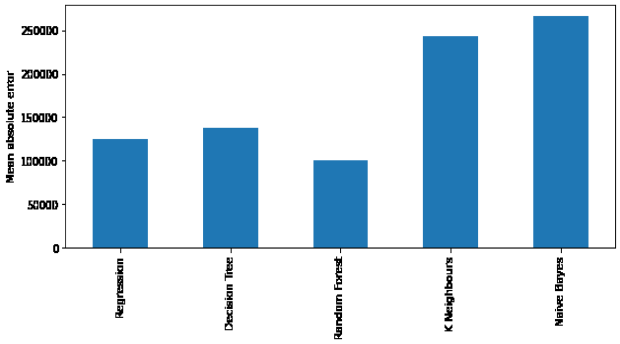


Fig 5. Mean Absolute Error of all the models

For further evaluation of the model's performance mean absolute error which is again a method of finding deviation between actual and model results is used and results are shown in fig 5. with random forest giving the lowest error among all.

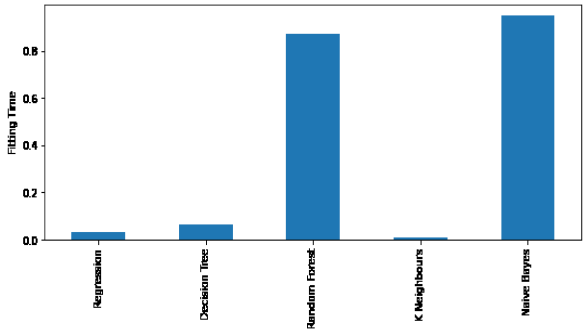


Fig 6. Training Time of all the models

Training time taken by each algorithm is shown in fig 6. with naïve bayes taking the highest time to train followed by the random forest algorithm.

TABLE III ACCURACY AND TRAINING TIME OF ALL ALGORITHMS

S.no.	Models/Algorithms	Accuracy (%)	Training time (s)
1.	Regression	76.26	0.032
2.	Decision Tree	68.72	0.0625
3.	Random Forest	84.75	0.873
4.	K Neighbors	4.16	0.0089
5.	Naïve Bayes	0.105	0.95

From the table, it is clear that Random Forest is highest with an accuracy of 84.25%, but the training time of k-neighbor is lowest and Naive Bayes as highest with random forest taking the second most fitting time. Further Fig 2, 3, 4, 5, 6 show the accuracy scores, mean squared errors, R2 scores, mean absolute errors, training time respectively.

After analyzing the scores and other factors, random forest algorithm seems to be the most suitable algorithm among others. Although the mean squared errors and mean absolute errors of naïve bayes, and k neighbors are the highest but their corresponding R2 scores and accuracy are significantly lower than the others. The R2 scores and accuracy of random forest are the highest with other scores as one of the highest. The features that were given the most importance is given in the figure with the importance scores:

TABLE IV TOP 6 FEATURES OF DATASET ATTRIBUTES

S.no.	Features
1.	Traffic
2.	Accidents
3.	ID
4.	Length
5.	Duration
6.	Nature

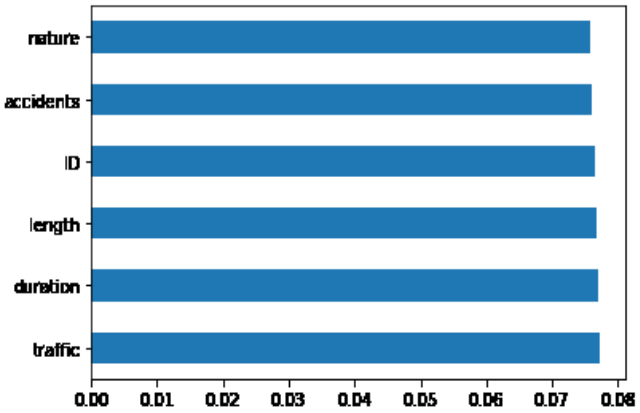


Fig 7. Top Features of dataset

V. CONCLUSION AND FUTURE WORKS

In this research, a GPS network dataset was evaluated to find the top features responsible for crime rates in network transmission. Features with most significance such as traffic, duration, length, etc. were calculated and filtered to evaluate the crimes and the data was preprocessed using various methods such as standardization, etc. The dataset was then used to train 5 different ML techniques. These techniques were used to create different models and those models were then fitted using training data and evaluated using testing data. The results were finally compared on the basis of highest and lowest values of parameters such as MAE, MSE, R2 scores and training time.

By comparing all the methods, Random Forest had the most accuracy of 84.25%. Also, the most relevant features of the dataset like nature, accidents, traffic, duration, length and ID. These characteristics could be useful in future research projects in areas like network administration and security.

Although many other techniques such as neural networks have been used to evaluate GPS networks mentioned in previous sections but crime evaluation can also be done which might further provide more refined

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 05,2022 at 05:45:53 UTC from IEEE Xplore. Restrictions apply.

results and features extraction for this study This study can be further extended with the use of location coordinates in a particular network route with crimes and then using CNNs to classify the features.

REFERENCES

[1] Thabit, Fursan, Sharaf Alhomdy, Abdulrazzaq HA Al-Ahdal, and Sudhir Jagtap. "A new lightweight cryptographic algorithm for enhancing data security in cloud computing." *Global Transitions Proceedings* 2, no. 1, pp.91-99., 2021, DOI: 10.1016/j.gltp.2

[2] Jeon, Jin Hyeong, Ki-Hyung Kim, and Jai-Hoon Kim. "Block chain based data security enhanced IoT server platform." In *2018 International Conference on Information Networking (ICOIN)*, pp. 941-944., 2018, DOI: 10.1109/ICOIN.2018.8343262

[3] Kaur, Preet Chandan, Tushar Ghorpade, and Vanita Mane. "Analysis of data security by using anonymization techniques." In *2016 6th International Conference-Cloud System and Big Data Engineering (Confluence)*, pp. 287-293, 2016, DOI: 10.1109/CONFLUENCE.2016.7508130

[4] El Alaoui, Imane, and Youssef Gahi. "Network Security Strategies in Big Data Context." *Procedia Computer Science*, *175*, pp.730-736, 2020, DOI: 10.1016/j.procs.2020.07.108

[5] Shaikh, Rizwana, and M. Sasikumar. "Data Classification for achieving Security in cloud computing." *Procedia computer science*, *45*, pp.493-498., 2015, DOI: 10.1016/j.procs.2015.03.087

[6] Fu, Yulong, Hanlu Chen, Qinghua Zheng, Zheng Yan, Raimo Kantola, Xuyang Jing, Jin Cao, and Hui Li. "An Adaptive Security Data Collection and Composition Recognition method for security measurement over LTE/LTE-A networks." *Journal of Network and Computer Applications*, *155*, p.102549., 2020, DOI: 10.1016/j.jnca.2020.102549

[7] Azami, Hamed, and Saeid Sanei. "GPS GDOP classification via improved neural network trainings and principal component analysis." *International Journal of Electronics*, *101*(9), pp.1300-1313., 2014, DOI: 10.1080/00207217.2013.832390

[8] Dabiri, Sina, Chang-Tien Lu, Kevin Heaslip, and Chandan K. Reddy. "Semi-supervised deep learning approach for transportation mode identification using GPS trajectory data." *IEEE Transactions on Knowledge and Data Engineering*, *32*(5), pp.1010-1023., 2019, DOI: 10.1109/TKDE.2019.2896985

[9] Dabiri, Sina, Nikola Marković, Kevin Heaslip, and Chandan K. Reddy. "A deep convolutional neural network based approach for vehicle classification using large-scale GPS trajectory data." *Transportation Research Part C: Emerging Technologies*, *116*, p.102644., 2020, DOI: 10.1016/j.trc.2020.102644

[10] Manesh, Mohsen Riahi, Jonathan Kenney, Wen Chen Hu, Vijaya Kumar Devabhaktuni, and Naima Kaabouch. "Detection of GPS spoofing attacks on unmanned aerial systems." In *2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pp. 1-6., 2019, .

[11] Sun, Rui, Guanyu Wang, Wenyu Zhang, Li-Ta Hsu, and Washington Y. Ochieng. "A gradient boosting decision tree based GPS signal reception classification algorithm." *Applied Soft Computing, 86*, p.105942., 2020, DOI: 10.1016/j.asoc.2019.105942

[12] Hu, Sheng, Song Gao, Liang Wu, Yongyang Xu, Ziwei Zhang, Haifu Cui, and Xi Gong. "Urban function classification at road segment level using taxi trajectory data: A graph convolutional neural network approach." *Computers, Environment and Urban Systems, 87*, p.101619., 2021, DOI: 10.1016/j.compenvurbsys.2021.101619

[13] Joon, Jo Hae, Dionysius Aldion Renata, Su Hyun Yoon, Choi Youngjune, and Yalew Zelalem Jembre. "Data Mining for Speed Bump Detection from Car Wheels and GPS Sensors using Random Forest." In *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 740-743., 2019., DOI: 10.1109/ICTC46691.2019.8939736

[14] Carroll, James V. "Vulnerability assessment of the US transportation infrastructure that relies on the global positioning system." *The Journal of Navigation* 56, no. 2, pp.185-193., 2003

[15] "scikit-learn: machine learning in Python — scikit-learn 0.24.2 documentation." https://scikit-learn.org/stable/

[16] "scikit-learn: metrics and scoring: quantifying the quality of predictions — scikit-learn 0.24.2 documentation" https://scikit-learn.org/stable/modules/model_evaluation.html

[17] "scikit-learn: regression techniques" https://scikit-learn.org/stable/modules/tree.html#regression

[18] "GitHub - Learning-to-code12/GPS-Trajectory-Evaluation" https://github.com/Learning-to-code12/GPS-Trajectory-Evaluation

[19] Adhikari, Surabhi. "Nlp based machine learning approaches for text summarization." In *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 535-538. IEEE, 2020, DOI: 10.1109/ICCMC48092.2020.ICCMC-00099

[20] Bhat, Aruna. "An analysis of Crime data under Apache Pig on Big Data." In *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, pp. 330-335. IEEE, 2019, DOI: 10.1109/I-SMAC47947.2019.9032565

[21] Gupta, Amit, Anupam Raj, and Mayank Arora. "IP Traffic Classification of 4G Network using Machine Learning Techniques." In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 127-132. IEEE, 2021, DOI: 10.1109/iccmc51019.2021.9418397

[22] "Crawded dataset: london trajectories dataset" https://crawdad.org/ufrj/london-trajectories/2021-03-22/index.html

[23] Bestak, Robert. "INTELLIGENT TRAFFIC CONTROL DEVICE MODEL USING AD HOC NETWORK." Journal of Information Technology 1, no. 02 (2019): 68-76

[24] Chen, Joy Iong Zong. "Optimal Multipath Conveyance with Improved Survivability for WSN's In Challenging Location." Journal of ISMAC 2, no. 02 (2020): 73-82.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 05,2022 at 05:45:53 UTC from IEEE Xplore. Restrictions apply.

Decoding the Brain Waves using EEG signals for classifying Body Gestures by applying suitable ML & DL Techniques

Dr Ruchika Malhotra*
Delhi Technological University, Delhi
ruchikamalhotra@dtu.ac.in

Dhananjay Deswal
Delhi Technological University, Delhi
dhananjaydeswal_2k18se054@dtu.ac.in

Jai Chaudhry
Delhi Technological University, Delhi
jaichaudhry_2k18se069@dtu.ac.in

ABSTRACT

This work aims to use EEG brain signals captured from the scalp of subjects at 500 Hz for classification of grasping tasks which is useful for BCI (Brain-Computer Interface) and comparative analysis of appropriate Machine Learning and Deep Learning Algorithms for the same. The dataset consists of 32 features and 6 output classes. Multinomial Logistic Regression, LSTM, CNN, RNN and ANN Algorithms were used for this purpose.

CCS CONCEPTS

• **Computing methodologies** → Machine learning; Machine learning approaches; Neural networks; Machine learning; Learning paradigms; Supervised learning; Supervised learning by classification.

KEYWORDS

EEG, ANN, LSTM, CNN, RNN, LR, Comparative Analysis

ACM Reference Format:
Dr Ruchika Malhotra, Dhananjay Deswal, and Jai Chaudhry. 2022. Decoding the Brain Waves using EEG signals for classifying Body Gestures by applying suitable ML & DL Techniques. In *2022 4th Asia Pacific Information Technology Conference (APIT 2022), January 14–16, 2022, Virtual Event, Thailand*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3512353.3512370

1 INTRODUCTION

The ability to detect mental states and the thinking process of the brain is useful for implementation of BCI (Brain-Computer Interface) software. This can be achieved using EEG signals from the brain. The electroencephalogram (EEG) is a recording of electrical activity of the brain from the scalp which serves as a non-intrusive approach for measurement of brain activity.

In other words, it's a sensor to measure human reaction to various sensory inputs to the human body. The main frequencies of human EEG waves are Delta (below 3Hz), Theta (3.5 – 7.5Hz), Alpha (7.5 – 13Hz) and Beta (above 14Hz).

One of the main issues of using EEG signals is the amount of data needed to properly predict and classify signals since the signals are random, complex and non-linear in nature. Therefore, training was done with short windows of the data to enable the ML and DL algorithms to learn properly on the features.

*HOD, Software Engineering Department, Delhi Technological University

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.
APIT 2022, January 14–16, 2022, Virtual Event, Thailand
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9557-1/22/01…$15.00
https://doi.org/10.1145/3512353.3512370

2 EXPERIMENT DESIGN

2.1 Dataset

The dataset has been taken from https://www.kaggle.com/c/grasp-and-lift-eeg-detection/data. The dataset includes 192 files (8 input and 8 output files for each of 12 subjects). Each input file has 32 features, corresponding to the 32 electrodes which are connected to the scalp of a subject for recording the EEG signals at 500 Hz frequency.

Each output file consists of 6 columns in the form of one hot encoded vector. The 6 columns correspond to 6 body gestures, namely, HandStart (starting the hand movement), FirstDigitTouch (touching a can with the hand), BothStartLoadPhase (grasping the can and starting to lift it), LiftOff (lifting the can), Replace (placing the can back), BothReleased (releasing the can).

For training we used a total 24 x 2 input and output files (4 x 2 input and output files for each of 6 subjects).

For testing we used a total 12 x 2 input and output files (2 x 2 input and output files for each of 6 subjects).

Total training data size: 48,50,019 rows. Total testing data size: 25,15,223 rows. ROC-AUC score had been used as the performance measure.

2.2 Preprocessing

The Fig-1 shows the fluctuations in the data. In order to smoothen the curve and to reduce the short-term fluctuations, rolling mean was used with window size as 4. Rolling mean replaces each element with the mean of adjacent 4 elements thus bringing consistency in the signal and ensuring removal of instantaneous high spikes to a good extent. It can be seen that the processed data is smooth and has lesser short-term fluctuations which is good for capturing the long-term sequences and trends.

2.3 Activation Functions Used

2.3.1 Sigmoid.

$$S(x) = 1/\left(1 + e^{-x}\right)$$

2.3.2 Relu.

$$F(x) = max\,(0, x)$$

2.4 Training Parameters

The training hyper-parameters chosen can be seen in Fig-4.They were kept constant during the training of all chosen Algorithms so that a comparative study too can be conducted.

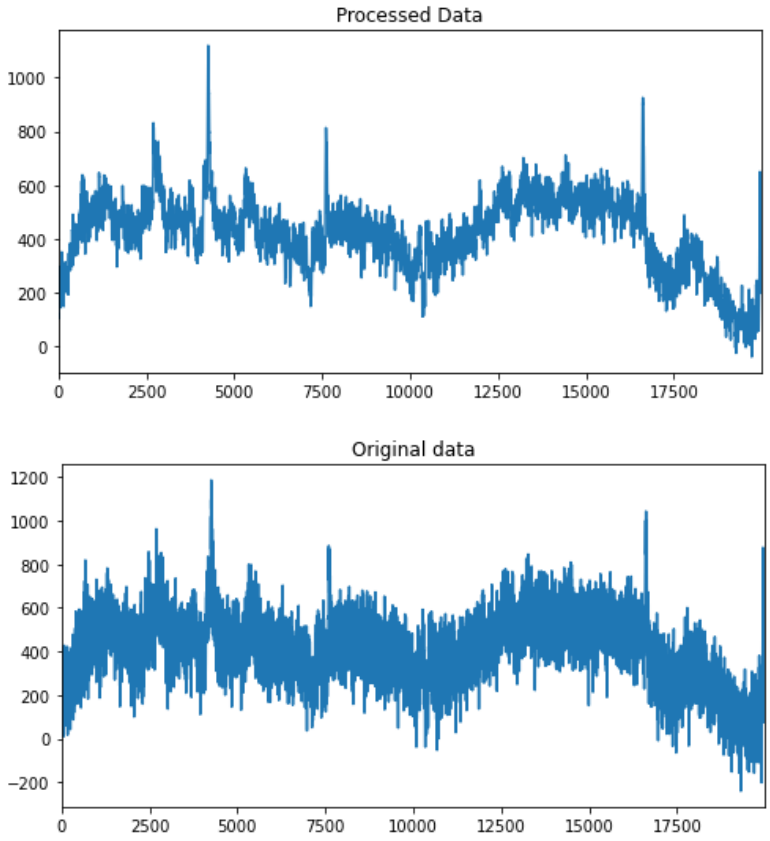


Figure 1: Signal Processing

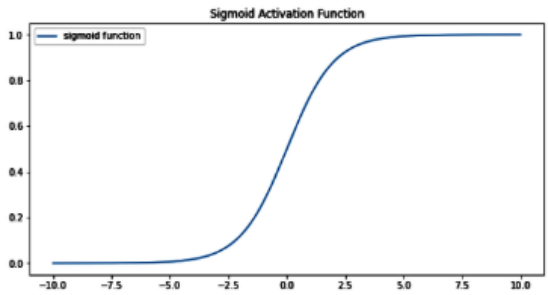


Figure 2: Sigmoid Activation Function

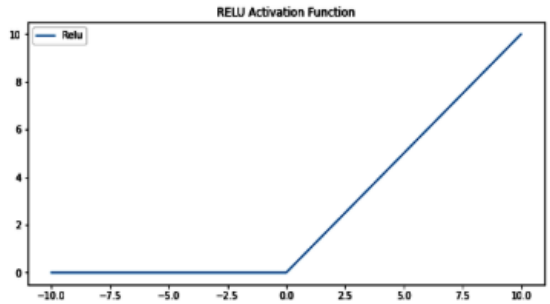


Figure 3: Relu Activation Function

2.5 ML and DL Techniques

2.5.1 Multinomial Logistic Regression. Multinomial Logistic Regression is a machine learning classification algorithm which is a generalized form of Logistic Regression. Simple Logistic Regression or the Binomial Logistic Regression is used to classify 2 classes (generally nominal classes). Multinomial Logistic Regression is used when there are more than 1 classes to be classified.

Logistic Regression (LR) is the most elementary machine learning classification algorithm. We selected it to get a baseline of performance. Since there are a total 6 target classes, hence Multinomial Logistic Regression was used.

Its performance was very poor because of the fact that even though it is a classification algorithm, still it has some limitations like it cannot train on a lot of features with considerable noise and in this case, the EEG signal data is sequential (time series data)

Training Hyper Parameters

Parameter	Chosen Value
Optimizer	Adam
Learning rate(s)	0.01, 0.005, 0.001, 0.0005
Batch size	32
Batches per epoch	5000
epochs	20 for each learning rate

Figure 4: Training Hyperparameters

so the output depends on the previous states. LR isn't meant for sequential data.

Still it was worth a try.

2.5.2 ANN. ANN (Artificial Neural Network) is a mathematical structure which was made after taking inspiration from the human brain itself. Just like the human brain has neurons for processing information, similarly, ANN has a number of processing units called the neurons which are spread out in an arbitrary number of layers and are used for capturing the relationship between the input data and the output data. It has one input layer for feeding the input data, one output layer for generating an output and an arbitrary number of intermediate layers which are called the hidden layer.

ANN is flexible and can effectively capture any kind of relationship (if provided enough training) may it be linear or nonlinear. It was selected to check its efficiency in a classification problem and to compare it with its other forms like RNN, LSTM and CNN. 6 Dense layers had been used. For re-centring and re-scaling, batch normalization layers had been used after each dense layer. The neurons in the layers were in a sequence: 512, 256, 128, 64, 32, 6.

Purelin activation function had been used in the first 4 dense layers. The RELU activation function had been used in the 5th dense layer. Sigmoid activation function had been used in the last layer.

2.5.3 RNN. RNN (Recurrent Neural Network) is a type of neural network in which the previous state is used for generating the new state. Unlike ANN, in which all the layers are independent of each other, RNN has a special feature of making use of previous states. Thus, RNN is said to have a memory which is very useful for capturing the relationship between the input and the output data, especially when the data is sequential.

EEG data is sequential data. Since RNN was especially made for capturing the relationship between the input and the output data when the data is sequential, so we selected it so that we can compare its performance with other types of neural networks. Simple RNN layer had been used as the input layer with 512 units.

5 Dense layers had been used. For re-centring and re-scaling, batch normalization layers had been used after each dense layer. The neurons in the dense layers were in a sequence: 256, 128, 64, 32, 6. Purelin activation function had been used in the first 3 dense layers. The RELU activation function had been used in the 4th dense layer. Sigmoid activation function had been used in the last layer.

Layer (type)	Output Shape	Param #
dense_6 (Dense)	(None, 20, 512)	16896
batch_normalization (BatchNo	(None, 20, 512)	2048
dense_7 (Dense)	(None, 20, 256)	131328
batch_normalization_1 (Batch	(None, 20, 256)	1024
dense_8 (Dense)	(None, 20, 128)	32896
batch_normalization_2 (Batch	(None, 20, 128)	512
dense_9 (Dense)	(None, 20, 64)	8256
batch_normalization_3 (Batch	(None, 20, 64)	256
flatten_1 (Flatten)	(None, 1280)	0
dense_10 (Dense)	(None, 32)	40992
batch_normalization_4 (Batch	(None, 32)	128
dense_11 (Dense)	(None, 6)	198

Total params: 234,534
Trainable params: 232,550
Non-trainable params: 1,984

Figure 5: ANN Architecture

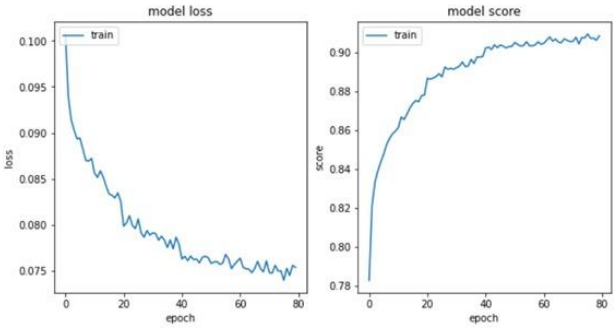


Figure 6: ANN Training Curves

EEG data is sequential data. Since RNN was especially made for capturing the relationship between the input and the output data when the data is sequential, so we selected it so that we can compare its performance with other types of neural networks. Simple RNN layer had been used as the input layer with 512 units.

5 Dense layers had been used. For re-centring and re-scaling, batch normalization layers had been used after each dense layer. The neurons in the dense layers were in a sequence: 256, 128, 64, 32, 6. Purelin activation function had been used in the first 3 dense layers. The RELU activation function had been used in the 4th dense layer. Sigmoid activation function had been used in the last layer.

2.5.4 CNN. CNN's have gained a lot of popularity because of its ability to extract features from image data. It uses convolution kernels for feature extraction from the input data. Each layer contains multiple filters and the number of filters is proportional to the complexity of the input data.

More filters and layers help in better feature extraction but too many filters and layers make it difficult to backpropagate the loss and hence the model doesn't learn.

Layer (type)	Output Shape	Param #
simple_rnn_22 (SimpleRNN)	(None, 20, 512)	279040
batch_normalization_35 (Batc	(None, 20, 512)	2048
dense_25 (Dense)	(None, 20, 256)	131328
batch_normalization_36 (Batc	(None, 20, 256)	1024
dense_26 (Dense)	(None, 20, 128)	32896
batch_normalization_37 (Batc	(None, 20, 128)	512
dense_27 (Dense)	(None, 20, 64)	8256
batch_normalization_38 (Batc	(None, 20, 64)	256
flatten_8 (Flatten)	(None, 1280)	0
dense_28 (Dense)	(None, 32)	40992
batch_normalization_39 (Batc	(None, 32)	128
dense_29 (Dense)	(None, 6)	198

Total params: 496,678
Trainable params: 494,694
Non-trainable params: 1,984

Figure 7: RNN Architecture

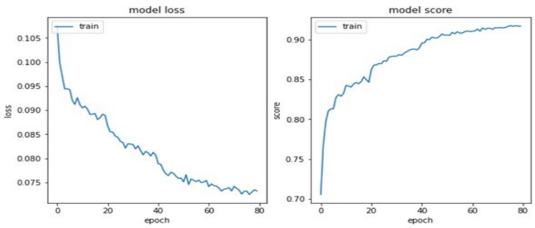


Figure 8: RNN Training Curves

Now EEG isn't any image type of data and moreover its sequential, still CNN can be applied because it shows good performance in time series prediction with appropriate window size.

Window length of sequential data during training was chosen to be 20 as 10 led to overfitting and 30 led to underfitting. Dropout layers were used to prevent overfitting. Batch Normalization has been used to improve performance of backpropagation as it does re-centring and re-scaling, thus speeding the training time.

CNN was Flattened after 4 layers. Then these parameters were fed to the final output layer. This output layer consists of a fully connected layer (6 nodes) with sigmoid activation. Each node of the Flattened 4th layer of CNN is connected with the 6 output nodes of the output layer, and this ensures adequate merger of features extracted by CNN convolutions.

2.5.5 LSTM. LSTMs are special kind of RNNs which are capable of learning long-short term dependencies of a time series data.

Each LSTM unit consists of 3 gates (input gate, output gate and forget gate). All these gates work collectively to ensure long term memory storage of sequential data. The forget gate decides what information is to be erased from the cell state by operating on inputs given to this unit. It uses the sigmoid function on the unit's inputs and then outputs a number between 0 and 1. (1 means keep it completely and 0 means discard it completely).

Layer (type)	Output Shape	Parameters	Hyper-parameters	
conv2d	(None, 12, 24, 32)	2624	kernel size filters	(9,9) 32
batch_normalization	(None, 12, 24, 32)	128		
dropout	(None, 12, 24, 32)	0	dropout %	10%
conv2d_1	(None, 7, 19, 48)	55344	kernel size filters	(6,6) 48
batch_normalization	(None, 7, 19, 48)	192		
Dropout	(None, 7, 19, 48)	0	dropout %	10%
conv2d_2	(None, 3, 15, 64)	76864	kernel size filters	(5,5) 64
batch_normalization	(None, 3, 15, 64)	256		
Dropout	(None, 3, 15, 64)	0	dropout %	10%
conv2d_3	(None, 1, 13, 128)	73856	kernel size filters	(3,3) 128
batch_normalization	(None, 1, 13, 128)	512		
Dropout	(None, 1, 13, 128)	0	dropout %	20%
flatten	(None, 1664)	0		
dense	(None, 6)	9990	nodes	6

Total params: 219,766
Trainable params: 219,222
Non-trainable params: 544

Figure 9: CNN Architecture

Then a decision is made about which values will be updated in the cell state. (via input gate).

Now, the raw LSTM (single layer) model cannot deal with multiple temporal series within the data so Multi-Layer LSTM architecture was used with relu activation and the last layer had softmax activation since the network needs to predict the probabilities of the output classes.

Manual tuning was performed to explore best model architecture for detection and classification. It was observed that those architectures which had layers in the powers of 2 with decreasing layer size performed way better than other architectures. A sequence length of 20 was chosen which gave better results as compared to other sequence lengths like 30 (led to underfitting), 10 (led to overfitting).

3 RESULTS AND DISCUSSIONS

Figure - 13 shows the final results of all the ML and DL algorithms. Figure - 14 shows a comparison graph of all the algorithms on the basis of ROC_AUC scores.

It can be seen from Figure - 13 and Figure - 14 that LR gave an ROC_AUC score of around 50% in both training and testing phase. This was already expected as LR cannot capture trends in sequential data as it deals with all the examples independently.

When we compare ANN, RNN, LSTM and CNN on the basis of their training ROC_AUC scores, we find that CNN gave the highest score of 98.23%, followed by LSTM with a score of 92.26%, then RNN with a score of 91.74% and at last ANN with a score of 90.84%. The order of complexity of models also corresponds to this scoring order. CNN and LSTM are more complex than RNN and further, RNN is more complex than ANN. Thus the ordering according to training scores can be attributed to complexity of the models, more complex models can capture relationships better than less complex models. Another factor that can explain the training performance order is the ability to capture relationships in sequential data. The dataset of EEG signals consists of sequential data. CNN can capture spatial relationships, while LSTM and RNN can capture temporal relationships in sequential data, making them better suited for classification of body gestures as compared to simple ANN.

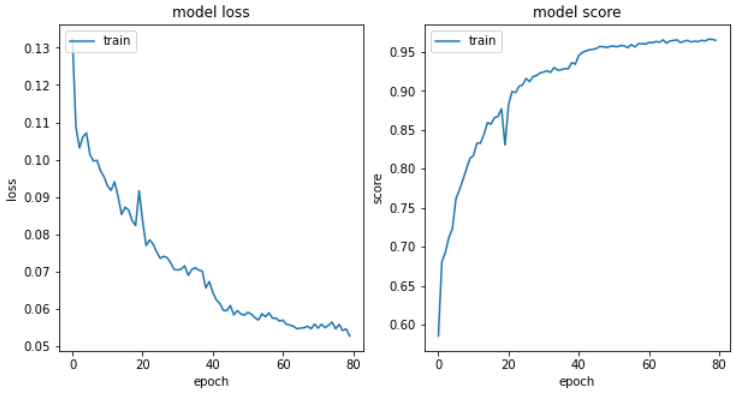


Figure 10: CNN Training Curves

Layer (type)	Output Shape	Parameters	Hyper-parameters	
lstm	(None, 20, 256)	295936	nodes	256
lstm_1	(None, 20, 128)	197120	nodes	128
lstm_2	(None, 20, 64)	49408	nodes	64
flatten	(None, 1280)	0		
dense	(None, 6)	7686	nodes	6

Total params: 550,150
Trainable params: 550,150
Non-trainable params: 0

Figure 11: LSTM Architecture

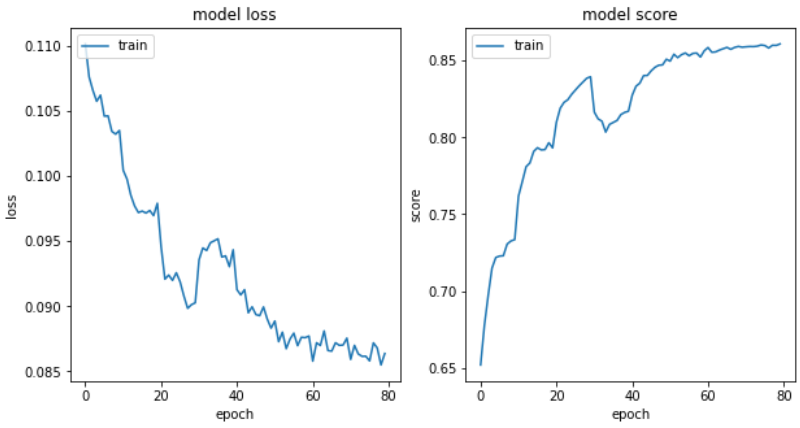


Figure 12: LSTM Training Curves

When we compare the algorithms on the basis of their testing scores, we find that ANN gave the best score of 87.22%, followed by CNN with a score of 85.31%, then RNN with a score of 83.42% and then LSTM with a score of 80.12%. This order is nearly the opposite of the training performance order. Overfitting can be one possible reason for the reversal of the performance order. More complex models tend to overfit the training data and hence give relatively poor results on the testing set due to loss of generalization. The performance order on the test set can also be explained using the principle of Occam's Razor, according to which, simple solutions are better than the complex ones.

Despite overfitting due to high complexity, CNN gave the second highest score on the test set, which suggests that its ability to capture spatial relationships in sequential data outweighed the

Final Results

Algorithm	Training AUC Score(%)	Testing AUC Score(%)
LR	49	50
RNN	91.74	83.42
ANN	90.84	87.22
LSTM	92.26	80.12
CNN	98.23	85.31

Figure 13: Final Results

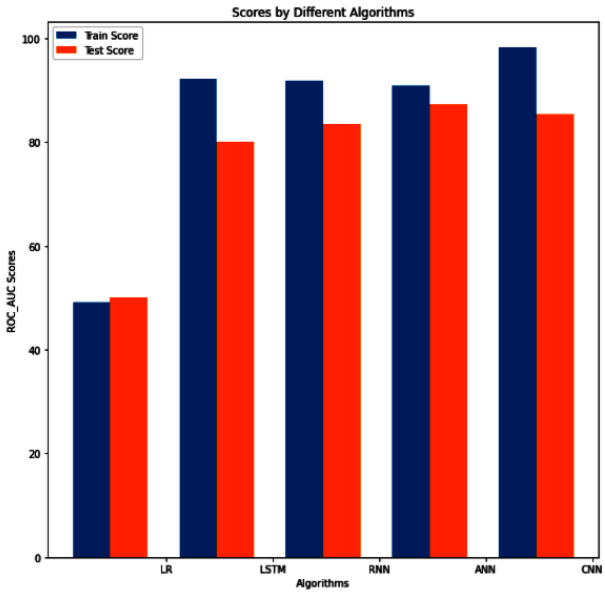


Figure 14: Comparative Analysis on the basis of ROC-AUC Score

negative impact of its overfitting on the training set. It also suggests that spatial relations are more prominent in the dataset than temporal relations and that is why CNN performed better than LSTM and RNN, which are capable of capturing temporal relationships.

4 CONCLUSION AND FUTURE WORK

We applied Multinomial Logistic Regression, Artificial Neural Network (ANN), Recurrent Neural Network (RNN), Long Short-Term Memory network (LSTM) and Convolutional Neural Network (CNN) on EEG brain waves data and classified the data into 6 categories of body gestures, namely, HandStart (starting the hand movement), FirstDigitTouch (touching a can with the hand), BothStartLoadPhase (grasping the can and starting to lift it), LiftOff (lifting the can), Replace (placing the can back), BothReleased (releasing the can). We compared all 5 algorithms by using ROC-AUC score as the performance metric. All the deep learning algorithms performed better than the machine learning algorithm, multinomial logistic regression.

CNN and LSTM performed the best on the training set by virtue of their high complexity and their ability to capture spatial and temporal relationships, respectively, in sequential data.

The performance score of CNN and LSTM got reduced on the test set because of overfitting on the training set. ANN gave the highest score on the test set because of its relatively less complexity which allowed it to maintain a level of generalization instead of overfitting on the training set. CNN gave the second highest score on the test set, which suggests that its ability to capture spatial relationships outweighed the negative impact of its overfitting on the training set. Hence, CNN can be considered as the overall best performing model for the purpose of decoding the brain waves for classification of body gestures.

For future work, we will regularize the deep learning models to reduce overfitting. We will further increase the training epochs for better comparison among the models. We will try to come up with a multi - architecture model for even better results. We also intend to use the best performing algorithm for diagnosis of the disease Schizophrenia using the EEG signals of the patients and for computing the stress levels of workers in real time using their EEG signals.

5 CITING RELATED WORK

In 1996, 9 prominent EEG stages were identified during sleep along with the durations and occurrences of each of these stages and the relationship amongst them was analyzed.[10]

In 2000, research was done on using EEG signals to measure correlation between cortex of brain and muscle activities and it was found that cortical beta and gamma rhythms are involved in the motor command.[2]

In 2006, D. O. Bos published research to establish suitable approach to determine optimal placement of limited electrodes on brain for emotion recognition using the captured EEG signals.[1]

In 2007, Knowledge based system for seizure detection in newborn infants was developed using features from band pass filtered EEG signals, via multi-layer neural network.[8]

In 2010, research was done on using EEG signals to classify human emotions using statistical features from the data via multilayer perceptron.[9]

In 2012, another methodology was introduced for rest stage ID which depended on the new norm of the American foundation of rest medication (AASM). The proposed approach utilizes time-recurrence examination and entropy measures to include extraction from a solitary electroencephalograph (EEG) channel. Component extraction was done from single EEG signals utilizing time-recurrence examination. Arbitrary Forest classifier was utilized for grouping the provisions removed from single EEG signal. Precision and kappa coefficient were utilized as execution measures

In 2012, research based on mental state classification using EEG based signals was carried out with Algorithms SVM and Random Forest.[4]

In 2017 a portable low-cost EEG system (MUSE) was developed to provide an excellent opportunity to improve the research on EEG signals. [3]

EEG has also been used to monitor and detect strokes [5], Epileptic seizures using time frequency analysis [7] and for rehabilitation of upper limb strokes when combined with robotic feedback.[6]

REFERENCES

[1] Bos, D.O., 2006. EEG-based emotion recognition. The Influence of Visual and Auditory Stimuli, 56(3), pp.1-17

[2] Mima Steger Salman Gerloff and Hallett, "EEG measurement of motor cortex control of muscle activity in humans," Clin. Neurophysiol., 2000.

[3] Krigolson, O.E., Williams, C.C., Norton, A., Hassall, C.D. and Colino, F.L., 2017. Choosing MUSE: Validation of a low-cost, portable EEG system for ERP research. Frontiers in neuroscience, 11, p.109.

[4] Plotnikov, A., Stakheika, N., De Gloria, A., Schatten, C., Bellotti, F., Berta, R., Fiorini, C. and Ansovini, F., 2012, July. Exploiting realtime EEG analysis for assessing flow in games. In 2012 IEEE 12th International Conference on Advanced Learning Technologies (pp. 688-689). IEEE.

[5] Jordan, K.G., 2004. Emergency EEG and continuous EEG monitoring in acute ischemic stroke. J. of Clinical Neurophys., 21(5), pp.341-352.

[6] Ang, K.K., Guan, C., Chua, K.S.G., Ang, B.T., Kuah, C., Wang, C., Phua, K.S., Chin, Z.Y. and Zhang, H., 2010, August. Clinical study of neurorehabilitation in stroke using EEG-based motor imagery braincomputer interface with robotic feedback. 2010 Annual International Conference of the IEEE (pp. 5549-5552).

[7] Tzallas, A.T., Tsipouras, M.G. and Fotiadis, D.I., 2009. Epileptic seizure detection in EEGs using time–frequency analysis. IEEE transactions on information technology in biomedicine, 13(5), pp.703- 710

[8] Aarabi, A., Grebe, R. and Wallois, F., 2007. A multistage knowledge based system for EEG seizure detection in newborn infants. Clinical Neurophysiology, 118(12), pp.2781-2797

[9] Chai, T.Y., Woo, S.S., Rizon, M. and Tan, C.S., 2010. Classification of human emotions from EEG signals using statistical features and neural network. In International (Vol. 1, No. 3, pp. 1-6). Penerbit UTHM

[10] Tanaka, H., Hayashi, M. and Hori, T., 1996. Statistical features of hypnagogic EEG measured by a new scoring system. Sleep, 19(9), pp.731-738.

[11] Fraiwan, L., Lweesy, K., Khasawneh, N., Wenz, H. and Dickhaus, H., 2012. Automated sleep stage identification system based on time– frequency analysis of a single EEG channel and random forest classifier. Computer methods and programs in biomedicine, 108(1), pp.10-19.



Journal of Modern Optics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tmop20

Design of 31 and 37 cores trench-assisted and air-hole assisted multi-core fibre for high-density space-division multiplexing

Garima Singh, Amin Khorshidahmad, Gurjit Kaur & Ahmad Atieh

To cite this article: Garima Singh, Amin Khorshidahmad, Gurjit Kaur & Ahmad Atieh (2022): Design of 31 and 37 cores trench-assisted and air-hole assisted multi-core fibre for high-density space-division multiplexing, Journal of Modern Optics, DOI: 10.1080/09500340.2022.2042746

To link to this article: https://doi.org/10.1080/09500340.2022.2042746

Published online: 04 Mar 2022.

Submit your article to this journal

Article views: 16

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tmop20

JOURNAL OF MODERN OPTICS
https://doi.org/10.1080/09500340.2022.2042746

Taylor & Francis
Taylor & Francis Group

Check for updates

Design of 31 and 37 cores trench-assisted and air-hole assisted multi-core fibre for high-density space-division multiplexing

Garima Singh[a], Amin Khorshidahmad[b], Gurjit Kaur[a] and Ahmad Atieh[b]

[a]Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India; [b]Optiwave Systems Inc., Ottawa, Ontario, K2E 8A7, Canada

ABSTRACT

In this paper, we have designed two multi-core fibres (MCFs) structures with 31 and 37 trench-assisted and air-hole-assisted cores with reduced inter-core crosstalk (XT). The air-hole pairs and triplets are placed between adjacent trench-assisted cores in order to further suppress modal field overlap. The XT analysis of the 31-core design underlines the significance of enclosing the outer cores with the air-hole structure. The effect of core pitch on the XT is also investigated. Placing the low-index air-hole shield closer to the trench enhances the modal confinement within relatively large trench-assisted core structures. As a result, the lowest possible core-pitch MCF (air-holes touching outer trench boundaries) yielded a minimum inter-core XT of between −60 dB and −70 dB for 100 km of fibre length. Therefore, our 37-core MCF design is suitable for high-density space-division multiplexing applications.

ARTICLE HISTORY
Received 24 August 2021
Accepted 8 February 2022

KEYWORDS
Multi-core fibre; trench-assisted (TA) core; air-assisted hole; MCF Crosstalk; space division multiplexing

1. Introduction

To meet the increasing capacity demands of next generation applications, a need to overcome the capacity crunch of current optical transmission systems drives the market [1]. Space division multiplexing (SDM) is recognized as a possible solution enabling high-capacity data transmission using multi-core fibres that enabled a record capacity of 1 *Pb/s* [2]. MCFs have multiple cores enclosed in a single cladding, which increases the transmission capacity by as many folds as the number of cores they have. To increase the number of cores, it is necessary to minimize the core-to-core distance parameter known as core pitch (△) and the distance between the fibre core and cladding known as cladding thickness (CT). However, reducing the core pitch and CT will result in increased crosstalk (XT) between the signals propagated in the different cores. Larger cladding can accommodate more cores, which is undesirable due to several mechanical and structural limitations.

Researchers in Ref. [3] recommended a maximum cladding diameter (CD) of 230 μm to guarantee a failure probability less than 1% over 20 years assuming a minimum bending radius of 60 mm and 200 turns. Trench-assisted cores have been proposed to reduce the crosstalk in MCFs [4–6]. In Ref. [7], a 7-core homogeneous MCF with trench-assisted pure silica cores arranged in hexagonal lattice and offered a crosstalk below −30 dB even after 10,000 *Km* propagation. In Ref. [8], the authors analysed theoretically the performance of 12-core MCF arranged in one ring. They showed that the single-ring structure has better crosstalk performance than hexagonal closed packed structure with crosstalk of −40 dB. In Ref. [9], the authors have designed 12-core and 19-core MCFs used to carry 1 *Pb/s/fibre*. The crosstalk parameter for a 100 *km* length was −40 dB and −30 dB, respectively.

Employing the air holes between the cores is also an attractive technique for reducing the crosstalk as air-holes creates a wall between cores. MCF with air-hole assisted fibres have been proposed to suppress the inter core crosstalk at least −20 dB in comparison to MCFs without air-holes [10,11]. In Ref. [12], the authors introduced an MCF structure that suppressed unwanted modes and reduced crosstalk to levels below −62 dB. The authors in [13] showed that XT can be further reduced by incorporating air holes between the cores with a trench profiles. In Ref. [14], the authors have illustrated the effect of air holes on crosstalk for a 19-core arranged in different rings and square lattice of 165 μm cladding MCF. They showed that crosstalk can be reduced by 20dB even after adding a single air hole. In Ref. [15],

CONTACT Gurjit Kaur gurjitkaur@dtu.ac.in; gurjeet_kaur@rediffmail.com Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India

© 2022 Informa UK Limited, trading as Taylor & Francis Group

Table 1. Characteristics of reported and designed MCFs.

No. of cores	CD(μm)	Core-air-hole arrangement	XT (dB)	References
12	200	Air-trench cores	−55 to 60	[18]
19	200	12 Air-holes on each core	−60	[19]
70	200	6 Air-holes on each core	−46	[2]
07	175	Air-trench cores	−50	[20]
06	125	6 Air-holes on each core	−55.5	[11]
07	192	6 Air-holes on each core	−60	[15]
07	125	6 Air-holes on each core	−45	[21]
19	250	1 Air-hole between each core	−20	[22]
31	200	Triangle lattice	−25	Proposed
37	200	Hexagon lattice	−70	Proposed

the authors have successfully suppressed the crosstalk by −40 and −60 dB for LP_{11} and LP_{01}, respectively. The authors in [16] have achieved low crosstalk by surrounding each trench-assisted core of a 7-core MCF in 192 μm cladding diameter (CD) with six air holes. In Ref. [17], the authors have proposed a 6-core fibre within a 125 μm cladding employing air hole-assisted structure. They maintained a crosstalk lower than −30 dB over 100 *km* fibre.

Researchers in the literature showed that increasing the number of cores result in extremely high crosstalk as adjacent cores will almost touch each other limiting the fibre core density. Moreover, there is a need to add air holes between cores to minimize crosstalk, which also limits the core density in the MCF. A summary of MCF designs in the literature is illustrated in Table 1.

In this work, we propose a novel 31 and 37 core trench-assisted and air-hole assisted MCFs. The cores are place in a hexagon or triangle lattice structure. Also, each core in the structure is surrounded by air holes placed in different ways to suppress crosstalk between the neighbouring cores. The design of the air-hole-assisted MCFs is adjusted to accommodate maximum number of cores with minimum crosstalk within 200 μm cladding diameter.

The theoretical and experimental MCFs as shown in Table 1 proves that placing air-holes between cores is an effective and reliable approach to reduce the crosstalk [19].

The remainder of this paper is arranged as follows. Section 2 discusses the design of the MCF structures. The crosstalk analysis of the proposed design structures is described in Section 3. The analysis of crosstalk suppression using air-hole assisted is discussed in Section 4. Finally, the conclusions are provided in Section 5.

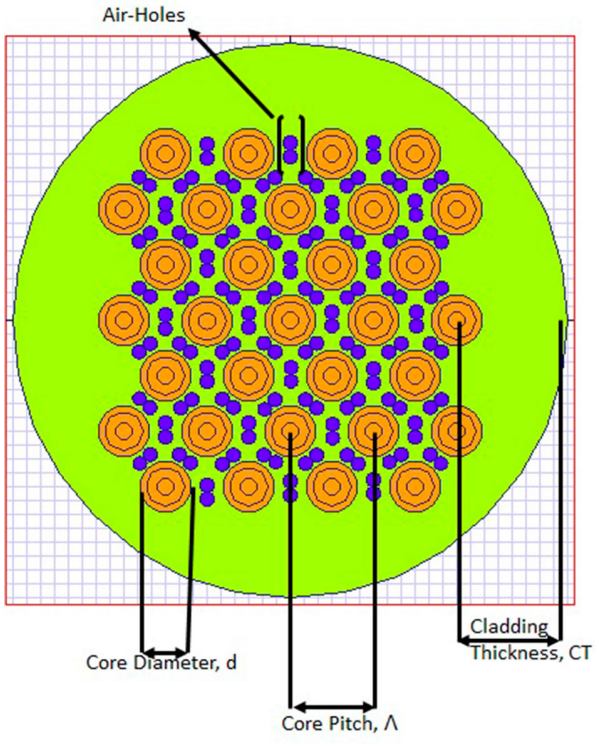


Figure 1. 31 Trench-assisted cores with air holes arranged in triangle lattice.

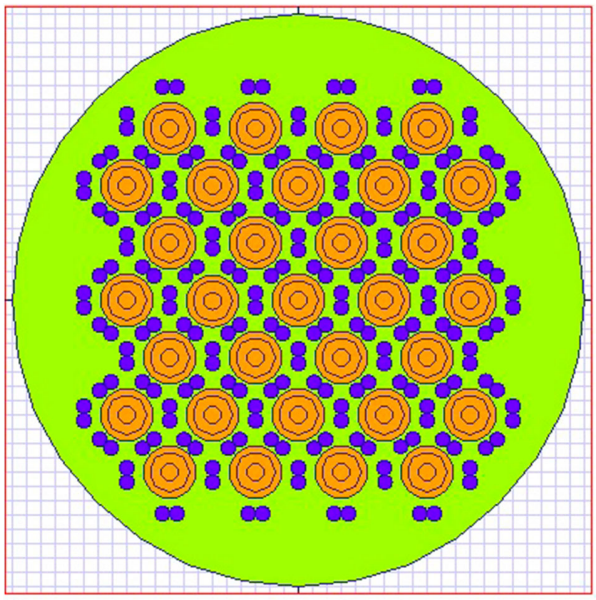


Figure 2. 31 Trench-assisted cores fully surrounded by air holes in triangle lattice.

2. Air-hole assisted trench-assisted MCF design

High-density multi-core fibre structures have been designed with 31-and 37- trench assisted air-hole assisted cores are placed in hexagon and triangle lattice structure

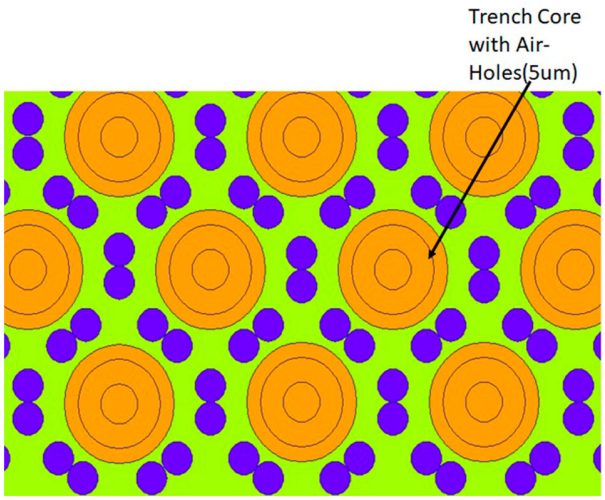


Figure 3. Single core view of 31 trench-assisted cores with air holes.

in cladding diameter of 200 μm as shown in Figures 1–3. Trench-assisted refractive index profile is used to suppresses the overlapping of the electromagnetic fields among the cores which results in suppressed crosstalk. However, air-holes have been placed uniquely in pairs of three on horizontal axis and two on diagonal side. The particularity of our design is the air-hole placement which forms a groove-like structure around the cores, which can evidently prevent the leakage of core energy and isolate the light from other cores. The air-holes are placed in such a manner to minimize their number and achieve minimum crosstalk.The MCF structures having air-holes all around the cores are also good in manufacturability [18,19]. Subsequently, six different MCF designs have been considered in this work; two designs with 31-cores as shown in Figures 1 and 2 and four designs with 37-cores with different horizontal and diagonal core pitch as shown in Table 2. Figure 1 shows the cores along the edges have no air-holes as proposed in Refs [23,24] to minimize the micro-bending losses in outer cores using minimum CT and core pitch. The reduction in the value of CT may increase the bending and confinement losses leading to transmission loss in outer cores [25]. Therefore, the values of CT and core pitch are chosen precisely in this work and we have covered all the trench-assisted cores with air-holes as shown in Figures 2 and 3.

Table 2. Designed fibre parameters.

Parameters	Value
Cladding diameter	200 μm
Cladding thickness (CT)	25μm
Core diameter(d)	18 μm
Hole diameter	5 μm
r_1	3 μm
r_2	$1.25 * r_1$=3.75 μm
Trench width (w_t)	$0.75 * r_1$=2.25 μm
Relative R.I.(Δ_1,Δ_2)	0.304%, 0.69%
Cladding Index	1.45
Core Index	1.4544
Trench Index	1.4399

Full vector finite element method is used in the simulation of the different designs of MCF structures. Table 2 illustrates the parameters used for the different designs investigated in this work. The core pitch and cladding thickness parameters are mainly varied to accommodate these designs for 31-cores and 37-cores. The core pitch parameter is defined using (Λ_x and Λ_{dig}) parameters shown in the table.

3. Crosstalk analysis

The optimal design parameters summarized in Tables 1 and 3 are used to achieve high-density multi-core fibre structures. When optical power is launched into one core, it will coupled to neighbouring cores during propagation. The inter-core crosstalk (ICXT) parameter is described by Equation (1). IXCT is the one of the main parameters

Table 3. Characteristics of designed MCFs.

Cores	Air-hole arrangement	Λ_x (μm)	Λ_{dig} (μm)	CT (μm)
31	Partially surrounded	30	25	40
31	Fully surrounded	30	25	40
37	Fully surrounded	25	23.585	25
37	Fully surrounded	25	25	25
37	Fully surrounded	23.585	25	25
37	Fully surrounded	23.585	23.585	25

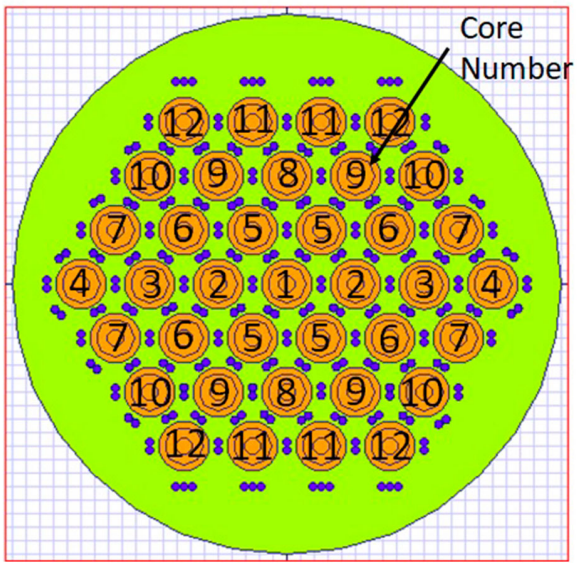


Figure 4. 37 Trench assisted cores with air holes arranged in hexagon lattice.

that needs to be investigated for MCFs.

$$XT = 10\log_{10}\left(\frac{P'}{P}\right)\,\mathrm{d}B, \tag{1}$$

where P and $\acute{P}$ represents the output power from the input and neighbouring core, respectively [26].

The structural parameters of the MCF need to be also optimized to achieve minimum inter-core crosstalk values. The average power in core m at a point z, $P_m(z)$ is defined as [27]

$$P_m(z) = K_{mn}^2 P_n(0)\int_0^z \mathrm{d}\eta \times\left[\int_{-\eta}^{z-\eta}\exp\left(j\Delta\beta'_{mn}\zeta\right)R(\zeta)\,\mathrm{d}\zeta\right]. \tag{2}$$

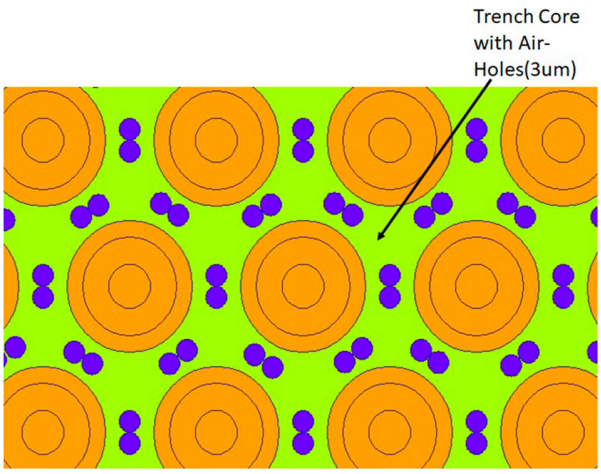


Figure 5. Single core view of 37 trench-assisted cores with air holes.

where K_{mn} is field coupling coefficient, $\Delta\beta_{mn}$ is propagation constant difference between core m and core n and $R(\zeta)$ is autocorrelation function of length ζ. The field profiles for each set of adjacent cores m and n are calculated first using the commercial modes solver FEM offered by OptiMode. The field coupling coefficient parameter K_{mn} is given by [27].

$$K_{mn} = \frac{\omega\varepsilon_0\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(N^2 - N_n^2\right)E_m^*.E_n\,\mathrm{d}x\,\mathrm{d}y}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}U_z.\left(E_m^*H_m + E_mH_m^*\right)\,\mathrm{d}x\,\mathrm{d}y}. \tag{3}$$

The coupling coefficients are used to calculate the average power coupling coefficient (PCC) based on the exponential autocorrelation model assuming a correlation length of 50 mm to get minimum inter-core XT [27]. The PCCs are used to analytically calculate the crosstalk as explained using [27].

$$h_{mn} = \frac{2k_{mn}^2 R_b}{\beta\Lambda}, \tag{4}$$

Therefore, the average crosstalk between the cores m and n can be written as [7]

$$XT \cong h_{mn}L \cong \frac{2k_{mn}^2 R_b}{\beta\Lambda}L, \tag{5}$$

where R_b is the bending radius and L is the length of fibre. Equation (5) shows the dependence of average inter-core crosstalk (ICXT) on bending radius, core pitch, fibre length and mode coupling coefficient. The ICXT is inversely proportional to the effective refractive index and the propagating wavelength.

4. MCF inter-core crosstalk analysis

The inter-core crosstalk for the different MCF structures are calculated and analysed in this section. The ICXT

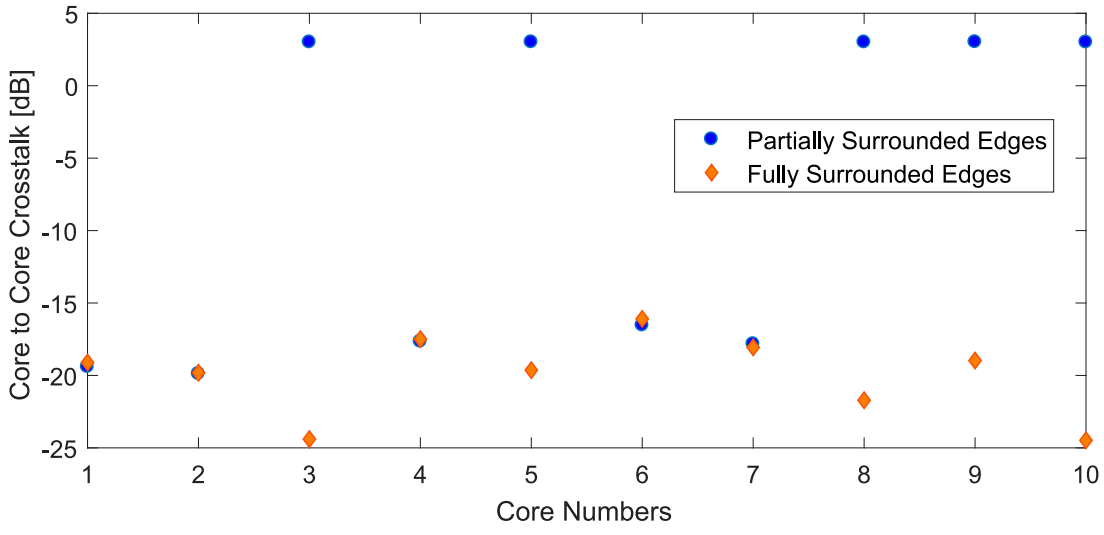


Figure 6. Crosstalk for 31 trench-assisted cores for partially and fully surrounded edges.

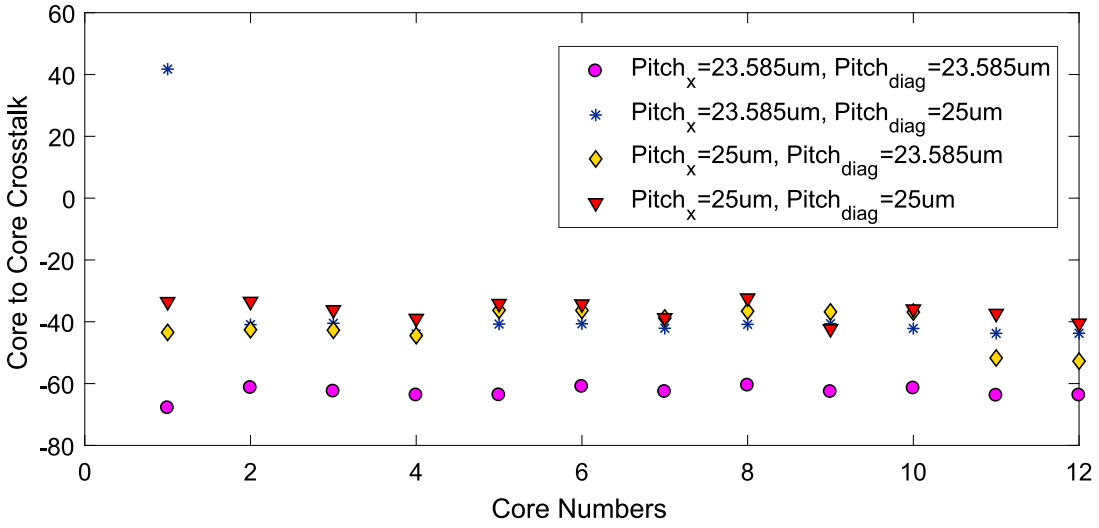


Figure 7. Crosstalk for 37 trench-assisted cores for different core pitch.

parameter is crucial for identifying the maximum possible spatial density of the MCF structure that can be achieved. Based on the symmetry of the designed structures, there are 12 distinct cores combinations in the 37-core structure as shown in Figure 3. Each core has an aureole of air holes that is created to suppress the crosstalk for the structure in 200μm as shown in Figure 5. We have calculated the crosstalk for each core considering all the core pitches for 100 km of fibre length as given in Table 3. Based on the design symmetry for the 31-cores design, there are 10 distinct cores. The ICXT is calculated for both cases of partially surrounded and fully surrounded edge cores as shown in Figure 4. The results shown in Figure 6 and 7 illustrate that the inner cores like 1, 2, 4, 6, 7 and 9 has the same crosstalk for both cases while the cores at edges like 3, 5 and 10 have decreased crosstalk by at least 25 dB for the fully surrounded air-holes case.

The result shows that in spite of reducing the core pitch, the achieved crosstalk can be as low as −40 dB compared to the maximum investigated core pitch. These results show that we can accommodate more number of cores by reducing the core pitch when surrounding each core with air-holes. Up to 20 cores have been placed in CD of 200 um with minimum XT of −60 dB demonstrated in the literature. However, we have successfully placed 37-cores using trench assisted and air-hole assisted in the same CD with a minimum XT of −70 dB.

5. Conclusion

We have designed a 31-core and 37-core trench-assisted air-holes assisted multi-core fibres. We found that the smallest pitch offered the best inter-core crosstalk values. The crosstalk of the minimum pitch value is reduced by −40 dB in comparison with the largest investigated core pitch. This is because of the use of air-holes placed between the cores. As a result, more cores can be accommodated inside the cladding with reduced core pitch even down to 20 μm with air-holes. Otherwise, without air holes, core pitch should be at least 40 μm. The presented analytical model provides a powerful tool for designing high-count homogeneous MCFs as high as 37 trench-assisted cores for optimum CD with the minimum XT of −70 dB for 100 km of fibre length.

Acknowledgments

The authors would also like to thank the anonymous reviewers for the feedback.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The authors would like to thank the “Research project grant to the faculty of DTU” under Delhi Technological University for its financial support of this work under file no. DTU/Council/BOM-AC/Notification/31/2018/5738.

References

[1] Kaur G, Srivastava D, Singh P, et al. Development of a novel hybrid PDM/OFDM technique for FSO system and its performance analysis. Opt Laser Technol. 2019;109:256–262.

[2] Watanabe T, Kokubun Y. Ultra-large number of transmission channels in space division multiplexing using few-mode multi-core fiber with optimized air-hole-assisted

double-cladding structure. Opt Express. 2014;22(7): 8309–8319.
[3] Matsuo S, Takenaga K, Sasaki Y, et al. High-spatial-multiplicity multicore fibers for future dense space-division-multiplexing systems. J Light Technol. 2016 Mar;34(6):1464–1475.
[4] Takara H, Sano A, Kobayashi T, et al. 1.01-Pb/s (12 SDM/222 WDM/456 Gb/s) crosstalk-managed transmission with 91.4-b/s/Hz aggregate spectral efficiency. European Conference and Exhibition on Optical Communication; Optical Society of America; 2012, Sept. p. Th-3.
[5] Hayashi T, Taru T, Shimakawa O, et al. Characterization of crosstalk in ultra-low-crosstalk multi-core fiber. J Light Technol. 2011;30(4):583–589.
[6] Matsuo S, Sasaki Y, Akamatsu T, et al. 12-core fiber with one ring structure for extremely large capacity transmission. Opt Express. 2012;20(27):28398–28408.
[7] Hayashi T, Taru T, Shimakawa O, et al. Design and fabrication of ultra-low crosstalk and low-loss multi-core fiber. Opt Express. 2011;19(17):16576–16592.
[8] Kumar D, Rakesh R. Optimal design for crosstalk analysis in 12-core 5-LP mode homogeneous multicore fiber for different lattice structure. Optical Fiber Technol. 2018;41:95–103.
[9] Matsuo S, Sasaki Y, Ishida I, et al. Recent progress on multi-core fiber and few-mode fiber. Optical Fiber Communication Conference; Optical Society of America; 2013 Mar. p. OM3I-3.
[10] Yao B, Ohsono K, Shiina N, et al. Reduction of crosstalk by hole-walled multi-core fibers. Optical Fiber Communication Conference; Optical Society of America; 2012, Mar. p. OM2D-5.
[11] Sakamoto T, Saitoh K, Hanzawa N, et al. Crosstalk Suppressed Hole-assisted 6-core Fiber with Cladding Diameter of 125 μ m. IET Conference Proceedings. The Institution of Engineering and Technology; 2013 Sept.
[12] Watanabe T, Kokubun Y. High density and low cross talk design of heterogeneous multi-core fiber with air hole assisted double cladding. OptoElectronics and Communications Conference and Photonics in Switching. Optical Society of America; 2013, Jun..
[13] Nozoe S, Fukumoto R, Sakamoto T, et al. Low crosstalk 125 μ m-cladding multi-core fiber with limited air-holes fabricated with over-cladding bundled rods technique. 2017 Optical Fiber Communications Conference and Exhibition (OFC); IEEE; 2017 Mar. p. 1–3.
[14] Goto Y, Tsujikawa K, Aozasa S, et al. Impact of air hole on crosstalk suppression and spatial core density of multi-core fiber. J Light Technol. 2018;36(20):4819–4825.
[15] Xia C, Amezcua-Correa R, Bai N, et al. Hole-assisted few-mode multicore fiber for high-density space-division multiplexing. IEEE Photonics Technol Lett. 2012;24(21): 1914–1917.
[16] Kaur G, Dhamania M, Tomar P, et al. Efficient "Integration of High-Order Models Using an FDTD?TDMA Method for Error Minimization". International Conference on Communications and Cyber Physical Engineering 2018; Singapore: Springer; 2018 Jan. p. 311–323.
[17] Sakamoto T, Saitoh K, Hanzawa N, et al. Crosstalk suppressed hole-assisted 6-core fiber with cladding diameter of 125 ?m. In 39th European Conference and Exhibition on Optical Communication (ECOC 2013); IET; 2013 Sept. p. 1–3.
[18] Xie Y, Pei L, Zheng J, et al. Low-DMD and low-crosstalk few-mode multi-core fiber with air-trench/holes assisted graded-index profile. Opt Commun. 2020;474:126155.
[19] Ziolowicz A, Szymanski M, Szostkiewicz L, et al. Hole-assisted multicore optical fiber for next generation telecom transmission systems. Appl Phys Lett. 2014;105(8):081106.
[20] Xie Y, Pei L, Zheng J, et al. Impact analysis of a dense hole-assisted structure on crosstalk and bending loss in homogeneous few-mode multi-core fibers. Opt Express. 2020;28(16):23806–23819.
[21] Ziolowicz A, Szostkiewicz L, Ostrowski L, et al. Overcoming the capacity crunch: ITU-T G. 657. B3 compatible 7-core and 19-core hole-assisted fibers. Next-Generation Optical Communication: Components, Sub-Systems, and Systems VI. Vol. 10130, International Society for Optics and Photonics; 2017 Jan. p. 101300C.
[22] Goto Y. Impact of air hole on core-to-core crosstalk suppression. Proceedings of the International Wire and Cable Symposium; 2015 Oct. p. 161–166.
[23] Sasaki Y, Akenaga K, Matsuo S, et al. Few-mode multicore fibers for long-haul transmission line. Opt Fiber Technol. 2017;35:19–27.
[24] Tobita Y, Fujisawa T, Sakamoto T, et al. Optical design of 4LP-mode multicore fibers for high spatial multiplicity. Opt Express. 2017;25:5697–5709.
[25] Goto Y, Nakajima K, Matsui T, et al. Influence of cladding thickness on transmission loss and relationship with multi core fiber. J Light Technol. 2015;33:4942–4949.
[26] Saitoh K, Matsuo S. Multicore fiber technology. J Light Technol. 2016;34(1):55–66.
[27] Koshiba M, Saitoh K, Takenaga K, et al. Analytical expression of average power-coupling coefficients for estimating intercore crosstalk in multicore fibers. IEEE Photonics J. 2012;4(5):1987–1995.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/339681027

Detection of Flooding Attacks Using Multivariate Analysis

Chapter · March 2020

DOI: 10.1007/978-3-030-43192-1_36

CITATIONS
0

READS
39

2 authors:



Priyanka Meel
Delhi Technological University
36 PUBLICATIONS **171** CITATIONS

SEE PROFILE



Tanmay Singh
ixigo
5 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Fake News and Rumour Detection View project

All content following this page was uploaded by Priyanka Meel on 14 January 2022.

The user has requested enhancement of the downloaded file.

Springer Link

Detection of Flooding Attacks Using Multivariate Analysis

International conference on Computer Networks, Big data and IoT

ICCBI 2019: Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI - 2019) pp 314-324 | Cite as

- Priyanka Meel (1) Email author (priyankameel@dtu.ac.in)
- Tanmay Singh (1)

1. Department of Information Technology, Delhi Technological University, , New Delhi, India

Conference paper
First Online: 05 March 2020

- 517 Downloads

Part of the Lecture Notes on Data Engineering and Communications Technologies book series (LNDECT, volume 49)

Abstract

In this paper, we propose a multivariate statistical analysis method namely the Hotelling's T^2 Method for the analysis of common network flooding attacks. The method analyses the behavior of system resources and network protocols and builds a baseline profile for its normal operation. We validated the proposed mechanism by carrying out flooding attacks on a wired network with Windows. We generated and sent attack packets through codes to a host machine, analyzed them (using Wireshark) and used a multivariate statistical method for testing the attack. This method effectively differentiates between normal and attack traffic and sets an alert in case of any abnormality in behavior.

Keywords

Abnormality distance metric Anomaly detection Multivariate statistical analysis

This is a preview of subscription content, log in to check access.

References

1. Forouzan, B.A.: Data Communications and Networking (McGraw-Hill Forouzan Networking). McGraw-Hill Higher Education (2007)
Google Scholar (https://scholar.google.com/scholar?q=Forouzan%2C%20B.A.%3A%20Data%20Communications%20and%20Netw

orking%20%28McGraw-Hill%20Forouzan%20Networking%29.%20McGraw-Hill%20Higher%20Education%20%282007%29)

2. Biswas, A.: Impact Analysis of System and Network Attacks (2008). All Graduate Theses and Dissertations. Paper 199. http://digitalcommons.usu.edu/etd/199 (http://digitalcommons.usu.edu/etd/199)

3. Li, M., Chi, C.H., Jia, W., Zhao, W., Zhou, W., Cao, J., Long, D., Meng, Q.: Decision analysis of statistically detecting distributed denial-of-service flooding attacks. Int. J. Inf. Technol. Decis. Making **2**(3), 397–405 (2003)
CrossRef (https://doi.org/10.1142/S0219622003000720)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Decision%20analysis%20of%20statistically%20detecting%20distributed%20denial-of-service%20flooding%20attacks&author=M.%20Li&author=CH.%20Chi&author=W.%20Jia&author=W.%20Zhao&author=W.%20Zhou&author=J.%20Cao&author=D.%20Long&author=Q.%20Meng&journal=Int.%20J.%20Inf.%20Technol.%20Decis.%20Making&volume=2&issue=3&pages=397-405&publication_year=2003)

4. Schildt, H.: The Complete Reference Java 2, 5th edn (2002)
Google Scholar (https://scholar.google.com/scholar?q=Schildt%2C%20H.%3A%20The%20Complete%20Reference%20Java%202%2C%205th%20edn%20%282002%29)

5. Xiaoming, L., Sejdini, V., Chowdhury, H.: Denial of Service (DoS) attack with UDP Flood (2010)
Google Scholar (https://scholar.google.com/scholar?q=Xiaoming%2C%20L.%2C%20Sejdini%2C%20V.%2C%20Chowdhury%2C%20H.%3A%20Denial%20of%20Service%20%28DoS%29%20attack%20with%20UDP%20Flood%20%282010%29)

6. Ligh, M., Adair, S., Hartstein, B., Richard, M.: Malware analyst's cookbook and DVD: tools and techniques for fighting malicious code (2011)
Google Scholar (https://scholar.google.com/scholar?q=Ligh%2C%20M.%2C%20Adair%2C%20S.%2C%20Hartstein%2C%20B.%2C%20Richard%2C%20M.%3A%20Malware%20analyst%E2%80%99s%20cookbook%20and%20DVD%3A%20tools%20and%20techniques%20for%20fighting%20malicious%20code%20%282011%29)

7. Alam, M.O., Adnan, A., Aktaruzzaman, A.K.M.: TCP SYN Flood DoS Attack Experiments in Wireless Networks (2007)
Google Scholar (https://scholar.google.com/scholar?q=Alam%2C%20M.O.%2C%20Adnan%2C%20A.%2C%20Aktaruzzaman%2C%20A.K.M.%3A%20TCP%20SYN%20Flood%20DoS%20Attack%20Experiments%20in%20Wireless%20Networks%20%282007%29)

8. Noureldien, N.A.: Protecting web servers from DoS/DDoS flooding attacks. A technical overview (2002)
Google Scholar (https://scholar.google.com/scholar?q=Noureldien%2C%20N.A.%3A%20Protecting%20web%20servers%20from%20DoS%2FDDoS%20flooding%20attacks.%20A%20technical%20overview%20%282002%29)

9. Antoniou, S.: The ping of death and other dos network attacks, 14 May 2009
Google Scholar (https://scholar.google.com/scholar?q=Antoniou%2C%20S.%3A%20The%20ping%20of%20death%20and%20other%20dos%20network%20attacks%2C%2014%20May%202009)

10. Eddy, W.M., Verizon Federal Network Systems: Defenses against TCP SYN flooding attacks. Internet Protocol J. **9**(4), 2–16 (2006)
Google Scholar (https://scholar.google.com/scholar?q=Eddy%2C%20W.M.%2C%20Verizon%20Federal%20Network%20Systems%3A%20Defenses%20against%20TCP%20SYN%20flooding%20attacks.%20Internet%20Protocol%20J.%209%284%29%2C%202%E2%80%9316%20%282006%29)

11. Wireshark User's Guide (2008)
Google Scholar (https://scholar.google.com/scholar?q=Wireshark%20User%E2%80%99s%20Guide%20%282008%29)

12. Jing, X., Yan, Z., Jiang, X., Pedrycz, W.: Network traffic fusion and analysis against DDoS flooding attacks with a novel reversible sketch. In: Information Fusion 2019 (2019)
Google Scholar (https://scholar.google.com/scholar?q=Jing%2C%20X.%2C%20Yan%2C%20Z.%2C%20Jiang%2C%20X.%2C%20Pedrycz%2C%20W.%3A%20Network%20traffic%20fusion%20and%20analysis%20against%20DDoS%20flooding%20attacks%20with%20a%20novel%20reversible%20sketch.%20In%3A%20Information%20Fusion%202019%20%282019%29)

13. Zheng, J., Li, Q., Gu, G., Cao, J., Yau, D.K., Wu, J.: Realtime DDoS defense using COTS SDN switches via adaptive correlation analysis. IEEE Trans. Inf. Forensics Secur. **13**(7), 1838–1853 (2018)
CrossRef (https://doi.org/10.1109/TIFS.2018.2805600)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Realtime%20DDoS%20defense%20using%20COTS%20SDN%20switches%20via%20adaptive%20correlation%20analysis&author=J.%20Zheng&author=Q.%20Li&author=G.%20Gu&author=J.%20Cao&author=DK.%20Yau&author=J.%20Wu&journal=IEEE%20Trans.%20Inf.%20Forensics%20Secur.&volume=13&issue=7&pages=1838-1853&publication_year=2018)

14. Kwon, D., Kim, H., An, D., Ju, H.: DDoS attack volume forecasting using a statistical approach. In: 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM) (2017)
Google Scholar (https://scholar.google.com/scholar?q=Kwon%2C%20D.%2C%20Kim%2C%20H.%2C%20An%2C%20D.%2C%20Ju%2C%20H.%3A%20DDoS%20attack%20volume%20forecasting%20using%20a%20statistical%20approach.%20In%3A%202017%20IFIP%2FIEEE%20Symposium%20on%20Integrated%20Network%20and%20Service%20Management%20%28IM%29%20%282017%29)

15. Hoque, N., Bhattacharyya, D.K., Kalita, J.K.: A novel measure for low-rate and high-rate DDoS attack detection using multivariate data analysis. In: 2016 8th International Conference on Communication Systems and Networks (COMSNETS) (2016)
Google Scholar (https://scholar.google.com/scholar?q=Hoque%2C%20N.%2C%20Bhattacharyya%2C%20D.K.%2C%20Kalita%2C%20J.K.%3A%20A%20novel%20measure%20for%20low-rate%20and%20high-rate%20DDoS%20attack%20detection%20using%20multivariate%20data%20analysis.%20In%3A%202016%208th%20International%20Conference%20on%20Communication%20Systems%20and%20Networks%20%28COMSNETS%29%20%282016%29)

16. Tan, Z., Jamdagni, A., He, X., Nanda, P., Liu, R.P.: A system for denial-of-service attack detection based on multivariate correlation analysis. IEEE Trans.

Parallel Distrib. Syst. **25**, 447–456 (2014)
CrossRef (https://doi.org/10.1109/TPDS.2013.194)
Google Scholar (http://scholar.google.com/scholar_lookup?title=A%20system%20for%20denial-of-service%20attack%20detection%20based%20on%20multivariate%20correlation%20analysis&author=Z.%20Tan&author=A.%20Jamdagni&author=X.%20He&author=P.%20Nanda&author=RP.%20Liu&journal=IEEE%20Trans.%20Parallel%20Distrib.%20Syst.&volume=25&pages=447-456&publication_year=2014)

17. Bogdanoski, M., Shuminoski, T., Risteski, A.: Analysis of the SYN Flood DoS Attack (2013)
Google Scholar (https://scholar.google.com/scholar?q=Bogdanoski%2C%20M.%2C%20Shuminoski%2C%20T.%2C%20Risteski%2C%20A.%3A%20Analysis%20of%20the%20SYN%20Flood%20DoS%20Attack%20%282013%29)

18. Tan, Z., Jamdagni, A., He, X., Nanda, P., Liu, R.P.: Denial-of-service attack detection based on multivariate correlation analysis. In: Lu, B.L., Zhang, L., Kwok, J. (eds.) Neural Information Processing, ICONIP 2011. LNCS, vol. 7064. Springer, Heidelberg (2011)
CrossRef (https://doi.org/10.1007/978-3-642-24965-5_85)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Denial-of-Service%20Attack%20Detection%20Based%20on%20Multivariate%20Correlation%20Analysis&author=Zhiyuan.%20Tan&author=Aruna.%20Jamdagni&author=Xiangjian.%20He&author=Priyadarsi.%20Nanda&author=Ren%20Ping.%20Liu&pages=756-765&publication_year=2011)

19. Praveena, A., Smys, S.: Anonymization in social networks: a survey on the issues of data privacy in social network sites. J. Int. J. Eng. Comput. Sci. **5**(3), 15912–15918 (2016)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Anonymization%20in%20social%20networks%3A%20a%20survey%20on%20the%20issues%20of%20data%20privacy%20in%20social%20network%20sites&author=A.%20Praveena&author=S.%20Smys&journal=J.%20Int.%20J.%20Eng.%20Comput.%20Sci.&volume=5&issue=3&pages=15912-15918&publication_year=2016)

Copyright information

© Springer Nature Switzerland AG 2020

About this paper

Cite this paper as:

Meel P., Singh T. (2020) Detection of Flooding Attacks Using Multivariate Analysis. In: Pandian A., Palanisamy R., Ntalianis K. (eds) Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI - 2019). ICCBI 2019. Lecture Notes on Data Engineering and Communications Technologies, vol 49. Springer, Cham. https://doi.org/10.1007/978-3-030-43192-1_36

- First Online 05 March 2020
- DOI https://doi.org/10.1007/978-3-030-43192-1_36
- Publisher Name Springer, Cham
- Print ISBN 978-3-030-43191-4

- Online ISBN 978-3-030-43192-1
- eBook Packages Intelligent Technologies and Robotics Intelligent Technologies and Robotics (R0)

- Buy this book on publisher's site
- Reprints and Permissions

Personalised recommendations

SPRINGER NATURE

© 2020 Springer Nature Switzerland AG. Part of Springer Nature.

Not logged in Delhi Technological University (2000607895) - Convener, UGC-Infonet Digital Library Consortium (3000132959) - INDEST-AICTE-Level III (3000168247) - Information and Library Network (INFLIBNET) Centre (3994475188) 14.139.251.100

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/TIM.2022.3154832, IEEE Transactions on Instrumentation and Measurement

DIAT-RadHARNet: A Lightweight DCNN for Radar based Classification of Human Suspicious Activities

Mainak Chakraborty, *Graduate Student Member, IEEE*, Harish C. Kumawat, *Graduate Student Member, IEEE*, Sunita Vikrant Dhavale, *Member, IEEE*, and Arockia Bazil Raj A, *Member, IEEE*

***Abstract*—Recognizing suspicious human activities is one of the critical requirements for national security considerations. Nowadays, designing the deep convolution neural network (DCNN) models suitable for micro-Doppler (m-D) signatures based human activity classification is rapidly growing. However, their high computation cost and a huge no. of parameters limit their direct/effective employability for field applications. This paper introduces a m-D signatures dataset "DIAT-μRadHAR" covering army crawling, boxing, jumping while holding a gun, army jogging, army marching, and stone-pelting/grenade-throwing, generated using a X-band CW radar. This paper also introduces a lightweight DCNN model, "DIAT-RadHARNet," designed for those human suspicious activity classification. To reduce the computation cost and to improve the generalization ability, the DIAT-RadHARNet is designed with four design principles: depth-wise separable convolutions, channel weighting based on the importance, different size filters in the depth-wise part, and operating different size kernels on the same input tensor. The network has a 213,793 no. of parameters with a total of 55 layers. Our extensive experimental analysis demonstrates that the DIAT-RadHARNet model efficiently classifies the activities with 99.22% accuracy, giving minimal false positive and false negative outcomes. The time complexity of the proposed DCNN model observed during the testing phase is 0.35 s. The same accuracy and time complexity are obtained even at adverse weather conditions, low-lighting environments, and long-range operations.**

***Index Terms*—X-band CW radar, micro-Doppler signatures, human suspicious activity, convolutional neural network, DCNN based classification**

I. Introduction

HUMAN activity recognition (HAR) has drawn significant attention in numerous fields such as security monitoring, counterterrorism, border surveillance, biomedical-patient health monitoring, and public violent protests' /attacks' early detection, etc., [1], [2]. The need of intelligent systems having the capability to detect and classify suspicious human activities becomes essential for triggering the necessary counter-action mechanisms, to control the situation, and/or for post situation/scenario analysis [3]. Nowadays, several HAR systems are available based on vision sensing (smart, closed circuit television (CCTV), surveillance video, infrared, thermal, and acoustic sensors) and RF sensing (radars). More common vision-based approaches for HAR has relied on visual data, and hence, have several limitations such as sensitivity to weather conditions, short-range operations [4], narrow field of view (for good accuracy), range focus/collimation tuning issues, accuracy issues of finder-scope/red-dot-finder based initial boresight alignment, and dependency on surrounding lighting conditions, etc., [5].

On the other hand, HAR using a radar system offers unique advantages, like, robustness for even adverse weather and/or low-lighting conditions, through-wall capability, foliage penetration/target-imaging, tunnel intruders detection, and long-range operations, etc., [6], [7]. A more common human suspicious activities from the security point of view are (i) person fight punching (boxing) during the one-to-one attack, person intruding for pre-attack surveillance (army marching), person training (army jogging), person shooting (or escaping) with a rifle (jumping with holding a gun), stone/hand-grenade throwing for damage/blasting (stone-pelting/grenades-throwing), and person hidden translation for attack execution or escape (army crawling) [8], [9]. Detecting/classifying these kinds of suspicious human activities correctly, better at the earliest, by processing their micro-Doppler (m-D) signatures is very much essential [10].

In the recent decades, researchers have used machine learning (ML) techniques that require a separate time-frequency (T-F) map features retrieve/classification algorithms [11]; like principal component analysis (PCA), multilayer perceptron, linear predictive coding (LPC), and support vector machine (SVM), etc., for HAR [12], [13]. However, the efficiency of extracted features using such traditional techniques for classification is limited by prior information and the intricacy of categorization problems [14], [15]. Also, the classification performance of this kind of approach degrades if the m-D signature similarity between the classes are high and/or the number of classes are more [16]. Therefore, nowadays, deep learning (DL) gains the researchers' attention since it does concurrent automatic features extraction, with improved self-actuation, self-regulation, and self-control due to its intrinsic computation and decision-making skills [17], [18], which motivated us to use the DL technique for the purpose of human suspicious activity classification. The main contribution reported in this paper are as follows:

1) Indigenous development of a X-band continuous wave (CW) 10 GHz radar and its uses to prepare the m-D

This research is supported by the Life Sciences Research Board (LSRB) in association with Defence Institute of Psychological Research (DIPR) sanction letter no. LSRB/o1/15001/M/LSRB-381/PEE& BS/2020, dated 15.03.2021.

M.Chakraborty, S.V Dhavale* is with the Department of Computer Science and Engineering, Defence Institute of Advanced Technology (DIAT), Girinagar, Pune 411025, INDIA (e-mail: mainak.mail@gmail.com, sunitadhavale@gmail.com).

H.C.Kumawat, AA Bazil Raj*, is with the Department of Electronics Engineering, Defence Institute of Advanced Technology (DIAT), Girinagar, Pune 411025, INDIA (e-mail: harish.kumawat1@gmail.com, brazilraj.a@diat.ac.in). * corresponding authors

0018-9456 (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 March 05,2022 at 05:54:34 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/TIM.2022.3154832, IEEE Transactions on Instrumentation and Measurement

signatures dataset, named as "DIAT-μRadHAR", corresponding to suspicious human activities: army marching, jumping with holding a gun, army jogging, army crawling, grenades-throwing/stone-pelting, and boxing.
2) Designing a lightweight deep convolutional neural network (DCNN), named as "DIAT-RadHARNet", using a diversified dataset ("DIAT-μRadHAR"), validating its classification/performance accuracy, and comparing the computational complexities of the designed DCNN model with the state-of-the-art CNN models.
3) Following four design principles: depth-wise separable convolutions, channel weighting based on the importance, different size filters in the depth-wise part, and operating different size kernels on the same input tensor, to reduce the computation cost and to improve the generalization ability of "DIAT-RadHARNet".
4) Computing the ablation study and classification report containing the statistics of layers, parameters, floating point operations (FLOPs), accuracy, precision, recall, F1 score, receiver operating characteristics (ROC), and categorical cross-entropy loss, etc., in order to have the thorough performance/classification accuracy analysis of proposed DCNN model.
5) Test results analysis of designed DCNN model's classification accuracy by conducting open-field experiments during different environmental conditions: normal/adverse weather conditions, low-lighting environments, and long-range operations.

The rest of this paper is organized as follows: a brief review summary of related recent works on HAR are provided in Section II, Section III describes the developed RF sensor and dataset formation methodology, Section IV elucidate the developed "DIAT-RadHARNet" architecture and its implementation details, Section V presents the experimental results with performance analysis metrics, and finally, Section VI gives the conclusion.

II. Related Recent Works

Recently, radar-based systems for recognizing the different human activities (hand gesture [19], walking, running, jumping, crawling, cycling, sitting-still/standing-still [20], patient walk (with/without cane/walker), walking with a concealed rifle, and walking while carrying bag/suitcase/laptop [21], etc.,) are becoming very popular. In [8], Zenaldin et al. reported the human activity classification based on a continuous wave (CW) radar (6.5 GHz; target operation ranges are 0.5-9 m) using a SVM classifier for running, walking, and crawling classes. In [9], Chen and Ye developed a 1D CNN using a non-diversified dataset prepared using a commercial K band CW radar ($\approx$ 24 GHz; target ranges are 0.5-5 m inside the laboratory) and used it for the classification of seven different types of human activities. In [13], Javier and Kim have introduced a non-diversified m-D signatures dataset, of size 1008, obtained using a S-band CW radar (2.4 GHz; targets ranges are 2-8 m inside the laboratory). The LPC is used to extract the features, followed by a SVM-based categorization. In [16], Zhang and Cao suggested a DCNN model trained using m-D samples, of size 1896, prepared using a w-band radar (77 GHz; target range is $\approx$17.7 m) for the recognition of walking, swinging hands, and falling human activities.

In [22], Z Gu et al. designed a 5.3 GHz one transmitter and two receivers based CW radar system to classify the hand gestures in the presence of human body movements. In [23], B Li et al. proposed a sign language/hand gesture classification system using an ultrawideband (UWB) radar module. In [18], Shrestha et al. prepared a dataset using a frequency modulated indoor CW radar (5.8 GHz; with 15 human targets) and classified six different human activities using a long short-term memory (LSTM) network. The model was trained and validated with two different 35-45 seconds duration sequences, respectively. In [15], Alnujaim et al. presented a six-layer DCNN architecture designed using a 2D convolutional and fully linked layers with fixed-size filters for HAR. This kind of design increases computational cost and is incapable of capturing multi-level features. Further, the developed model is trained using a non-diversified dataset of size 144, generated using a 2.4 GHz CW radar and 12 human targets. In [24], Lee et al. presented the human activity classification based on the incremental learning of m-D signatures. The authors used a S-band CW radar (2.4 GHz; 12 human targets) and prepared a dataset (of size 1008) containing spectrogram images to classify seven different classes.

Thus, to the best of our knowledge, numerous works are being conducted to classify general human activities, but only fewer works on suspicious human activities. Further, from the literature, it is noted that the majority of the works have some flaws as (i) physical parameters (i.e., orientation) of the human target are not taken into account when acquiring the samples, (ii) samples are collected in a laboratory environment and with short-range targets operations, (iii) smaller size datasets are used to train the models, (iv) datasets are not available to the access of academicians/researchers, (v) multi-level features are not considered, (vi) computationally costly, and (vii) not suitable for on-device implementation, etc. The proceeding sections address these flaws to a greatest extend. In addition, we wish to find the answer for the following research questions through our extensive DCNN designs and validation experimental results/analysis:

1) does the use of different size filters in depth-wise separable convolutional layer raise the computational cost?
2) does weighting the channels depending on the feature map's importance help improving the model's generalization capabilities?
3) is multi-scale filtering on the same input tensor useful in collecting information about human motion signals from a spectrogram image of suspected human activity?
4) does combining channel weighting with multi-scale feature maps increase network performance?

III. Radar Experimental Setup and Dataset Formulation

This section describes the radar design and dataset preparation.

0018-9456 (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 March 05,2022 at 05:54:34 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/TIM.2022.3154832, IEEE Transactions on Instrumentation and Measurement

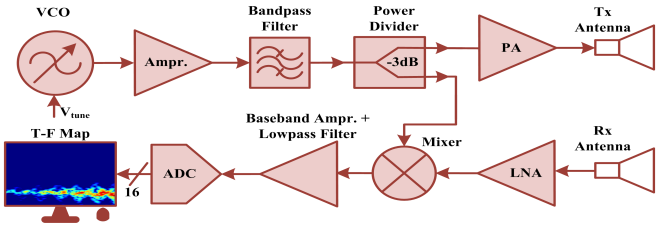


Fig. 1: Top-level schematic of the developed X-band CW Radar.



Fig. 2: Photograph of an indigenously designed CW (X-band) radar in use during field tests.

TABLE I: Main Specifications of developed X-band CW Radar

S.No.	RF Sub-system	Specification
1	VCO	RF Frequency: X-Band(10 GHz) V_{tune}: 10 V Power: 0 dBm
2	Amplifier	Bandwidth: 6 -18 GHz Gain: 24 dB
3	Bandpass Filter	Bandwidth: 9.75 - 11.25 GHz
4	Power Divider	Bandwidth: 1.8 - 12.6 GHz Isolation: 11.7-19.9 dB
5	Power Amplifier	Bandwidth: 8-12 GHz Gain: 50 dB
6	Mixer	Bandwidth: 3.7 -10.5 GHz
7	LNA	Bandwidth: 0.5 -12 GHz Gain: 17 dB
8	Tx/Rx Antenna	Bandwidth: 8 - 12 GHz Gain: 9 dBi Beamwidth: 15^o
9	Baseband Amplifier	Bandwidth: DC-3 kHz Gain: 50 dB
10	Lowpass filter	Bandwidth: DC-2 kHz
11	ADC	Sampling Frequency: 10 kHz

A. Radar Design and Specifications

An X-band CW radar, as shown in Fig. 1, operating at 10 GHz, is developed at our radar system design laboratory and is used to collect the experimental m-D signatures via different open-field experiments (refer Fig.2). The developed radar consists of an X-band voltage controlled oscillator (VCO), source amplifier, bandpass filter, power divider, power amplifier (PA), and Tx horn antenna at the transmitter chain and Rx horn antenna, low noise amplifier (LNA), and a mixer at the receiver chain. The main specifications of these subsystems are given in Table I. The radar signal acquisition module is designed with a baseband amplifier, low-pass filter, and analog to digital converter (ADC) [25]. This section receives the signal from the radar Rx chain and sends it, in digital form at the sampling rate of 10 ksps, to the m-D signature extraction algorithm developed in PC (MATLAB environment).

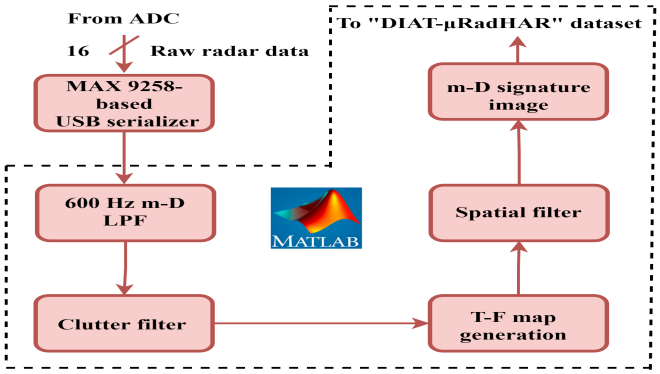


Fig. 3: Pre-processing pipeline on radar raw data to generate m-D signature images.

B. DIAT-μRadHAR Dataset Preparation

To have a diversified dataset, thirty different human targets, including males and females of different weights and heights, are directed to perform the suspicious activities in the open experimentation field, at different ranges between 10 m and 0.5 km, in front of the radar. To generate the DIAT-μRadHAR dataset as a realistic one, all the suspicious activities are performed at seven different orientations (tilted positions): 0^o, $\pm 15^o$, $\pm 30^o$ and $\pm 45^o$, in radar run-time. The radar raw data, i.e., the output of ADC (16 bits), is serialized by a MAX9258 based USB serializer and passed into the MATLAB environment. Radar time-domain signals, acquired during the normal/adverse weather conditions, low-lighting environments, and long-range operations, are recorded for the radar-run time of 3 s. Radar data acquired for 3 s is passed through a 600 Hz m-D software low pass filter (LPF), as shown in Fig. 3 which removes the high-frequency components and allows only the m-D spectrum for which our radar is designed, i.e., the maximum radial velocity is 9 m/s [26]. Since in coherent CW radar, the background clutters; bias the radar signal with a DC voltage [27], the clutter (stationary) effect in the input signal is filtered using a "dsp.DCBlocker" (first stage in clutter filter) MATLAB object. During the experiments, the radar-cross section (RCS) (i.e., reflections to the X-band radar's CW signal) of grasses, tree leaves, wind fluctuations, or similar slow/random moving targets is very low; hence, they introduced a very weak (very low power) also very low Doppler components (clutter power is concentrated about the zero frequency band). Thus, the Doppler effects of such clutters are removed by an appropriate design of a digital transversal filter (second stage in clutter filter), as discussed in [27], giving a high degree of attenuation around 0 Hz (i.e., deep stop-band around DC), and a high gain throughout the rest of the considered Doppler spectrum; to get the improved signal-to-clutter ratio (SCR). The clutter effect filtered radar signal is applied into short-time Fourier transform (STFT), given in Eq.(1), for generating the T-F map.

$$X[m,k] = \sum_{n=-\infty}^{+\infty} x[n]w[m-n]e^{-i2\pi nk/N} \quad (1)$$

Where, x[n] is the discrete signal, k is the index value for frequency, w[n] is the window function, n is the sample number, and N is the number of samples in analyzing window.

0018-9456 (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 March 05,2022 at 05:54:34 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/TIM.2022.3154832, IEEE Transactions on Instrumentation and Measurement

In our work, the window length is 2048 (Hamming window), 4096 points discrete Fourier transform (DFT), 2000 points (97.65%) overlapping, and signal length is 40000. Noises, sourced from slight movements of the background clutter and radar subsystems, present in the T-F map are further removed using a spatial filter (third stage, as shown in Fig.3) designed as described in [28]. The noise cancelled and background clutter effects filtered spectrogram samples are stored in the respective class for designing our novel DCNN architecture. The class-wise data collection details can be found in Table II. Fig. 4 shows the event photographs and their sample spectrogram images for each suspicious activity. The total number of spectrogram samples in our dataset is 3780. Out of the collected samples, 80% and 20% are used for training and validation purposes, respectively. The spectrogram images clearly exhibit how the features of different human activities, due to torso, hand, leg, and body movements, are correlated (i.e., interclass correlations) [29].

In order to facilitate the worldwide researchers working in this field, we decided to publish our complete dataset (as .mat file) and proposed DCNN network (source code) at our institute website (https://www.diat.ac.in/view-profile/?id=98) for free access for their academic/research purposes.

IV. A Proposed Novel DCNN Architecture

This section explains the design of a proposed novel DIAT-RadHARNet architecture along with its design considerations and its implementation details.

A. Proposed DCNN Architecture

The DIAT-RadHARNet architecture, shown in Fig. 5, is designed from the inspiration of MobileNetV2 [30], InceptionV3 [31], SENet [32], and BlazeFace [33] architectures with considering the following five important design principles:

1) the depth of the feature map in a convolution operation denotes the number of convolutional kernels used to learn the various feature representations of the input tensor. Some of the feature maps are more significant than others. Thus, the simplest way to improve a DCNN's generalization capabilities is to weight the channels based on the importance of the feature map [32].
2) depth-wise separable convolutions reduce parameters of the model and extra computing overhead substantially [30].
3) most of the existing DCNN architectures utilized 3×3 filters in the depth-wise separable convolutional layers. The computational costs on such convolution operation is dominated by the point-wise convolution. For example; if the input tensor size is $D_s \times D_s \times M$ and the filter size is $D_k \times D_k$, then, the total number of multiplication and addition operation for the depth-wise part is ${D_s}^2 M {D_k}^2$ while the subsequent point-wise convolution on N channels require a total of ${D_s}^2 MN$ multiplication and addition operation; which is $\frac{N}{{D_k}^2}$ times than the depth-wise convolution. Thus utilizing different size $(1\times1, 3\times3, 5\times5)$ filters in the depth-wise part does not increases extra computational overhead [33].
4) the occupied area of m-D signatures of suspicious human activities on the spectrogram images differ based on the type of activity, radar aspect angle, i.e., activities orientation with respect to radar boresight, physical parameters, and the range at which the activities take place, etc. Thus, finding the optimal filter size to learn those signatures is difficult. A larger filter produces a global distribution, while a smaller one produces a local distribution. On the other hand, increasing the depth is prone to over-fitting. The solution to this problem is to operate different size kernels on the same input tensor [31].
5) the introduction of Global Average Pooling (GAP) layers in any DCNN model, instated of the traditional fully connected layer, before the final classifier reduces the extra burden of trainable parameters [34].

The proposed DIAT-RadHARNet DCNN model comprises 55 layers as: 13 separable convolutional layers, 1 max-pooling layer, 3 multiply layers, 6 add layers, 13 batch normalization layers, 6 ReLU activation layers, 3 dense layers with ReLU activation, 3 dense layers with sigmoid activation, 1 dropout layer, 4 GAP layer, 1 softmax, and 1 input layer. To reduce the computational overhead, we used separable convolutional layers throughout our network instead of 2D DCNN. All the convolutional layers are activated by ReLU activation, and we use the same padding to maintain the same output feature-maps dimension as the input tensor. We included batch normalization layers in our model to improve the generalization error and accelerate the training process. The proposed DIAT-RadHARNet DCNN model can be expressed as in Eq. (2).

$$\text{DIAT-RadHARNet(x)} = \beta(f^{FC}(f^{D}(f^{G}(f^{MSFCA_i}(f^{M}(f^{B}(\sigma(f^{C_i}(x * w_i + b_i))))))))) \tag{2}$$

Where β, f^{FC}, f^{D}, f^{G}, f^{MSFCA_i}, f^{M}, f^{B}, σ, f^{C_i}, x, w_i and b_i represents softmax activation, classification layer, dropout layer, global average pooling layer, multi-scale filtering and channel attention (MSFCA) layers, max-pooling layer, batch normalization layer, ReLU activation, first convolutional layer, input image, weight matrix and bias, respectively. The proposed DCNN model takes the preprocessed spectrogram images as an input in the first phase. During preprocessing phase, images are first downsized to $256 \times 256 \times 3$, normalized using a 1/255 scaling ratio, and then augmented with zooming, brightening, and horizontal flipping [10], on-the-fly, using the "ImageDataGenerator" class of Keras library. We utilized separable convolutional layers to reduce the computational overhead and to make the model lightweight. On the other hand, introducing batch normalization layers after convolutional layers in our model helps to reduce the generalization error and to speed up the training process.

The processed spectrogram samples are then passed through the first convolutional layer of the proposed DCNN model, where we apply 96 filters with stride two of size (3×3) to learn the basic feature representations. The output of this layer of dimension $(128 \times 128 \times 96)$ is then given to the

0018-9456 (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 March 05,2022 at 05:54:34 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/TIM.2022.3154832, IEEE Transactions on Instrumentation and Measurement

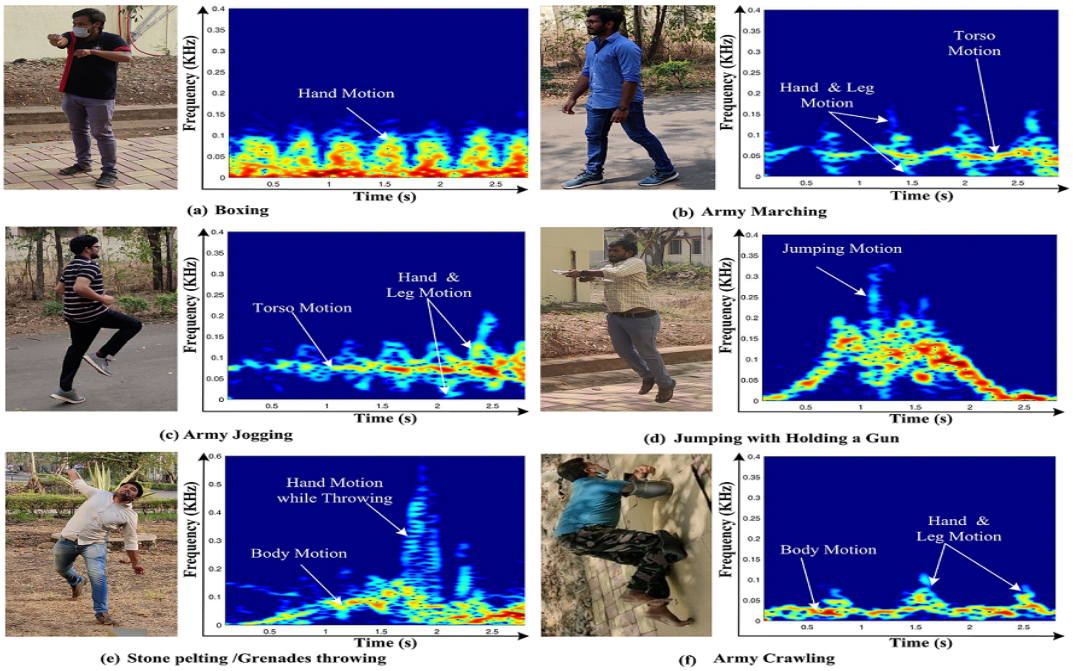


Fig. 4: Micro-Doppler spectrogram samples of each suspicious human activity.

TABLE II: Human targets class wise parameters. The total number of samples is 3780.

Class Name	Human Target's Physical Parameters			Different Angle of Targets Orientation	Normal weather condition	Low-light environments	Adverse weather condition	Long-range operations	Total No. of Samples
	Height (cm)	Weight (kg)	Age (years)						
Boxing	165-190	51-90	24-35	0°, ± 15°, ± 30°, ± 45°	104	145	230	150	629
Army crawling	170-190	65-87	26-35	0°, ± 15°, ± 30°, ± 45°	150	150	149	152	601
Army jogging	160-189	58-85	24-40	0°, ± 15°, ± 30°, ± 45°	157	139	170	149	615
Army marching	147-183	55-93	24-52	0°, ± 15°, ± 30°, ± 45°	187	145	152	208	692
Stone pelting/ Grenades throwing	155- 190	51-90	25-35	0°, ± 15°, ± 30°, ± 45°	153	169	172	148	642
Jumping with holding a gun	170-185	65-90	25-38	0°, ± 15°, ± 30°, ± 45°	168	145	144	144	601

batch normalization layer, followed by the first max-pooling layer. The max-pooling layer with stride two reduces the input volume's spatial dimension to $(64 \times 64 \times 96)$ for subsequent layers. Then, we add three consecutive MSFCA blocks to its end as shown in Fig. 5. The MSFCA blocks are designed with the proposed design principles said in points 1 to 4 above. The MSFCA block takes the output of the previous max-pooling layer or activation layer. This block helps the model to learn the multilevel feature representations from the same input tensor and improve the channel interdependency. As a result, the overall performance of the network increased with minimal computational cost. Each MSFCA block consists of four parallel paths. Path one, two, and three are made up with one convolutional layer followed by one batch normalization layer which can be expressed by Eqs. (3), (4), and (5), respectively.

$$f^{P_1}(x) = f^{B}(\sigma(f^{C_i}(f^{x} * w_i + b_i))) \quad (3)$$

$$f^{P_2}(x) = f^{B}(\sigma(f^{C_i}(f^{x} * w_i + b_i))) \quad (4)$$

$$f^{P_3}(x) = f^{B}(\sigma(f^{C_i}(f^{x} * w_i + b_i))) \quad (5)$$

where f^{C_i}, σ, f^{x}, w_i, b_i, i, and f^{B} represents convolutional layer, ReLU activation, output feature map of the previous max-pooling layer or activation layer, weights, bias, layer index, and batch normalization layer, respectively. To capture multilevel feature representation from the input tensor, we applied different size filters in paths one, two, and three; but the stride was one for all convolution operations as per the proposed design principles in points 2, 3 and 4, above. The filter size for paths one, two, and three was 5×5, 3×3, and 1×1, respectively. The output tensors of the three paths are then element-wise added in the addition layer using Eq. (6).

$$f^{A_1}(x) = f^{P_1}(x) \oplus f^{P_2}(x) \oplus f^{P_3}(x) \quad (6)$$

where $\oplus$ is the element-wise addition operation. The multilevel feature representation map $f^{A_1}(x)$, generated by the element-wise addition operation is then passed through a batch normalization layer f^{B} followed by a ReLU activation layer, which is then element-wise added with the output of path four $f^{P_4}(x)$, as given in Eq. (7).

$$f^{A_2}(x) = \sigma(f^{B}(f^{A_1}(x))) \oplus f^{P_4}(x) \quad (7)$$

where σ is the ReLU activation function. Path four, as given in Eq. (8), helps to model channel interdependencies with only a small increase in the computational cost; but including this path in the MSFCA block directly improves the overall network performance.

$$f^{P_4}(x) = \sigma(f^{C_i}(f^{Mul}(\alpha(f^{FC}(\sigma(f^{FC}(f^{G}(f^{x})))))) * f^{x})) \quad (8)$$

where σ, f^{C_i}, f^{Mul}, α, f^{FC}, f^{G}, and f^{x} denotes ReLU activation, convolutional layer, multiplication layer, sigmoid activation, dense layer, global average pooling layer, and

0018-9456 (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 March 05,2022 at 05:54:34 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/TIM.2022.3154832, IEEE Transactions on Instrumentation and Measurement

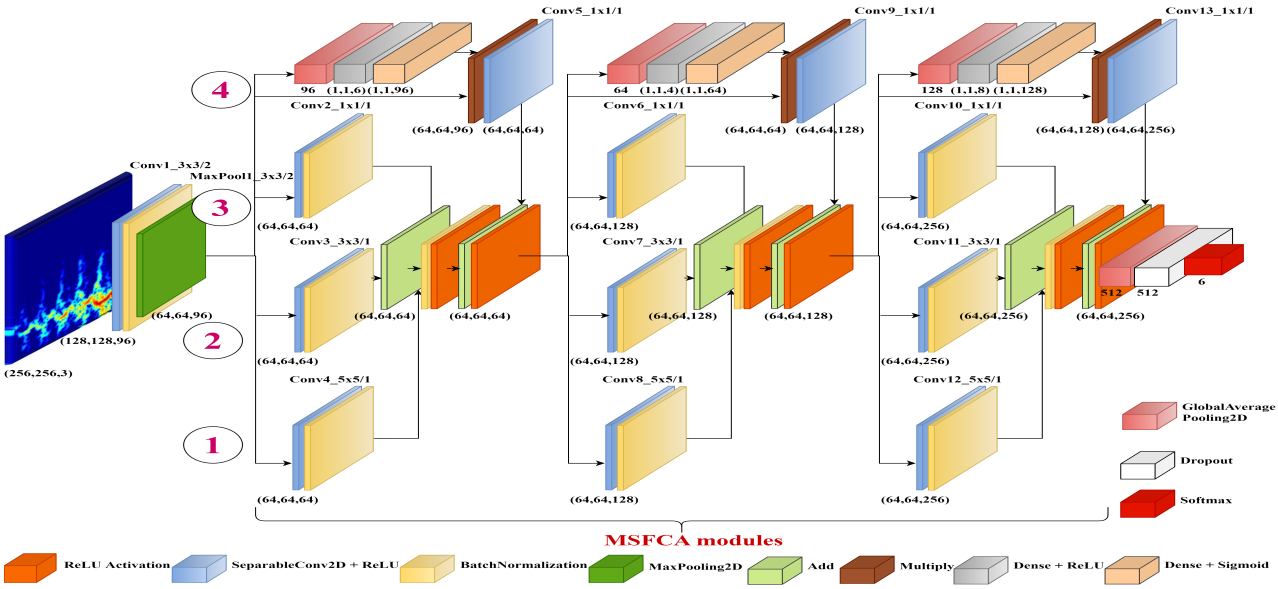


Fig. 5: Overall framework of the proposed DCNN architecture.

output of the previous max-pooling layer or activation layer, respectively. Path four takes an input feature map x of dimension $D_s \times D_s \times M$, reduces it to dimension $1 \times 1 \times M$ by GAP layer, and then passes this M-length tensor through a multi-layer perceptron, which projected back to the same dimension $1 \times 1 \times M$ as the dimension of the input feature map, which is then element-wise multiplied with the input feature map x. The first dense layer in this path reduces the complexity of its output channel by a factor of sixteen. The advantage of including this path in the MSFCA block is said above in proposed design principal 1. The resultant tensor is then subsequently passed into a 1×1 convolutional layer to maintain the same dimension as $f^{A_1}(x)$. Finally, by combining Eq. (3) to Eq. (8), we can express the MSFCA block as Eq. (9).

$$f^{MSFCA_i} = \sigma(f^{A_2}(x)) \tag{9}$$

where f^{A_2} is the second addition layer of that block, and x is the output feature map of the previous layer, which can be represented by $f^x \in \mathbb{R}^D$. The D is the output dimension of the previous layer. After that, we fed the output tensor of f^{MSFCA_3} block to a GAP layer (for the reason said in design principle 5). Then, the output feature map of the GAP layer is fed into the dropout layer to prevent overfitting. Finally, a dense layer of six neurons triggered by the softmax activation function is added to the proposed DCNN network to anticipate the final action label. The proposed network has 213,793 parameters. The developed new DCNN model is trained end-to-end using 3780 m-D images of six classes of suspicious human activities. Further, the proposed model employs depth-wise separable convolutional layers and a global average pooling layer to reduce computation costs (lightweight DCNN), which makes the proposed architecture more suitable for on-device (for e.g., field-programmable gate array (FPGA)) implementation.

B. Implementation details

The proposed novel DIAT-RadHARNet architectural model is implemented on a Windows 10 workstation with an Intel Xeon Gold 6140 2 Processor (2.30 GHz, 2.29 GHz), 256 GB memory, NVIDIA Quadro RTX 6000(24.0 GB) Graphics card, CuDNN v7.6.0, and CUDA v10.0.130 Tool kit. Our proposed architecture is implemented using Keras 2.2.4, python 3.5.6, matplotlib 2.2.2, and TensorFlow-GPU v1.14.0. In our work, we employed a trial-and-error method to tweak the model hyper-parameters. The DIAT-RadHARNet network is end-to-end trained on the captured m-D signature dataset, and the Adam optimizer is used to optimize the network. We employed zooming by 0.1 scale, brightening between 0.9 to 1.1, and horizontal flipping data augmentation technique to increase the size of a training dataset. The learning rate of the model is initially set to 0.001 and reduced by 0.5 factor up to 0.00001 when validation accuracy saturated in subsequent two epochs. The batch size is set to be eight, and the maximum epoch is 300. The weights and bias in all convolutional layers are initialized with glorot uniform and a constant value of 0.2, respectively. On the other hand, the normal weight initializer is used in the dense layers of the MSFCA block. The number of filters applied in each layer and its output dimension is shown in Fig. 5. We employed the same padding in all convolution operations to maintain the same dimension as the input tensor. To make the network more generalized and to prevent overfitting, 40% dropout is applied before the softmax layer. We also monitor the validation loss; if no progress is seen in the next ten epochs, we use early stopping to avoid over-fitting. The categorical cross-entropy loss function, as shown in Eq. (10), is used to measure the performance of the DIAT-RadHARNet network.

$$L = -\sum_{k=1}^{c} t_k \log(p_k) \tag{10}$$

where p_k denotes the softmax probability for the k^{th} class, c is the total number of class, and t_k denotes the truth label.

0018-9456 (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 March 05,2022 at 05:54:34 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/TIM.2022.3154832, IEEE Transactions on Instrumentation and Measurement

V. Experimental results and Performance Analysis

To illustrate the efficacy of the proposed DCNN model, ROC curves are created for each of six distinct activity classes for a given set of predicted probabilities and the ground truth labels. We have used a technique known as "one vs all" to obtain the ROC curve. In this technique, we have used "roc_curve" metrics from "sklearn package" which takes (i) true binary labels, (ii) target scores, and (iii) positive class, and returns (i) false positive rate, and (ii) true positive rate. In this way, the ROC curve, for every class, is obtained against all other classes. The resulting ROC plots are depicted in Fig. 10; from which, we can observe that the area under the curve (AUC) value is one for all classes, which proves the model's classification excellency. As we know, a high AUC score indicates that the model is good at differentiating the six human suspicious activity classes.

A detailed/statistical performance study is conducted to evaluate the class-wise performance and estimate the classification error of our DCNN model. The class-wise F1-score, recall, and precision of our DCNN model are presented in the classification report given in Table (III), which evidences for the 99.22% of overall classification accuracy of the proposed DCNN model. The loss and accuracy plot shown in Fig. 11 depicts the training process and the learning direction of our DCNN model. The training and validation curves in the accuracy plot converge, indicating that the network is not overfitted.

In order to perform a thorough test analysis on the classification accuracy of our proposed DCNN model, all the suspicious human activities shown in Fig. 4 are performed in front of the radar, with a new set of human targets, at different ranges between 10 m and 0.5 km during normal/adverse weather conditions, low-light environments and long-range operations, and the respective (on-the-spot) test samples are collected. We used these on-the-spot test samples: normal (380 samples)/adverse (314 samples) weather conditions, low-light (211 samples) environments, and long-range (140 samples) target operations, to get the thorough performance analysis. The classification statistics obtained during these open-field different-environmental experiments are as follows:

1) Normal weather conditions: Fig. 6 gives the classification statistics of on-the-spot test results of the proposed DCNN model, obtained from the test experiments conducted during normal weather conditions. It can be observed from Fig. 6 that the model achieves 99.22% test classification accuracy, and the miss classification rate is only 0.78%. It also reveals that only three samples belonging to boxing, jumping with holding a gun, and stone pelting are miss classified as army crawling.

2) Adverse weather conditions: The confusion matrix associated to the test experiments conducted during adverse weather conditions is shown in Fig. 7. The overall on-the-spot test accuracy of the model in adverse weather condition is 99.05% and miss classification rate is 0.95%. We can notice that the model's false negative and positive score for each class is very low.

3) Low-light environments: The performance statistics of the model in the low-light conditions are given in the confusion matrix shown in Fig. 8. The model achieves 99.06% classification accuracy in these experimental conditions with only 0.94% of miss classification. It can also be observed that the model's false positive and negative results are minimal.

4) Long-range target operations: The proposed DCNN model achieves 99.29% of test classification accuracy in long-range target operations as reported in the confusion matrix shown in Fig. 9. The miss classification rate of our model in long-range operations is only 0.71%. We can notice that only

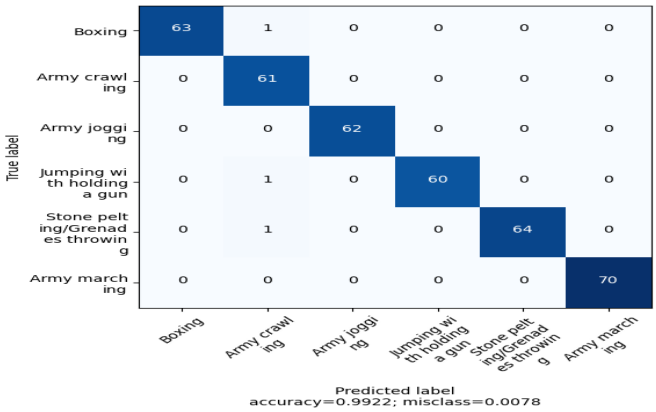


Fig. 6: DIAT-RadHARNet: Confusion matrix-normal weather conditions.

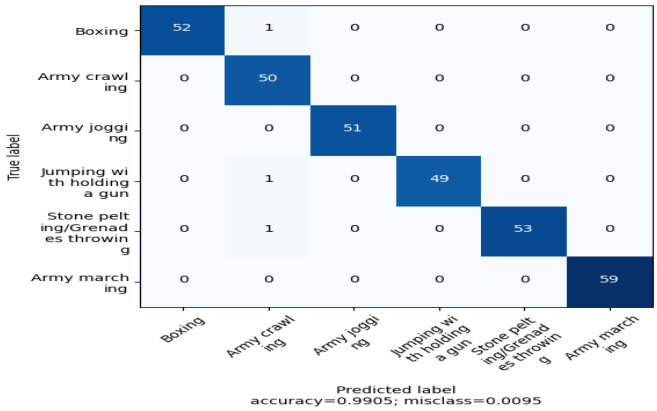


Fig. 7: DIAT-RadHARNet: Confusion matrix-adverse weather conditions.

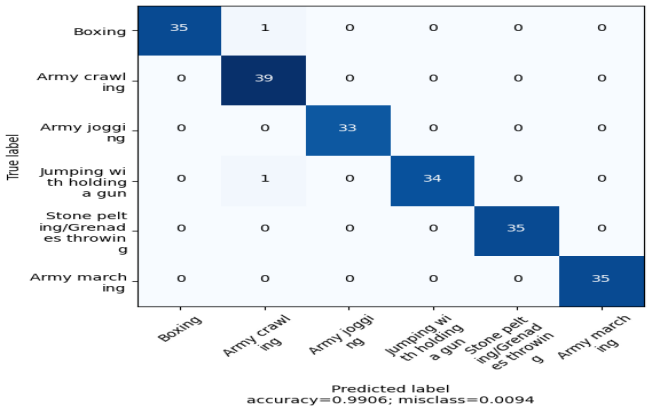


Fig. 8: DIAT-RadHARNet: Confusion matrix-low light conditions.

0018-9456 (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 March 05,2022 at 05:54:34 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/TIM.2022.3154832, IEEE Transactions on Instrumentation and Measurement

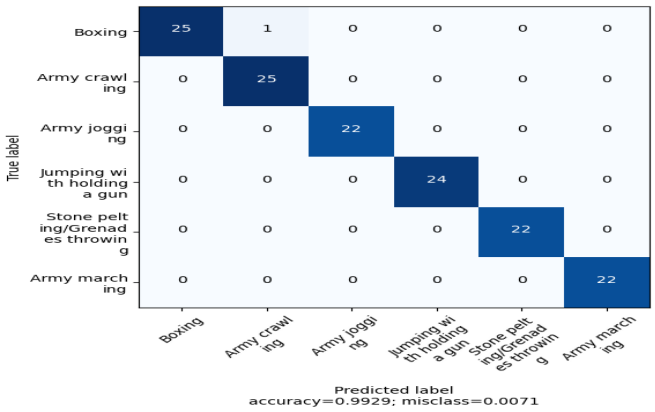


Fig. 9: DIAT-RadHARNet: Confusion matrix-long range operations.

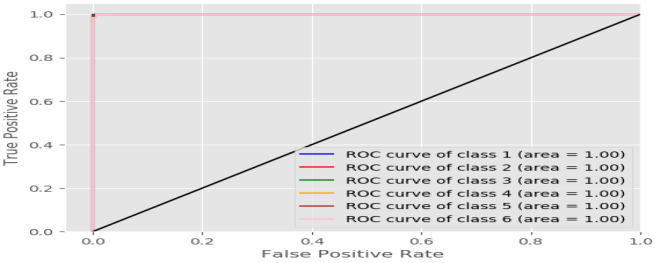
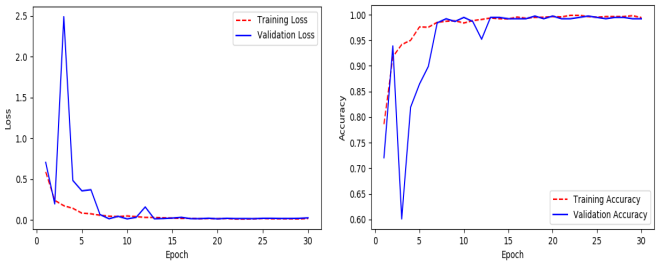


Fig. 10: ROC curve of the proposed DIAT-RadHARNet model.

TABLE III: Classification report of the proposed DIAT-RadHARNet model.

Action categories	Precision	Recall	F1-Score
Boxing	1.00	0.98	0.99
Army crawling	0.95	1.00	0.98
Army jogging	1.00	1.00	1.00
Jumping with holding a gun	1.00	0.98	0.99
Stone pelting / Grenades throwing	1.00	0.98	0.99
Army marching	1.00	1.00	1.00
Accuracy			99.22
Macro avg	0.99	0.99	0.99
Weighted avg	0.99	0.99	0.99



(a) (b)

Fig. 11: Categorical cross-entropy loss: (a) and accuracy (b) plots; of the proposed DIAT-RadHARNet model.

one sample belonging to the boxing category is miss classified as army crawling.

Thus, these open-field different environmental test experimental results evidence the good class-wise classification performance/accuracy of our DCNN model on the newly, i.e., on-the-spot, collected test datasets/samples. It also reveals that the model's false negative and positive score for each class is very low. Further, the time complexity observed during the testing phase is 0.35 s, which is quite low, and it evidences the suitability of the designed DCNN model for on-the-fly classification applications.

Based on the overall performance evaluations, the DIAT-RadHARNet model achieved a superior accuracy (99.22%) with 213,793 parameters and 55 layers. Table (IV) gives the comparison of the DIAT-RadHARNet with that of other well-known models, trained, validated, and tested with our datasets. It is noticed from the Table (IV) that the (layers), storing capacity (size), and computational complexity (parameters and FLOPs) are at the optimum trade-off giving the classification accuracy at 99.22%. Even though the input dimension is high, i.e., [256x256x3] in our DCNN model, it gives a greater accuracy with the size of 2.75Mb, parameters of 0.21M, and FLOPs of 0.83G. Though the layers of VGG-16 and VGG-19 are relatively low, all other design parameters are very high, and the classification accuracy is low, i.e., 97% and 98%, respectively. Similarly, FLOPs for 1.0 MobileNet-224, MobileNetV2, and BlazeFace are low as 0.57G, 0.30G, and 0.30G, respectively, but the respective classification accuracy are also low, i.e., 93%, 94.67%, and 96.74%, respectively.

Further, through the extensive designs and performance validation experiments, we found the answer to all our research questions raised in Section II. The answers (findings) are as follows:

1) the optimum number of MSFCA blocks for our DCNN model is determined using an ablation study, as given in Table (V). Therein, we can notice that three consecutive MSFCA blocks are sufficient to have the test accuracy of 99.22%. Further, it is observed that increasing the number of MSFCA blocks doesn't lead to increasing the accuracy, but it increases the computational complexity. Hence, we finalized only three MSFCA blocks in our proposed architecture. Adjusting the filter size in the depth-wise separable convolutional layer resulted in no extra processing burden (as shown in Table (V), the total FLOPs are only 0.62G when three MSFCA blocks are used without channel weighting (CW)).
2) weighting the feature maps adds a minimal trainable parameter to the model, but the benefit is that it improves the model's generalization capabilities (as shown in Table (V), the total FLOPs is only 0.62G when using three MSFCA blocks without channels weighting (CW) and 0.83G when using three MSFCA blocks with CW, implying that the difference between these two cases is only 0.21G.).
3) as shown in Table (V), when three MSFCA blocks are used without CW, the overall test accuracy of the model is 98.13%, indicating that the benefit of multi-level feature representation extraction plays a vital role

0018-9456 (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 March 05,2022 at 05:54:34 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/TIM.2022.3154832, IEEE Transactions on Instrumentation and Measurement

TABLE IV: Comparison of the DIAT-RadHARNet model with other popular DCNN models.

DCNNs	Input Dimension	Layers	Size(Mb)	#Params(M)	FLOPs	Accuracy
VGG-16 [35]	$224 \times 224 \times 3$	23	528	138.3	15.5G	97%
VGG-19 [35]	$224 \times 224 \times 3$	26	549	143.6	19.6G	98%
1.0 MobileNet-224 [36]	$224 \times 224 \times 3$	92	16	4.2	0.57G	93%
MobileNetV2 [30]	$224 \times 224 \times 3$	156	14	3.5	0.30G	94.67%
DenseNet-201 [37]	$224 \times 224 \times 3$	709	80	20.2	4.32G	94.34%
Inception-v3 [31]	$299 \times 299 \times 3$	313	92	23.8	5.73G	84%
ResNet-50 [38]	$224 \times 224 \times 3$	177	98	25.6	3.88G	76%
BlazeFace [33]	$128 \times 128 \times 3$	73	3	0.6	0.30G	96.74%
DIAT-RadHARNet	$256 \times 256 \times 3$	55	2.75	0.21	0.83G	99.22%

TABLE V: Ablation study of the proposed DIAT-RadHARNet model. FLOPs denote floating point operations.

No of MSFCA blocks with/without channels weighting (CW)	Layers	#Params(M)	FLOPs	Accuracy
1 MSFCA with CW	23	0.03	0.12G	96.87%
2 MSFCA with CW	39	0.07	0.27G	97.91%
3 MSFCA with CW	55	0.21	0.83G	99.22%
4 MSFCA with CW	71	0.49	1.96G	97.91%
5 MSFCA with CW	87	1.05	4.16G	98.96%
6 MSFCA with CW	103	2.16	8.55G	98.96%
3 MSFCA without CW	31	0.10	0.62G	98.13%

in extracting valuable characteristics from spectrogram images.

4) combination of feature map weighting based on priority and multi-scale feature map extraction from the m-D signatures of different human targets improves the network's generalization ability as well as overall accuracy (as shown in Table (V), when three MSFCA blocks are used with CW, thus, the overall test accuracy of the model is 99.22% at the FLOPs of 0.83G).

VI. Conclusion

This paper introduced a dataset "DIAT-μRadHAR" containing the m-D signatures corresponding to six different human suspicious activities and designed a new lightweight DCNN architecture "DIAT-RadHARNet" for their detection and classification. The new DCNN model is trained/validated with the m-D signature samples recorded using a X-band (10 GHz) CW radar's open-field; range-varying (10m-0.5km), different weather/lighting conditions, diversified human targets, and different aspect angles, experiments. In this work, we increased the generalization capability of the designed DCNN model by weighting the feature maps and adding negligible trainable parameters. We also found that changing the filter size in the depth-wise separable convolutional layer did not require additional computational overhead.

Designed DCNN model gives 99.22% classification accuracy with a very-low false negative and positive outcomes with 213,793 No. of parameters and a total no. of 55 layers. The time complexity observed during the testing phase of the designed DCNN model is 0.35s which is quite low; hence, it is more suitable for on-device, i.e., FPGA implementation, which is one of our ongoing research works. However, the work reported in this article doesn't address the classification of suspicious activities of multiple subjects, inverse synthetic aperture radar (ISAR) imaging, classification of suspicious human activities based on the ISAR signatures, cross-range/adaptive-beam-forming effect analysis on m-D signatures of suspicious human activities and imaging the target's activities on a range-velocity map to demonstrate the separability of subject targets and the slow-moving clutters, etc., are our near-future research works.

Acknowledgement

This research is supported by the Life Sciences Research Board (LSRB) in association with Defence Institute of Psychological Research (DIPR) sanction letter no. LSRB/o1/15001/M/LSRB-381/PEE& BS/2020, dated 15.03.2021. The authors want to thank Dr Suresh Kumar, DIPR for his support and NVIDIA for the academic GPU research grant. The authors wish to thank all the students who supported us during the data collection.

References

[1] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE communications surveys & tutorials*, vol. 15, no. 3, pp. 1192–1209, 2012.

[2] F. Fioranelli, S. A. Shah, H. Li, A. Shrestha, S. Yang, and J. Le Kernec, "Radar sensing for healthcare," *Electronics Letters*, vol. 55, no. 19, pp. 1022–1024, 2019.

[3] W. Huang, L. Zhang, W. Gao, F. Min, and J. He, "Shallow convolutional neural networks for human activity recognition using wearable sensors," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[4] L. Ren, L. Kong, F. Foroughian, H. Wang, P. Theilmann, and A. E. Fathy, "Comparison study of noncontact vital signs detection using a doppler stepped-frequency continuous-wave radar and camera-based imaging photoplethysmography," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 9, pp. 3519–3529, 2017.

[5] M. Chakraborty, A. Pramanick, and S. V. Dhavale, "Two-stream mid-level fusion network for human activity detection," in *International Conference on Innovative Computing and Communications*. Springer, 2021, pp. 331–343.

[6] H. C. Kumawat and A. B. Raj, "Extraction of doppler signature of micro-to-macro rotations/motions using continuous wave radar-assisted measurement system," *IET Science, Measurement & Technology*, vol. 14, no. 7, pp. 772–785, 2020.

[7] R. M. Narayanan and M. Zenaldin, "Radar micro-doppler signatures of various human activities," *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1205–1215, 2015.

[8] M. Zenaldin and R. M. Narayanan, "Radar micro-doppler based human activity classification for indoor and outdoor environments," in *Radar Sensor Technology XX*, vol. 9829. International Society for Optics and Photonics, 2016, p. 98291B.

0018-9456 (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 March 05,2022 at 05:54:34 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/TIM.2022.3154832, IEEE Transactions on Instrumentation and Measurement

[9] H. Chen and W. Ye, "Classification of human activity based on radar signal using 1-d convolutional neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 7, pp. 1178–1182, 2019.

[10] H. C. Kumawat, M. Chakraborty, A. A. B. Raj, and S. V. Dhavale, "Diat-μsat: Small aerial targets' micro-doppler signatures and their classification using cnn," *IEEE Geoscience and Remote Sensing Letters*, 2021.

[11] W. Gao, L. Zhang, W. Huang, F. Min, J. He, and A. Song, "Deep neural networks for sensor-based human activity recognition using selective kernel convolution," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.

[12] W. Li, B. Xiong, and G. Kuang, "Target classification and recognition based on micro-doppler radar signatures," in *2017 Progress in Electromagnetics Research Symposium-Fall (PIERS-FALL)*. IEEE, 2017, pp. 1679–1684.

[13] R. J. Javier and Y. Kim, "Application of linear predictive coding for human activity classification based on micro-doppler signatures," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 10, pp. 1831–1834, 2014.

[14] Z. Chen, C. Jiang, S. Xiang, J. Ding, M. Wu, and X. Li, "Smartphone sensor-based human activity recognition using feature fusion and maximum full a posteriori," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 7, pp. 3992–4001, 2019.

[15] I. Alnujaim, D. Oh, and Y. Kim, "Generative adversarial networks for classification of micro-doppler signatures of human activity," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 3, pp. 396–400, 2019.

[16] R. Zhang and S. Cao, "Real-time human motion behavior detection via cnn using mmwave radar," *IEEE Sensors Letters*, vol. 3, no. 2, pp. 1–4, 2018.

[17] S. K. Leem, F. Khan, and S. H. Cho, "Detecting mid-air gestures for digit writing with radio sensors and a cnn," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 4, pp. 1066–1081, 2019.

[18] A. Shrestha, H. Li, J. Le Kernec, and F. Fioranelli, "Continuous human activity classification from fmcw radar with bi-lstm networks," *IEEE Sensors Journal*, vol. 20, no. 22, pp. 13 607–13 619, 2020.

[19] J. Höysniemi, P. Hämäläinen, L. Turkki, and T. Rouvi, "Children's intuitive gestures in vision-based action games," *Communications of the ACM*, vol. 48, no. 1, pp. 44–50, 2005.

[20] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 3, pp. 257–267, 2001.

[21] X. Geng, G. Li, Y. Ye, Y. Tu, and H. Dai, "Abnormal behavior detection for early warning of terrorist attack," in *Australasian Joint Conference on Artificial Intelligence*. Springer, 2006, pp. 1002–1009.

[22] Z. Gu, J. Wang, F. Shen, K. Xu, D. Ye, J. Huangfu, C. Li, and L. Ran, "Blind separation of doppler human gesture signals based on continuous-wave radar sensors," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 7, pp. 2659–2661, 2019.

[23] B. Li, J. Yang, Y. Yang, C. Li, and Y. Zhang, "Sign language/gesture recognition based on cumulative distribution density features using uwb radar," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.

[24] D. Lee, H. Park, T. Moon, and Y. Kim, "Continual learning of micro-doppler signature-based human activity classification," *IEEE Geoscience and Remote Sensing Letters*, 2021.

[25] A. A. B. Raj, *FPGA-based embedded system developer's guide*. CRC Press, 2018.

[26] S. Kumar and A. B. Raj, "Design of x-band fmcw radar using digital doppler processor," in *2021 International Conference on System, Computation, Automation and Networking (ICSCAN)*. IEEE, 2021, pp. 1–5.

[27] G. L. Charvat, *Small and short-range radar systems*. CRC Press, 2014.

[28] I. Pitas, *Digital image processing algorithms and applications*. John Wiley & Sons, 2000.

[29] A. S. Akella and A. B. Raj, "Dcnn based activity classification of ornithopter using radar micro-doppler images," in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE, 2021, pp. 780–784.

[30] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[31] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[32] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[33] V. Bazarevsky, Y. Kartynnik, A. Vakunov, K. Raveendran, and M. Grundmann, "Blazeface: Sub-millisecond neural face detection on mobile gpus," *arXiv preprint arXiv:1907.05047*, 2019.

[34] M. Chakraborty, S. V. Dhavale, and J. Ingole, "Corona-nidaan: lightweight deep convolutional neural network for chest x-ray based covid-19 infection detection," *Applied Intelligence*, vol. 51, no. 5, pp. 3026–3043, 2021.

[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[36] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[37] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[38] 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*, 2016, pp. 770–778.



Mainak Chakraborty (Graduate Student Member, IEEE) presently pursuing his Ph.D. from the Defence Institute of Advanced Technology, Pune, INDIA. He is working as an Assistant Professor in the Department of C.S.E at Dream Institute of Technology, Kolkata, INDIA, since 2011. He received his M.Tech and B.Tech degrees in Computer Science & Engineering and Information Technology from the Maulana Abul Kalam Azad University of Technology, West Bengal, INDIA; in 2011 and 2008, respectively. He is a recipient of AICTE-INAE Teachers Research Fellowship from the Indian National Academy of Engineering. His research areas are Deep Learning, Machine Learning, Human Activity Detection, and Medical Image Analysis.



Harish C Kumawat (Graduate Student Member, IEEE) received his B.Tech. degree in Electronics and Communication Engineering from Rajasthan Technical University, Kota, India, and his M.Tech. degree in Microwave and Optical Communication from Delhi Technological University, Delhi, India. He worked as Assistant Professor in Government Engineering College Bikaner, Rajasthan from Aug. 2014 to Jan. 2018. Currently, he is working in the field of Radar Signal Processing in Defence Institute of Advanced Technology, Pune, India.



Sunita Vikrant Dhavale (Member, IEEE) presently associated with the Defence Institute of Advanced Technology (DIAT), Pune, as an Assistant Professor in the Computer Science & Engineering department. She received her M.E. and Ph.D. degrees in Computer Science from Pune University in 2009 and DIAT in 2015. She is a recipient of Atal Tinkering Lab Mentorship, ISACA Academic Advocate Membership, EC-Council's CEH-v8, IETE M. N. Saha Memorial Award, and Outstanding Woman Achiever Award from Venus International Foundation. Her research areas are computer vision, deep learning, and cybersecurity. She is a member of many professional bodies, including IEEE, AAAI, ACM, ISTE, IETE, IAENG, and ISACA.



Arockia Bazil Raj A (Member, IEEE) received his B.E degree in Electronics and Communication Engineering from the Bharathidasan University, Tiruchirappalli, India, his M.E degree in Communication Systems, and his Ph.D. degree in Information and Communication Technology from the Anna University, Chennai, India. Based on his teaching and research expertise, he wrote three books published by De Gruyter, GmbH, CRC Press, USA, and IET, UK. He was an Assistant Professor in Kings College of Engineering, Thanjavur, India, from 2002 to 2006. Presently, he is working as an Associate Professor in Defence Institute of Advanced Technology (DIAT), Pune, India. His current research interests cover radar system design, radar signal processing, FSO communications, digital signal processing, and RF photonics.

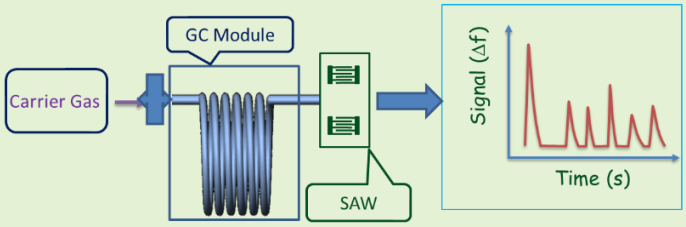
0018-9456 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 05,2022 at 05:54:34 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.2022.3162072, IEEE Sensors Journal

Effect of Carrier gas on sensitivity of Surface Acoustic Wave Detector

Jitender Kumar, A.T.Nimal, Upendra Mittal, Vinod Kumar and Vinay Kumar Singh

Abstract— **Enhancement of sensitivity is a hot pursuit for every kind of sensor/detector. While many methods are exploited for various sensors/detectors for this, the methods are specific in nature for the technology involved. This article discusses a novel study where enhancement of sensitivity, specific to a Surface Acoustic Wave (SAW) device used as a Gas Chromatographic (GC) detector is achieved by using a proper carrier gas. With an ultimate aim of implementing the outcome on SAW-GC based, handheld, Chemical Warfare Agent (CWA) Electronic-Nose vapour detector being developed at our laboratory, four different carrier gases i.e. Hydrogen, Helium, Nitrogen and Zero air were subjected to the scrutiny for its impact on sensitivity. Detection of a CWA simulant, 2-Chloro-Ethyl-Ethyl-Sulfide (CEES), 1,5-Dichloropentane (DCP) simulating Sulphur Mustard (H Agent), Di-methylmethyl phosphonate (DMMP), Diethyl cyanophosphonate (DECP), Triethyl pohosphate (TEP) simulating G-Agents along with three Volatile Organic Compounds (VOCs) i.e. Methanol, Toluene and Xylene were carried out in the study. A fast-GC with 70 seconds chromatogram, with appropriate GC parameters is used for the study in order to cater for the handheld battery operated rapid detector. A spread of more than two times in sensitivity could be seen while using common GC carrier gases in the range of concentration (200-2000ppm) studied. By virtue of GC with SAW detector, excellent selectivity, repeatability, and long term stability is achieved. A detailed discussion on the reasons for the same is presented.**



Index Terms*— *Carrier gas, Chemical Warfare Agents, Gas Chromatography, Surface Acoustic Wave

I. Introduction

In the detection of toxic vapours of Chemical Warfare Agents (CWA), the Lethal Dose LD50 (which when applied would kill 50% of the population) is of the order of few parts per million (ppm) to parts per billion (ppb) for numerous chemicals, which demands very low Limit of Detection (LOD).

During the development of such detectors, every means to enhance sensitivity a little further count. While additions such as pre-concentrator traps could increase the sensitivity of many folds, the basic detection mechanism needs to be highly sensitive. Optimization of various parameters and/or a little tuning can be useful in the ultimate pursuit of developing a sensitive trace level vapour detector.

The detection of toxic vapours in various environmental conditions, with high sensitivity equipment, is an essential demand from armed forces and home land security agencies [1], [2]. The commercially available trace level detectors for toxic vapours uses Ion Mobility Spectroscopy (IMS) [3], Flame Photometry[4], Mass Spectroscopy[5], Infrared Spectroscopy[6], Chemiresistive nanomaterial, [7]–[11] and Surface Acoustic Wave (SAW) [12] technologies. Many other detectors such as are reported to be in the advanced stages of research[10]. Since the present work is on Surface Acoustic Wave (SAW) detector using Gas Chromatography, further discussion is restricted to GC based sensors.

GC is one of the established technologies used for the separation and detection of volatile organic compounds (VOCs) by using various detectors and columns. In a capillary GC column, the mixture of vapours entering GC is retained to different extents according to their affinity with the internal coating (stationary phase) of the column, and the vapours exit at the various time called ‘retention time’. The selectivity is obtained by separation of vapours over time. The conventional GC system typically employs a selective capillary column of 10-100 m length resulting in higher resolution and enhanced

Jitender kumar is with Solid State Physics Laboratory (SSPL), Lucknow Road, Timarpur, New Delhi, India -110054 and with Department of Chemistry Delhi University Delhi, India (email: Jitender.tilwaliya@gmail.com)
A.T.Nimal is with SSPL, email: atnimal@yahoo.com
Upendra Mittal is with SSPL, email: upendra_mittal@yahoo.com
Vinod Kumar Special Centre for Nano Sciences, Jawaharlal Nehru University, Delhi-110067, India email: kumarv@jnu.ac.in
Vinay kumar Singh Chemistry department, Sri Aurobindo College, University of Delhi, Delhi-110017, India (Corresponding author)
The authors are grateful to Director SSPL, New Delhi, India for her constant support and encouragement for doing research work in the field of detection of CW agents.

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 March 29,2022 at 09:41:30 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.2022.3162072, IEEE Sensors Journal

selectivity with extended analysis time, extending from several minutes to a few hours. For field applications, to cut short the operation and analysis time, researchers have developed ultrahigh-speed chromatography by using short column lengths. Such systems are used in various fields of scientific, medical, and industrial applications for real-time detection of toxic VOCs [1], [13]–[20]. The GC based sensor has many advantages such as high reproducibility, less sensor training, good shelf life, enhanced selectivity, a low false alarm in the presence of interferents compared to non-GC techniques such as IMS, chemeresistisve and optical. The disadvantage such as bulkiness and longer response time is now being overcome by the use of miniature fast GC. Various GC based CWA detectors reported in the literature and their parameters pertinent to the present topic are listed in table 1[3]–[6], [11], [12], [18], [21], [22].

Table 1. GC Based CWA Detector

S.No	Product name	Detector	Carrier gas	Sensitivity	Response/Specification
1.	Accu Sense	TCD and IMS	DHGC technology	ppb	Takes 15 min for analysis with ppb range (Seer Technology, Inc.)
2.	zNose	SAW	Helium	ppb	Shows good response (<1 min) for the detection of ppb range (Sensor Technology)
3.	Canary-3	SAW	Ambient air	ppb	Based on MEMS GC module and having good response time (<1 min) towards the detection of CWA (Defiant Technologies)
4.	Chem-ID	TCD	Helium	ppb	Detect the CW agents within 15 min at ppb range (D-tect Systems)
5.	Explorer	PID/ECD	Carrier	ppm	Have dual detector configuration and detect the compounds within 15 min (INFICON)
6.	Guardion 7	TMS	Helium	ppm	MS based detector and takes the analysis time 15<min (Torion Technologies, Inc.)
7.	Micro GC Fusion Gas Analyzer	MEMS TCD	-	ppm	Takes <5 min for the detection of compounds in ppm range (INFICON)
8.	MINICAMS 3500/2010	FPD/ PFPDFPD	Helium	ng	Requires fuel gas along with carrier gas, show good response <30 sec. (OI Analytical)
9.	Sea PORT Mini GC	MEMS array	Air carrier	ppm	Operate with Air, takes <15 min for analysis (Seacoast Science, Inc.)
10.	TOCAM	PID	Ambient	ppb	Takes <15 min for analysing target with good sensitivity (Defiant Technologies)
11.	Mini-GC	Chemiresis-tive	Air	ppb	Fast response but slow recovery.

The GC based system needs carrier gas for the operation to carry vapours along the column. In literature, carrier gases such as nitrogen, helium, hydrogen, air, argon, carbon dioxide, ammonia, and even steam have been reported for GC applications[1], [17]–[20], [23]–[29].

Helium gas is extensively utilized as a carrier gas, because of its inert nature, high purity, and established methodologies [30]. However, its abundance on earth and hence its availability and high cost are the drawbacks. The physical properties of some of these carrier gases are mentioned in table 2 [31], [32].

Table 2. Physical properties of Carrier at Room temperature

Carrier gas	Thermal Conductivity (W/(m·K)	Density (g/L)
Hydrogen	0.1805	0.08988
Helium	0.1513	0.1786
Zero air	0.02588	1.29
Nitrogen	0.02583	1.2504

A promising GC detector for CWA is a solid state SAW device. Piezoelectric substrates are used in the fabrication of SAW devices by using the photolithography process [33]. SAW device comprises of comb-like structures called an inter-digital transducer (IDT). On applying an electrical signal to the input IDT, the SAW (mechanical wave) is formed and output IDT receives these waves and converts them into electrical signals.

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 March 29,2022 at 09:41:30 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.2022.3162072, IEEE Sensors Journal

The design of IDTs determines the type (such as delay lines, filters, resonators, etc.), frequency, and phase of operation of SAW device. Since the acoustic energy is confined to the surface any perturbation on or near the surface affects the SAW properties, and this is used to realize sensors. SAW devices are usually used in oscillator circuit configuration for sensor applications [33]–[35]. In a SAW –GC detector, the interaction of the chemical agents with the fine mechanical waves causes a reversible shift in frequency, which is the detector output.

In GC based technologies, it is a well-known fact that, as per Van Deemter curves, there are optimum carrier gases for separation efficiency owing to molecular weight, viscosity, and diffusivity of gases [36], [37]. However, there are no reports of its effect on peak heights (i.e., sensitivity).

The present study to investigate the impact of carrier gas on the sensitivity seems to be specific to GC-SAW based instruments, for reasons that would be discussed at the end of this report. To the best of our knowledge and available literature, the effect of carrier gas on the sensitivity of the SAW based detector has not been studied earlier.

II. Experimental Setup

It consists of SAW detector with thermoelectric cooler (TEC), SAW oscillator, frequency readout circuitry, carrier gases and details are as follows:

A. Fabrication of SAW-GC Detector

The 433.92 MHz SAW one-port SAW resonators (RO3101) obtained from RF monolithic are used in this study. It is fabricated on quartz substrate having aluminum Inter-digited transducer (IDT) of periodicity 7 µm. The typical insertion loss of the device was -1.5 dB and central frequency of 433.92 MHz. The SAW device is mounted onto the TEC to maintain the temperature of the device at 25ºC or lower. The uncoated SAW device mounted on TEC is the detector. The capillary GC column (used to separate and fix the retention time of target gas) is attached in the closest possible vicinity of the SAW detector (SAW Device) surface. The GC capillary column does not make any physical contact with the detector surface. The schematic of the arrangement is shown in Fig.1.

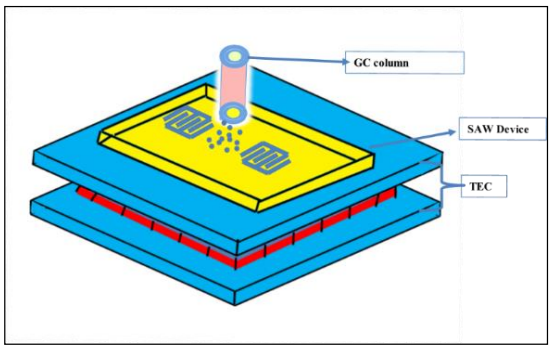


Fig. 1. Schematic design of SAW Detector with TEC

As the vapour elutes from the column it hits the detector surface causing the detector frequency to shift due to its interference with SAW on the surface. Since the SAW detector frequency may vary with temperature, a TEC is employed to control the temperature. The S_{11} and S_{21} parameters of the device were characterized by using Vector Network Analyzer (VNA, Anritsu, model no. MS4642A) shown in Fig.2. Prior to measurements, standard calibration was carried out on the VNA.

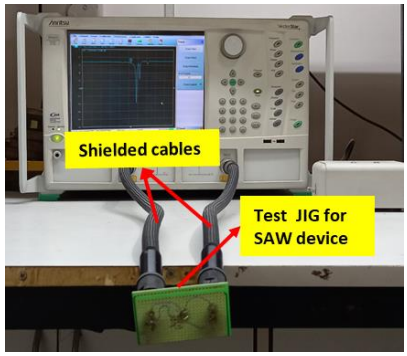


Fig.2a. Measurement setup (Vector Network Analyser)

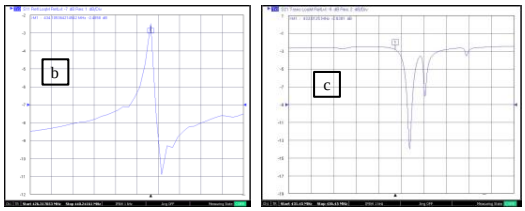


Fig.2. Characteristics of one-port SAW resonator (b) S_{11} (c) S_{21}

B. Electronic Circuit for SAW Detector

Two ST cut quartz-based SAW resonator devices of 433.92 MHz were used in oscillator circuit configuration, one as a detector and the other as a reference. The reference device is used to mix the two frequencies and bring down to an intermediate frequency of 50 to 200 kHz. The outputs of both the devices were mixed through a low-pass filter (LPF) to produce an intermediate frequency Fig.3. The detector is placed at the end of the GC column and the column is aligned to the center of the SAW detector. The electronic circuit with the SAW device is presented in Fig.4

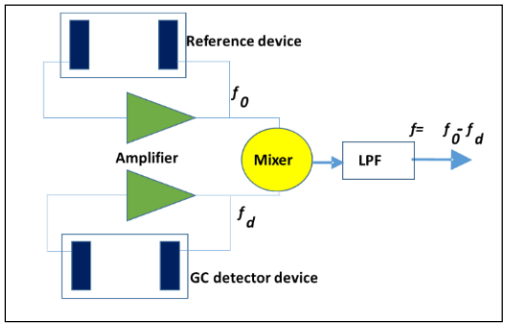


Fig. 3. SAW oscillator Configuration for SAW detector

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 March 29,2022 at 09:41:30 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.2022.3162072, IEEE Sensors Journal

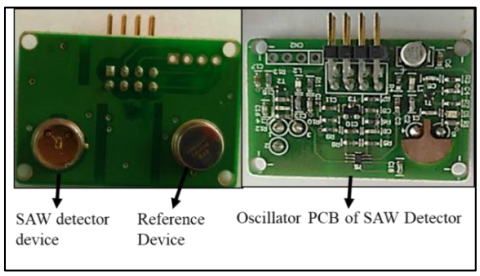


Fig. 4. SAW detector with electronics

III. Material and Method

For the present study, analytical grade Methanol, Toluene, p-Xylene, 2-CEES, DCP, DMMP, DECP, and TEP (Sigma Aldrich), Microliter syringe (Hamilton), DB-200 GC capillary column (Agilent), small TEC (Size - 12.3×12.3 mm, Laird Technologies, part No. 177-1753) and Chemito-1000 Gas Chromatograph instrument (Thermo Fisher) were used for performing the experimental work. A known concentration mixture of all the simulants and volatile organic compounds was prepared by mixing in equal quantities (2000ppm v/V) in acetone. A syringe was used to injected 0.1 microliters of the mixture of sample from vial to injector port for analysis. The parameters of GC i.e. oven temperature (80°C), injector port (120°C), SAW detector (25°C), flow (3 ml/min) and 3-meter GC capillary column of 0.32mm I.D. kept the same for the entire period of the experiment. The simplified diagram of the test setup is shown in Fig. 5 and test set-up is shown in Fig. 6.

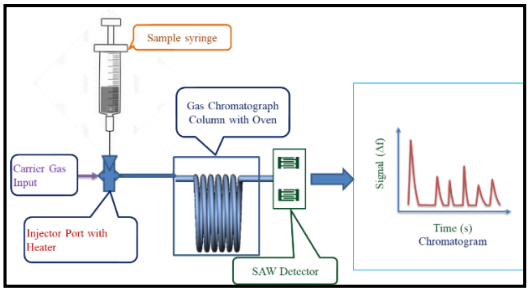


Fig. 5. Schematic of Test setup

The flow of each carrier gas was controlled at 3 ml/min and confirmed by a bubble meter at the end of the GC capillary column. The frequency shift of the SAW detector is recorded using the data acquisition system.

IV. Results and Discussion

The uncoated SAW detector gives a fast response, shows high sensitivity, and gives a linear response to the given analyte on the surface of the SAW detector. It also shows good baseline stability and fast recovery. The SAW detector was tested with a standard mixture of the sample using a different carrier gas

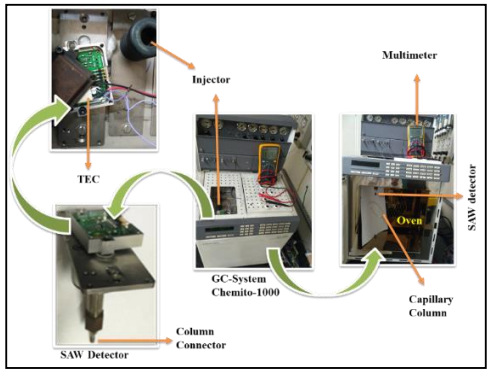


Fig. 6. Test setup for sensing

(Hydrogen, Helium, Nitrogen and Zero air) shown in Fig.7. Since fast GC is used the peaks are wide, though the widths are in seconds. The mixture of all eight compounds is separated and identified by their distinctive retention times (RT) by using a small GC capillary column. A small variation in the retention time of chromatogram was observed while using different carrier gases (Hydrogen, Helium, Nitrogen, and Zero air) due to their different physical properties, i.e., viscosity, diffusivity etc.

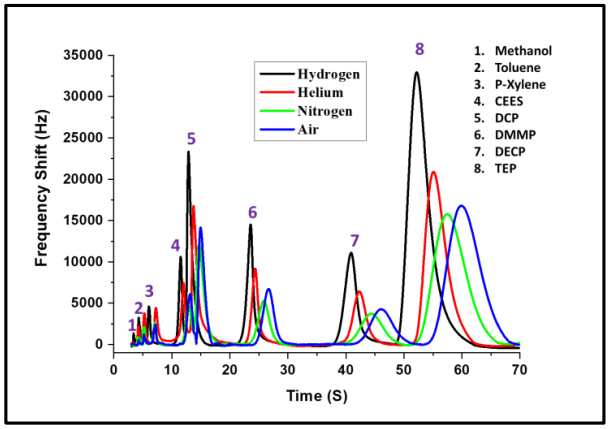


Fig. 7. Response of GC-SAW detector with mixture of 8 organic compound using different carrier gases.

It was observed that the detector gives more sensitivity when hydrogen is used as carrier gas. It was also observed that the detector gives almost similar sensitivity towards agents by using Zero air and Nitrogen as carrier gas. Similarly, the detector was also tested with different three concentrations i.e. 1x (~200ppm), 2x (~400ppm) and 3x (~600ppm) of the mixture of four simulants DCP, DECP, DMMP, and TEP using four different carrier gases is shown in Fig 8. It was found that the SAW detector response, in terms of sensitivity/frequency shift, is maximum when hydrogen is used as carrier gas.

Note that the sensitivity of the GC-SAW sensor is measured in terms of the shift in frequency per ppm concentration of analyte gas (Hz/ppm). However, in most practical cases sensitivity is used in synonymous to frequency shift [38].

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 March 29,2022 at 09:41:30 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.2022.3162072, IEEE Sensors Journal

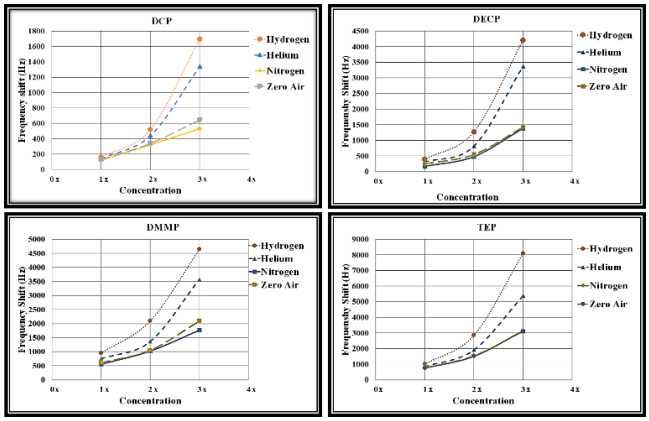


Fig. 8. Response of SAW detector with different simulants and concentration using Hydrogen, Helium, Nitrogen and Zero air as carrier Gas

Table 3 gives the detector sensitivities for various analytes, concentrations with different carrier gases. Each experiment has been repeated at least six times to ensure repeatability and the values given are the average sensitivities.

Table 3. Sensitivity of SAW in Hz/ppm against analyte using different gas H_2, He, N_2 and Air as carriers.

Analyte	Conc.	Carrier gas			
		H_2	He	N_2	Air
DCP	1x	2.5	1.5	1.0	1.2
	2x	2.8	1.8	1.1	1.1
	3x	4.3	2.5	1.3	1.5
DMMP	1x	5.0	4.0	2.8	3.0
	2x	5.5	3.5	2.6	2.7
	3x	8.0	6.0	3.5	3.5
DECP	1x	2.0	1.5	1.0	1.2
	2x	3.1	2.0	1.2	1.3
	3x	7.0	5.5	2.5	2.4
TEP	1x	2.5	2.5	2.5	2.5
	2x	2.5	2.5	2.5	2.5
	3x	2.5	2.5	2.5	2.5

The response curves were obtained by repeating over many days The detector shows good response, repeatability, and low noise (<±10Hz). Also in our experience with such GC SAW sensing, found that the sensitivity is highly repeatable [39]. The response and recovery time of the sensor are shown in table 4.

Table 4. Response and recovery time of the sensor towards the analyte using H_2 as carriers.

Analyte	Conc.	Response time (S)	Recovery time (S)
DCP	1x, 2x & 3x	13.1 (±0.5)	14.60 (±1.0)
DMMP		23.5 (±0.5)	24.8 (±1.2)
DECP		41.0(±0.5)	43.6 (±1.5)
TEP		52.5 (±0.5)	57.5 (±2.0)

The SAW based chemical sensor works on the principle of mass loading. The reason for higher sensitivity when using hydrogen as carrier gas is analysed below. Two probable reasons are considered for analysis:

1. The mass change on the surface of the detector.
2. The temperature changes on the surface of the detector.

Let us take the first case, and analyse how mass changes affect the sensitivity.

When the analyte is eluted from the column, it creates perturbation on the SAW propagation path on the surface of the SAW detector and results in change in phase velocity and amplitude, which leads to a frequency shift of the oscillator circuit. The frequency of SAW changed due to mass loading on the surface of the detector as per the Sauerbrey equation 1 [40], [41].

$$\Delta f = -\frac{2f_0^2}{A\sqrt{\rho_q \mu_q}}\Delta m \qquad (1)$$

Where Δf is the change in frequency, f_o is the resonant frequency, Δm is the mass change, A is the active crystal area, ρ_q is the density of substrate material, and μ_q is the Shear modulus of quartz.

The change in frequency is proportional to the mass change of adsorbed molecules on the device. A theoretical model of adsorption of vapors on uncoated SAW devices is described in [29]. Detector frequency is changed due to the adsorption and desorption of analyte and carrier gas on the surface of the detector. Normally the frequency of the SAW oscillator decreases with mass loading on the surface. Since dual oscillator configuration is used (Fig 3) the difference frequency increases due to mass loading.

Fig.9 attempts to illustrate the effect of three different carrier gases on SAW sensitivity. Here f_0 is the resonant frequency when no gas is passed on the SAW surface, f_1, f_2 & f_3 are the frequencies of the SAW device after passing H_2, He, and N_2 /Zero Air carrier gases respectively. When analyte vapor along with a carrier gas (H_2, He, or N_2) passes over the SAW device its frequencies shift to f_4, f_5 & f_6. Here $\Delta f_H = (f_4\text{-}f_1)$, $\Delta f_{He} = (f_5\text{-}f_2)$ & $\Delta f_N = (f_6\text{-}f_3)$ are the difference frequency signals corresponding to H_2, He and N_2 respectively.

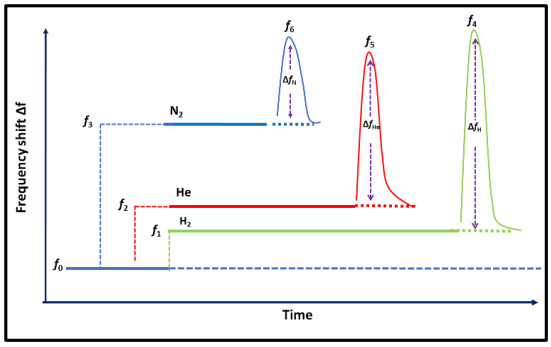


Fig. 9. Representative illustration of frequency shifts with mass loading of analyte with different carrier gases

The carrier gas molecules eluting from the GC capillary column impinges on the surface of the SAW device in a jet form at a velocity of ~62 cm/s. The new frequencies f_1, f_2, and f_3 are the baselines (or offsets) of the GC chromatogram. When the analyte molecules elute from the column along with carrier gas

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 March 29,2022 at 09:41:30 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.2022.3162072, IEEE Sensors Journal

the SAW frequency is disturbed due to a mass/density change between the carrier and the mixture (of carrier and analyte). It must be noted that the analytes are heavier than the carriers. The resultant frequency changes Δf_H, Δf_{He} & Δf_N are observed as the peak signal for the analyte. It may be suggested here that the frequency shift is more prominent in the case of low mass/density (hydrogen) carrier gas because of higher relative mass/density change. However, the mass of the analyte added being the same for different carrier gases, it may be argued that the shift should be the same irrespective of the carrier gas. In this case, the relative mass change will hold true only when the response of the SAW sensor for increasing mass is non-linear, (i.e., the shift for unit change in mass decreases with increasing mass). However, it may be noted that the SAW device responses are mostly linear except for extreme cases[42]. In a closer look, we may consider the fact that when the analyte is added to the carrier gas, part of the carrier gas is replaced by analyte vapor in the jet stream. When replacement is considered the relative change in mass is definitely higher for carrier gas with lower mass/density. Though the analyte is in parts per million and the replacement is trace level, it is understood that this is the quantity responsible for the frequency shift. Hydrogen being the lightest gas with low molecular weight and density, gives the maximum mass change (Δm) over the surface of SAW for a given concentration. From eq. (1) it is evident that this will result in a maximum frequency shift. Thus, the resultant sensitivity (frequency shift) of the SAW detector with GC is higher (H_2>He>Air>N_2) with hydrogen as carrier gas as compared to the other higher density carrier gases.

Now, let us analyse the second case of temperature change: -

The carrier gas receives heat from the hot GC capillary column depending on its thermal conductivity as it travels through it for a short time (~5s) and its temperature is raised slightly. The heat energy transferred by gas eluting from the GC column to the surface of the SAW device, will lead to change in the temperature of the SAW, and will result in frequency shift of the SAW oscillator.

$$\Delta f \propto \Delta T \qquad (2)$$

Where Δf is the difference frequency shift due to temperature variation ΔT on the surface of the detector due to heat flow.

When the carrier gas alone is flown the shift forms the offset. Now, when the analyte is introduced, a part of the carrier is replaced by the analyte and accordingly the thermal conductivity and thereby temperature is changed from that of the carrier. Here we needed to know the thermal conductivities of the analyte (k_A) used: while the thermal conductivities of Methanol, Toluene, and p-Xylene are known to be 0.202, 0.134 and 0.1313 respectively [43]–[45] the thermal conductivity of CEES, DCP, DMMP, DECP, and TEP are not available in literature. Compared with the thermal conductivities of the carrier gases (k_C) given in table 2, it can be understood that, all three conditions (i) $k_A > k_C$, (ii) $k_A < k_C$ and (iii) $k_A \approx k_C$ are applicable. Let us consider the cases one by one here. In case of $k_A > k_C$, when analyte vapor is introduced, it will take more heat from the GC capillary column and transfer more heat to the SAW device than the carrier gas. It is known that with increase in temperature the frequency of SAW oscillator decreases[46], [47]. In dual oscillator configuration, the difference frequency increases. By the same process, in case of $k_A < k_C$, when the analyte is introduced the SAW device cools slightly resulting in decrease of difference frequency. In case of $k_A \approx k_C$ the shift is negligible.

The shift in the present case seems to be a combination of mass and thermal loading. Let us take examples of methanol and toluene as analytes with H_2, He, and N_2 as carriers and look at the differential frequency shift due to loading (Δf). This is depicted in table 4.

Table 4. Differential frequency shift due to thermal loading for methanol and toluene as analytes with H_2, He and N_2 as carriers

Analyte & Carrier	Scenario		Δf	
	Thermal	Mass	Thermal	Mass
Toluene + N_2	$k_A > k_C$	$m_A > m_C$	↑	↑
Toluene + H_2	$k_A < k_C$		↓	↑
Methanol + H_2 /Toluene + He	$k_A \approx k_C$		-	↑

It may noted that with toluene as the analyte, in case N_2 as a carrier the thermal and mass loading cause Δf change in the same direction, whereas for H_2 it is on opposite directions. It may also be noted that Δf due to temperature change is much higher in the case of N_2 as a carrier because of large k_A-k_C compared H_2. Nevertheless, H_2 as carrier is found to offer higher sensitivity experimentally with all the analytes. Moreover, the ST cut quartz of the SAW device has zero first order temperature coefficient near 25°C. Thus the effect of mass loading overpowers that of thermal loading

Further studies may be carried out to find the contributions of these two factors, which may help in better understanding, leading to the design of more sensitive detectors.

V. Conclusion

This paper describes the sensitivity enhancement study of SAW-GC based chemical sensors using carrier gases with varying density and thermal conductivity. For the first time, it is reported that carrier gas plays a significant role on the sensitivity enhancement of the SAW detector, apart from the known separation and resolution capabilities. The response of the SAW detector was analyzed experimentally and it was observed that for any given concentration of analyte it increases more than 2 times, if hydrogen is used as carrier gas as compared to nitrogen. Therefore, it is proposed to use hydrogen as a carrier gas for SAW GC based chemical sensor to attain increase in the sensitivity. From the sensitivity data and considering the low noise, it can be extrapolated for the minimum detectable concentration as <10 ppm. By using pre-concentrator before the GC the sensitivity can be further increased and minimum detection level can be brought down to ppb levels. In addition, further cooling by TEC, use of high frequency SAW device, etc can also be employed to increase the sensitivity and bring down minimum detection level. However, the scope of this paper is to bring out the enhancement in sensitivity due to choice of carrier gas alone. For example, in Staple's work, if hydrogen is used instead of helium as a carrier gas, up to 1.7 times sensitivity enhancement is possible. This study will also be beneficial for the development of SAW-based hand-held gadgets for the

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 March 29,2022 at 09:41:30 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.2022.3162072, IEEE Sensors Journal

simultaneous detection of toxic gases in the environment at low concentrations. In recent years, the availability of helium has decreased and its cost has increased significantly. Therefore, this will also help to choose the alternate substitute of carrier gas for the SAW detector.

ACKNOWLEDGMENT

Authors are thankful to Director SSPL- DRDO, Dr. Seema Vinayak for her constant support. Authors are also thankful to SAW Group members- SSPL for valuable discussion.

REFERENCES

[1] L. Meciarova, S. Vilcekova, and M. Balintova, "Measurement of VOCs with a portable GC/SAW detector," *Chem. Eng. Trans.*, vol. 40, no. Special Issue, pp. 283–288, 2014, doi: 10.3303/CET1440048.

[2] H. Singh et al., "SAW mono sensor for identification of harmful vapors using PCA and ANN," Process Saf. Environ. Prot., vol. 102, pp. 577–588, 2016, doi: 10.1016/j.psep.2016.05.014.

[3] J. Puton and J. Namiesnik, "Ion Mobility Spectrometry : Current Status and Application for Chemical Warfare," *Trends Anal. Chem.*, 2016, doi: 10.1016/j.trac.2016.06.002.

[4] K. H. Mckelvie and K. B. Thurbide, "Micro - Flame Photometric Detection in Miniature Gas Chromatography on a Titanium Tile," *Chromatographia*, vol. 82, no. 6, pp. 935–942, 2019, doi: 10.1007/s10337-019-03723-y.

[5] C. A. Valdez, "Analysis of chemical warfare agents by gas chromatography-mass spectrometry : methods for their direct detection and derivatization approaches for the analysis of their degradation products," vol. 2, pp. 1–26, 2017, doi: 10.1515/revac-2017-0007.

[6] G. M. Murray, *Detection and Screening of Chemicals Related to the Chemical Weapons Convention*. 2013. DOI: 10.1002/9780470027318.a0403.pub2

[7] X. Liu, T. Ma, N. Pinna, and J. Zhang, "Two-Dimensional Nanostructured Materials for Gas Sensing," *Adv. Funct. Mater.*, vol. 27, no. 37, pp. 1–30, 2017, doi: 10.1002/adfm.201702168.

[8] N. Joshi, M. L. Braunger, F. M. Shimizu, and A. R. Jr, "Insights into nano-heterostructured materials for gas sensing : a review Insights into nano-heterostructured materials for gas sensing : a review," 2021.

[9] A. Phys, "Functional gas sensing nanomaterials : A panoramic view Functional gas sensing nanomaterials : A panoramic view," vol. 021301, no. February, 2020, doi: 10.1063/1.5123479.

[10] J. F. F. Jr, H. Hamaguchi, B. Yoon, and T. M. Swager, "Chemiresistor Devices for Chemical Warfare Agent Detection Based on Polymer Wrapped Single-Walled Carbon Nanotubes," 2017, doi: 10.3390/s17050982.

[11] J. H. Lee, H. Lee, W. Kim, and W. Lee, "Chemiresistive Gas Sensors for Detection of Chemical Warfare Agent Simulants," vol. 28, no. 3, pp. 139–145, 2019.

[12] F. Nabila *et al.*, "Sensing and Bio-Sensing Research A review of current advances in the detection of organophosphorus chemical warfare agents based biosensor approaches," *Sens. Bio-Sensing Res.*, vol. 26, no. October, p. 100305, 2019, doi: 10.1016/j.sbsr.2019.100305.

[13] S. Y. Oh, "Rapid Monitoring of Pharmacological Volatiles of Night-Flowering Evening-Primrose According to Flower Opening or Closing by Fast Gas Chromatography/Surface Acoustic Wave Sensor (Electronic zNose)," *Phytochem. Anal.*, vol. 29, no. 3, pp. 275–283, 2018, doi: 10.1002/pca.2740.

[14] N. Maouche *et al.*, "A surface acoustic wave sensor functionalized with a polypyrrole molecularly imprinted polymer for selective dopamine detection," *J. Mol. Recognit.*, vol. 28, no. 11, pp. 667–678, 2015, doi: 10.1002/jmr.2482.

[15] P. Korytar, H. G. Janssen, E. Matisová, and U. A. T. Brinkman, "Practical fast gas chromatography: Methods, instrumentation and applications," *TrAC - Trends Anal. Chem.*, vol. 21, no. 9–10, pp. 558–572, 2002, doi: 10.1016/S0165-9936(02)00811-7.

[16] G. Watson, W. Horton, and E. Staples, "Gas chromatography utilizing SAW sensors," *Proc. - IEEE Ultrason. Symp.*, pp. 305–309, 1991, doi: 10.1109/ULTSYM.1991.234175.

[17] E. J. Staples, T. Matsuda, and S. Viswanathan, "Real Time Environmental Screening of Air , Water and Soil Matrices Using a novel Field Portable GC / SAW System," in *Asia pacific Conference*, 1998, no. April, pp. 8–10, [Online]. Available: https://www.researchgate.net/publication/237493339_Real_Time_Environmental_Screening_of_Air_Water_and_Soil_Matrices_Using_a_novel_Field_Portable_GC_SAW_System.

[18] E. J. Staples and S. Viswanathan, "Ultrahigh-speed chromatography and virtual chemical sensors for detecting explosives and chemical warfare agents," *IEEE Sens. J.*, vol. 5, no. 4, pp. 622–631, 2005, doi: 10.1109/JSEN.2005.850990.

[19] X. Chen *et al.*, "Gas chromatograph-surface acoustic wave for quick real-time assessment of blood/exhaled gas ratio of propofol in humans," *Br. J. Anaesth.*, vol. 113, no. 5, pp. 807–814, 2014, doi: 10.1093/bja/aeu193.

[20] M. Akbar, M. Restaino, and M. Agah, "Chip-scale gas chromatography: From injection through detection," *Microsystems Nanoeng.*, vol. 1, no. November, pp. 1–8, 2015, doi: 10.1038/micronano.2015.39.

[21] "CBRNe Tech Index." https://www.cbrnetechindex.com/Chemical-Detection/Technology-CD/Chromatography-CD-T.

[22] G. W. Watson, E. J. Staples, and S. Viswanathan, "Real Time Chemical Process Measurements Using a Multi-port GC/SAW System," no. February, 2015, [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.498.9754&rep=rep1&type=pdf.

[23] A. Y. El Naggar, "Factors affecting selection of mobile phase in gas Chromatography," *Am. J. Res. Commun.*, vol. 1, no. 3, pp. 219–228, 2013, [Online]. Available: www.usa-journals.comwww.usa-journals.com,.

[24] A. V. Chuikin, S. V. Grigor'ev, and A. A. Velikov, "Steam as a carrier gas in capillary chromatography," *Pet. Chem.*, vol. 52, no. 3, pp. 208–212, 2012, doi: 10.1134/S0965544112030048.

[25] J. M. Sanchez and R. D. Sacks, "Performance characteristics of a new prototype for a portable GC using ambient air as carrier gas for on-site analysis," *J. Sep. Sci.*, vol. 30, no. 7, pp. 1052–1060, 2007, doi: 10.1002/jssc.200600367.

[26] R. Nasreddine, V. Person, C. A. Serra, and S. Le Calvé, "Development of a novel portable miniaturized GC for near real-time low level detection of BTEX," *Sensors Actuators, B Chem.*, vol. 224, pp. 159–169, 2016, doi: 10.1016/j.snb.2015.09.077.

[27] C. A. Cramers, H. G. Janssen, M. M. Van Deursen, and P. A. Leclercq, "High-speed gas chromatography: An overview of various concepts," *J. Chromatogr. A*, vol. 856, no. 1–2, pp. 315–329, 1999, doi: 10.1016/S0021-9673(99)00227-7.

[28] S. A. Greene and H. E. Roy, "Effect of Different Carrier Gases on Retention Times in Gas-Adsorption Chromatography," *Anal. Chem.*, vol. 29, no. 4, pp. 569–570, 1957, doi: 10.1021/ac50162a043.

[29] L. A. S. V.G Berkzkin, "Ammonia as a carrier gas and flame forming agent in Gas Chromatography," *J. Chromatogr. A*, vol. 119, pp. 33–39, 1976, doi: https://doi.org/10.1016/S0021-9673(00)86768-0.

[30] C. N. Nnaji, K. C. Williams, J. M. Bishop, and G. F. Verbeck, "Hydrogen as a GC/MS carrier and buffer gas for use in forensic laboratories," *Sci. Justice*, vol. 55, no. 3, pp. 162–167, 2015, doi: 10.1016/j.scijus.2015.01.003.

[31] "The thermal conductivity of gases." https://www.electronics-cooling.com/1998/09/the-thermal-conductivity-of-gases/.

[32] P. Evans, "Density of gases." http://www.theengineeringmindset.com.

[33] U. Mittal, T. Islam, A. T. Nimal, and M. U. Sharma, "Fabrication of high frequency surface acoustic wave (SAW) devices for real time detection of highly toxic chemical vapors," *Int. J. Smart Sens. Intell. Syst.*, vol. 8, no. 3, pp. 1601–1623, 2015, doi: 10.21307/ijssis-2017-821.

[34] Fahim, Mainuddin, U. Mittal, J. Kumar, A. T. Nimal, and M. U. Sharma, "Single chip readout electronics for SAW based gas sensor

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 March 29,2022 at 09:41:30 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.2022.3162072, IEEE Sensors Journal

systems," *Proc. IEEE Sensors*, vol. 2017-Decem, pp. 1–3, 2017, doi: 10.1109/ICSENS.2017.8233886.

[35] A. T. Nimal, M. Singh, U. Mittal, and R. D. S. Yadava, "A comparative analysis of one-port Colpitt and two-port Pierce SAW oscillators for DMMP vapor sensing," vol. 114, pp. 316–325, 2006, doi: 10.1016/j.snb.2005.05.021.

[36] J. J. van Deemter, F. J. Zuiderweg, and A. Klinkenberg, "Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography," *Chem. Eng. Sci.*, vol. 50, no. 24, pp. 3869–3882, 1995, doi: 10.1016/0009-2509(96)81813-6.

[37] Haseltine James, "Hydrogen as a Carrier Gas for GC and GC–MS." https://www.chromatographyonline.com/view/hydrogen-carrier-gas-gc-and-gc-ms.

[38] A. Faricha, Suwito, M. Rivai, M. A. Nanda, D. Purwanto, and R. P. R. Anhar, "Design of electronic nose system using gas chromatography principle and Surface Acoustic Wave sensor," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 16, no. 4, pp. 1458–1467, 2018, doi: 10.12928/TELKOMNIKA.v16i4.7127.

[39] X. Chen *et al.*, "Gas chromatograph-surface acoustic wave for quick real-time assessment of blood/exhaled gas ratio of propofol in humans," *Br. J. Anaesth.*, vol. 113, no. 5, pp. 807–814, 2014, doi: 10.1093/bja/aeu193.

[40] G. Sauerbrey, "Verwendung von Schwingquarzen zur Wägung dünner Schichten und zur Mikrowägung," *Zeitschrift für Phys.*, vol. 155, no. 2, pp. 206–222, 1959, doi: 10.1007/BF01337937.

[41] J. Liu and Y. Lu, "Response mechanism for surface acoustic wave gas sensors based on surface-adsorption," *Sensors (Switzerland)*, vol. 14, no. 4, pp. 6844–6853, 2014, doi: 10.3390/s140406844.

[42] T. Wang, R. Green, R. Guldiken, J. Wang, S. Mohapatra, and S. S. Mohapatra, "Finite element analysis for surface acoustic wave device characteristic properties and sensitivity," *Sensors (Switzerland)*, vol. 19, no. 8, 2019, doi: 10.3390/s19081749.

[43] "Engineering ToolBox, (2018). Methanol - Thermophysical Properties." https://www.engineeringtoolbox.com/methanol-methyl-alcohol-properties-CH3OH-d_2031.html.

[44] "Engineering ToolBox, Toluene - Thermophysical Properties," 2018, [Online]. Available: https://www.engineeringtoolbox.com/toluene-methylbenzene-properties-d_2095.html.

[45] B. Taxis, M. Zalaf, and W. A. Wakeham, "Thermal conductivity and thermal diffusivity of xylene isomers in the temperature range 308-360 K at pressures up to 0.38 GPa," *Int. J. Thermophys.*, vol. 9, no. 1, pp. 21–35, 1988, doi: 10.1007/BF00503997.

[46] L. Niobate, "sensor using surface acoustic wave device based on LiNbO 3 / SiO 2 / Si substrate," vol. 095317, no. July 2016, 2018.

[47] E. G. Turitsyna and S. Webb, "Simple design of FBG-based VSB filters for ultra-dense WDM transmission ELECTRONICS LETTERS 20th January 2005," *Electron. Lett.*, vol. 41, no. 2, pp. 40–41, 2005, doi: 10.1049/el.



Jitender Kumar, received his M.Sc. degree in organic chemistry from the University of Meerut in 2004. He is currently pursuing the Ph.D. from Delhi University, Delhi. He is currently a Senior Scientist with the Solid State Physics Laboratory, New Delhi. He is involved in the development & characterization of Gas Chromatography and SAW-based chemical sensor. He is also involved in the development of vapor sampler and preconcentrator for toxic industrial chemicals.



A.T. Nimal received his Ph.D. degree in physics from the National Physical Laboratory, University of Delhi, New Delhi, in 2004. He is working as a senior Scientist with the Solid State Physics Laboratory, New Delhi, since 1997. He was also involved in the design and development of SAW devices such as delay lines, dispersive delay lines, resonators, and filters for the above sensors and signal processing. He is currently involved in the development of SAW-based electronic nose for CW agents.



Upendra Mittal received his M.E. degree in electrical engineering from the Delhi College of Engineering, University of Delhi, in 2003, and the Ph.D. degree in electrical engineering from Jamia Millia Islamia University, New Delhi, in 2016. He is working as a senior Scientist with the Solid State Physics Laboratory, New Delhi. He is also working for the development of SAW-based E-nose systems. His current research interests RF circuit design, RF characterization, and the development of chemical vapor sensors.



Vinod Kumar Currently,he is an Assistant Professor in the Special Centre for Nanosciences in Jawaharlal Nehru University, Delhi, India. He completed his Ph.D. in Materials Chemistry from Department of Chemistry and M.Sc. and B.Sc. (Hons) Chemistry from Kirori Mal College, University of Delhi, India. His research focuses on the "Synthesis, structure and properties of binary/mixed metal oxides nanomaterials" and their applications in water splitting, electrode materials of Li-ion batteries, SAW based sensors, water purification and nanomedicine along with nanotoxicity.



Vinay Kumar Singh received his PhD from Delhi University of Delhi, in 1990. Presently, he is a Professor at Sri Aurobindo College, University of Delhi. He is involved in the development and characterisation of GC-SAW based chemical sensor for the detection of Toxic Industrial chemicals and gases.

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 March 29,2022 at 09:41:30 UTC from IEEE Xplore. Restrictions apply.

Applied Physics A (2022) 128:302
https://doi.org/10.1007/s00339-022-05440-0

Applied Physics A
Materials Science & Processing

Effect of Cr doping on $Li_2ZnTi_3O_8$ as alternative anode material to enhance electrochemical properties of lithium-ion batteries

Anchali Jain[1] · Amrish K. Panwar[1] · Pawan K. Tyagi[1]

Received: 23 January 2022 / Accepted: 25 February 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2022

Abstract
Cr-doped and undoped lithium zinc titanate (LZTO), $Li_2ZnCr_xTi_{3-x}O_8$ ($x = 0, 0.1, 0.3, 0.5$) has been prepared as alternative anode using a simple and facile high-temperature solid-state ball-milling route followed by calcination in the air atmosphere. The cubic spinel structure with $P4_332$ space group is confirmed through X-ray Diffraction (XRD) for all compositions indicating doping of Cr^{3+} does not cause any changes within the lattice as well as maintain the phase structure up to an optimized doping. The study of morphology and distribution of the particles in the synthesized LZTO and doped LZTO compositions are carried out via scanning electron microscopy and transmission electron microscopy (TEM). It reveals the formation of polyhedron-shaped spherical facets in the sub-micron to nanometer range as observed by TEM. The particle size reduces toward the nanometer range as the amount of Cr doping increases and it also decreases the agglomeration of particles. All the samples doped with Cr indicate good electrochemical performance as compared to pristine LZTO. Sample $Li_2ZnCr_{0.3}Ti_{2.7}O_8$ doped with Cr content 0.3 displays higher discharge capacity, better cyclability, and lower polarization among all other Cr doped samples. $Li_2ZnCr_{0.3}Ti_{2.7}O_8$ delivered excellent rate performances attaining capacities of 251.73 ± 4, 184.34 ± 5, 157.42 ± 5, and 119.03 ± 4 mAh g^{-1} at 0.1C, 0.5C, 1, and 2C, respectively, which is higher than the pristine LZTO sample. Electrochemical impedance spectroscopy shows $Li_2ZnCr_{0.3}Ti_{2.7}O_8$ acquired a lower value of charge-transfer resistance and the fastest Li-ion diffusivity which enhances its capacity at higher current rates and makes it suitable as an alternative anode for lithium-ion batteries.

Keywords Lithium-ion batteries · Anode · Lithium zinc titanate

1 Introduction

Rechargeable lithium-ion batteries (LIBs) have been tremendously used as a prominent power source for next-generation electric vehicles due to their distinct features such as high coulombic efficiency (99%), high energy/power densities, and low memory effect [1, 2]. To date, graphite has been extensively used as a commercialized anode material because of its cost-effectiveness, vast availability, and high theoretical capacity (372 mAh g^{-1}) [3]. However, graphite leads to the dendrites formation on the surface during charge–discharge, which creates severe safety issues [4, 5]. Additionally, graphite undergoes volume expansion (~ 10%) during Li^+ intercalation/de-intercalation which results in the capacity fading [4, 5]. To overcome these crucial issues, titanium-based compounds such as lithium titanate, $Li_4Ti_5O_{12}$ (LTO) are currently being investigated as alternative anode due to their "zero-strain characteristics", i.e., no structural change during the mechanism of inserting/de-inserting of Li^+ and better safety [6]. However, LTO also has several disadvantages such as higher-intercalation potential (~ 1.55 V), lower theoretical capacity (175 mAh g^{-1}), low electronic conductivity (~ 10^{-13} S cm^{-1}), and poor lithium-ion diffusion coefficient (~ 10^{-9} to 10^{-13} cm^2 s^{-1}) which prevents LTO to be used in high power applications [7, 8].

Recently, a new class of cubic spinel material $Li_2MTi_3O_8$; where M may be Zn, Co, Ni, or Mg, etc. has received considerable attention due to its non-polluting nature, safety measures, and low cost as compared to LTO containing less lithium content [9, 10]. Among the cubic spinel materials, $Li_2ZnTi_3O_8$ displays better properties as an alternative prospect of anode material (other than LTO) owing to high theoretical capacity, 227mAh g^{-1}, lowering in de-intercalation

✉ Amrish K. Panwar
amrish.phy@dtu.ac.in

[1] Department of Applied Physics, Delhi Technological University, Delhi, India

Published online: 15 March 2022

plateau, 0.5–0.8 versus Li/Li^+, high electronic conductivity, and low cost [11–13]. Usually, $Li_2ZnTi_3O_8$(LZTO) formed a three-dimensional tunnel network and can be described in space notation as $(Li_{0.5}Zn_{0.5})_{tet}[Li_{0.5}Ti_{1.5}]_{oct}$, where Li^+ and Zn^+ occupy tetrahedral sites in 1:1, here Li^+ and Ti^{3+}occupy octahedral sites in 1:3, respectively [11, 12]. The interaction with three Li^+ can be described as: $Li_2ZnTi_3O_8 + 3Li^+ + 3e^- \rightarrow Li_5ZnTi_3O_8$ [11]. However, one of the main disadvantages of LZTO is its poor ionic and electronic conductivity. Therefore, numerous efforts have been made to advance its conductivity and electrochemical performance using different approaches such as preparation of nano-sized material [13, 14], coating with conductive material [9, 15–17], and doping of other ions at different sites like Ag^+ [18], Al^{3+} [19], Cu^{2+} [20], Na^+ [21], Ce^{4+} [22] and many more. Doping can modify the bonding between the atoms by changing its length and strength, introducing the defects/voids into the lattice structure, and also build up the path for lithium-ion diffusion which improvises the migration of Li^+ and e^- within the lattice structure, thereby varying the electronic/ionic conductivity, diffusion coefficient and stability of the materials as well as enhancing the overall electrochemical performance of LIBs [23].

In this study, the synthesis of $Li_2ZnTi_{3-x}Cr_xO_8$ ($x = 0$, 0.1, 0.3, 0.5) has been attempted using a solid-state route, and its structural, morphological, and electrochemical properties are studied. The solid-state reaction method is chosen because of its low synthesis cost and simple reaction route. The effect of chromium dopant has been studied at the Ti site over other metal ions, due to its advantages of similar ionic radii with Ti^{4+} as well as the doping of Cr in place of Ti^{3+}/Ti^{4+} in the crystal structure provides more structural stability over octahedral sites in the spinel structure [23].

2 Experimental

2.1 Material synthesis

2.1.1 Preparation of $Li_2ZnTi_{3-x}Cr_xO_8$

Solid-state route by ball-milling was used to prepare Cr-doped $Li_2ZnTi_{3-x}Cr_xO_8$ ($x = 0$, 0.1, 0.3, 0.5). In this process, analytical grade, Li_2CO_3 (CDH, 98.5%), $Zn(CH_3COO)_2{\cdot}2H_2O$ (Sigma-Aldrich, > 98%), TiO_2 (Merck, > 99%), and Cr_2O_3 (Sigma Aldrich, > 98%) were used as starting precursors without any further purification. The stoichiometric amount of all precursors was ball milled for 4 h in presence of a small quantity of de-ionized water for homogeneous mixing. A 5% excess of lithium source is further added to balance its evaporation at higher temperatures. After mixing, the as-prepared sample was dried at 80 °C for 12 h to remove any water content. The obtained as-prepared dried material was grounded and calcined at 700 °C for 3 h and again treated at 800 °C for another 5 h in air. Lastly, the calcined material was re-grounded and used for further electrode preparation and characterizations.

2.1.2 Electrode preparation and Li coin cells

The electrochemical tests were performed with half cells, CR2016 cell assembly keeping pristine/doped LZTOs samples as working electrodes, and lithium metal as reference/counter electrodes. Hence, active material, Li metal, and polypropylene sheet Celgard 2400 were used as working electrode, reference electrode, and separator, respectively. Ethyl carbonate and dimethyl carbonate have been used to prepare 1 mol/L $LiPF_6$ electrolyte in a volume ratio of 1:1. Active material, super P, and polyvinylidene difluoride (PVDF) were homogeneously mixed in *N*-methyl-2-pyrrolidine (NMP) at a ratio by weight of 80:10:10 to prepare the slurry of the working electrode. Copper current collector was used for coating of resultant slurry using an automatic coating unit followed by drying at 120 °C for overnight. Further, the electrodes of 16 mm in diameter were punched and dried in a vacuum oven at 60 °C for overnight to eliminate any moisture content and brought in the argon-filled (O_2/H_2O levels less than 0.5 ppm) glove box work station made by MBraun and assembling of half cells, CR 2016 was attempted in the glove box.

2.2 Materials characterizations

Rigaku's Ultima IV-Powder X-ray Diffractometer was used to characterize all calcined samples over an angular range, 2θ of 10°–90° with a source of radiation, Cu$K\alpha$ of wavelength, 0.154 nm to investigate their crystal structure. Thermogravimetric experiment was done to study the thermal stability of the material via Perkin Elmer make model TGA 4000 equipment. ZEISS make scanning electron microscopy model EV018 was used to study the morphological analysis. FEI make 'FEI Tecnai G^2 20 S-Twin' microscope was used to observe transmission electron microscopy (TEM) images. For electrical measurements, all the samples were mixed separately with the polyvinyl alcohol (PVA, 2.5 wt%) and pressed into circular disks of 10 mm with the help of a hydraulic press by applying the pressure of 7 tons. To remove the PVA, pallets were heated at 250 °C for 2–3 h in air. To make electrical contacts, Silver (Ag) paste was applied on both sides of pallets followed by drying at 150 °C for 1 h in the air to remove any moisture content. D.C. Resistance, electronic conductivity, and activation energy measurements were done using the source voltage varied from − 10.0 V to 10.0 V and 1.0 V, respectively, with the help of Kiethley 6430 source meter at room temperature. Electrochemical tests were performed using Biologic Science Instruments make VMP3 cell test workstation at ambient temperature. Cyclic Voltammetry

(CV) was executed at a scanning rate of 0.05 mV s^{-1} which is further supported by Galvanostatic Charge/Discharge (GCD) tests within the voltage range 0.05–3.0 V(vs. Li/Li^+) for all assembled cells at different current rates. Investigation of the electrochemical impedance spectroscopy (EIS) was done within the frequency range of 0.1–10 MHz using an AC pulse signal of 5 mV.

3 Results and discussion

3.1 Thermogravimetric (TGA) and differential thermal (DTA) studies

To study the formation mechanism of the LZTO material sample, an as-prepared ball milled sample (reactants: Li_2CO_3, $Zn(CH_3COO)_2{\cdot}2H_2O$, TiO_2) was examined by TGA/DTA, and results are shown in Fig. 1. It can be noted from the TGA curve that up to 175 °C there is a slight loss in mass (~7.23%) which is accompanied by an endothermic peak around 335 °C due to the desorption or removal of water content in the sample. As the temperature rises, the weight loss gradually decreases to 700 °C accompanied with the broad exothermic peaks which can correspond to the loss of CO_2 from Li_2CO_3 as depicted by the chemical reaction involved in this whole process [24]:

$$Li_2CO_3 + ZnO + 3TiO_2 \rightarrow Li_2ZnTi_3O_8 + CO_2 \uparrow \quad (1)$$

After 700 °C, there is no further reaction occurs which shows that the LZTO phase formation has started beyond 700 °C.

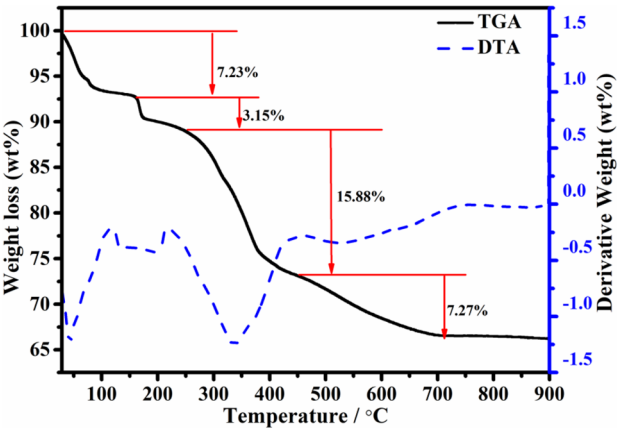


Fig. 1 TGA/DTA curve of as-prepared LZTO anode from room air temperature to 900 °C at a rise rate of 10 °C min^{-1} in the air atmosphere

3.2 Structural analysis

To confirm the formation and crystalline nature of samples, XRD analysis was carried out. Figure 2a displays the XRD pattern of pristine $Li_2ZnTi_3O_8$ (LZTO) and Cr-doped $Li_2ZnTi_3O_8$, all peaks are well indexed to cubic spinel system matching with JCPDS card no. #086-1512 and confirming the space group $P4_332$. All the diffraction peaks are sharp indicating the crystalline nature of all samples and confirming that the Cr^{3+} has substituted in the crystal lattice structure. However, as the content of Cr increases to 0.5, a small peak relating to the Cr_2O_3 phase and broadening of peaks can be detected. It indicates that after a certain limit, Cr^{3+} is not able to replace Ti^{4+} in the crystal lattice. Figure 2b shows the narrow range enlarge diffraction peak of the (111) plane corresponding to the diffraction angle (2θ), 18.3° for all the

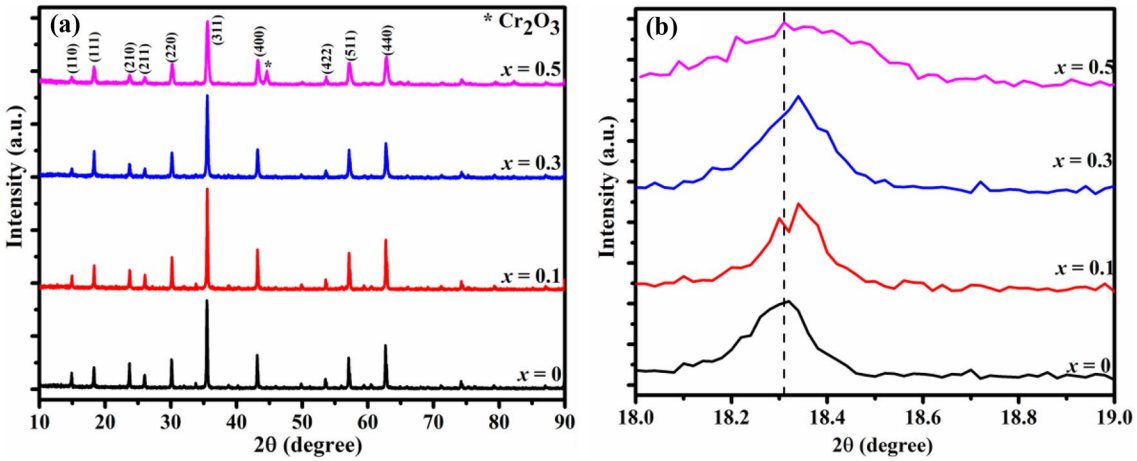


Fig. 2 **a** Wide range XRD patterns of pristine LZTO ($x=0$) and doped LZTO with Cr ($x=0.1$, 0.3, 0.5) prepared at 800 °C for 5 h, **b** Narrow range enlarged peak (111) of synthesized $Li_2ZnTi_{3-x}Cr_xO_8$ ($x=0$, 0.1, 0.3, 0.5) anode

samples which confirms the incorporation of Cr ion slightly shifts the peak toward the higher angle side, suggesting the decrease in lattice constant than pristine LZTO. Additionally, 0.3Cr doped LZTO shows sharp peaks compared to other samples, which suggests that this particular doping stabilizes the LZTO structure well. The lattice parameters of pristine and Cr doped LZTO are calculated and shown in Table 1. It is noticed that the lattice parameter decreases from 8.3689Å to 8.3610Å which could be associated with the smaller size of Cr^{3+} (0.064 nm) in comparison to Ti^{4+} (0.068 nm). As a result, crystal lattice could be contracted, causing the defects/distortions within the lattice structure, and therefore it is expected doping increase the lithium-ion diffusivity as well as electronic conductivity of the LZTO sample [19, 22, 23].

Further, for confirmation of size reduction, crystallite size corresponding to the dominant (311) peak has been calculated using Scherrer's formula described as in relation (2);

$$d = \frac{k\lambda}{\beta\cos\theta} \tag{2}$$

where d, λ, k, θ, and β are the size of crystallite in nm, wavelength of X-ray (around ~ 0.154 nm), shape constant (~ 0.94), grazing angle in degree, and FWHM corresponding to θ in radian, respectively. Table 1 shows that as the content of Cr^{3+} increases, crystallite size decreases linearly which makes Cr^{3+} doping favorable for improved electrochemical performance.

3.3 Morphological analysis

Scanning electron microscopy (SEM) micrographs of calcined pristine (0Cr) LZTO and Cr-doped samples (0.1Cr, 0.3Cr, 0.5Cr) are shown in Fig. 3a–d. It is visible from SEM micrographs that the pristine LZTO shows prismatic shaped agglomerated particle in the sub-micrometer range with a particle size of, 0.1–0.35 µm as calculated by Image-J software. It is noticed that on doping of Cr there is a decrement in the size of particles as 80–200 nm. It can also be seen that Cr doping also decreases the agglomeration of particles and increases the homogeneity of particles in the material

Table 1 Structural parameters calculated for $Li_2ZnTi_{3-x}Cr_xO_8$ ($x = 0$, 0.1, 0.3, 0.5)

Sample name	Lattice constant ($a = b = c$) (Å)	Volume V (cm^{-3})	Crystallite size 'd' (nm)
$Li_2ZnTi_3O_8$	8.3689	586.14	47.25
$Li_2ZnCr_{0.1}Ti_{2.9}O_8$	8.3675	585.86	46.39
$Li_2ZnCr_{0.3}Ti_{2.7}O_8$	8.3629	585.43	35.48
$Li_2ZnCr_{0.5}Ti_{2.5}O_8$	8.3610	584.48	23.60

LZTO. The transmission electron microscopy (TEM) images for undoped (0Cr) and doped (0.3Cr) samples are shown in Fig. 4a, b. The TEM image (Fig. 4a) shows the dense facets of a spherical-shaped polyhedron of pristine LZTO particles which is further modified into clear shape and faces of these spherical polyhedron facets with reduced size and less agglomeration of particles after the doping of 0.3Cr. This improvement in the morphology and particle size could be advantageous in enhancing the electrical conductivities and electrochemical characteristics of LZTO after Cr doping.

3.4 Conductivity measurements

The dc conductivity measurement is performed for all the pristine and Cr doped LZTO samples on the pallets and conductivity is calculated using relation, $\sigma = (L)/(R)\,(A)$, where σ, R, A, and L are conductivity, resistance in ohms (Ω), cross-section area (cm^2) and thickness of the pallet, respectively. Using the Arrhenius equation, activation energy (E_a) is calculated which is described as; $\sigma_{d.c.} = \sigma_o \exp(-E_a/K_B T)$, where T is the room temperature in Kelvin, and K_B is the Boltzmann constant in JK^{-1}. Table 2 displays the electronic conductivity and activation energy for all the samples. It is observed that as the content of Cr doping increases, the activation energy for all the samples decreases, and electronic conductivity increases. This suggests that enhanced electronic conductivity could be favorable for improved electrochemical performance synthesized material as an anode.

3.5 Electrochemical properties

3.5.1 Cyclic voltammetry (CV)

The electrochemical performance of the synthesized samples of pristine and Cr doped LZTO has been carried out in terms of CV, EIS, and GCD tests. Figure 5a, b shows the CV curves for cycles 1 and 3 of Cr-doped and undoped samples measured between 0.05 and 3.0 V at 0.05 mV s^{-1} scan rate. Both cathodic and anodic peaks of Cr-doped samples are similar to the pristine $Li_2ZnTi_3O_8$ sample, suggesting that doping does not alter the primary electrochemical mechanism of the LZTO material. In the initial cycle, all the curves show cathodic and anodic pair peaks around 1.0–2.0 V which can be associated with the oxidation reaction of Ti^{4+}/Ti^{3+} during the lithiation/de-lithiation process [13, 16, 25]. Also from the third cycle, it can be observed that peaks shift toward higher potential which is associated with the transition of phase from spinel to rock salt phase [17, 19, 25]. Additionally, from the first to third cycle, same pattern has been followed by all the CV curves which suggest after the initial cycle material stabilizes for further cycles. However, from both curves it can be concluded that after Cr doping cathodic/anodic peaks current intensifies which indicates

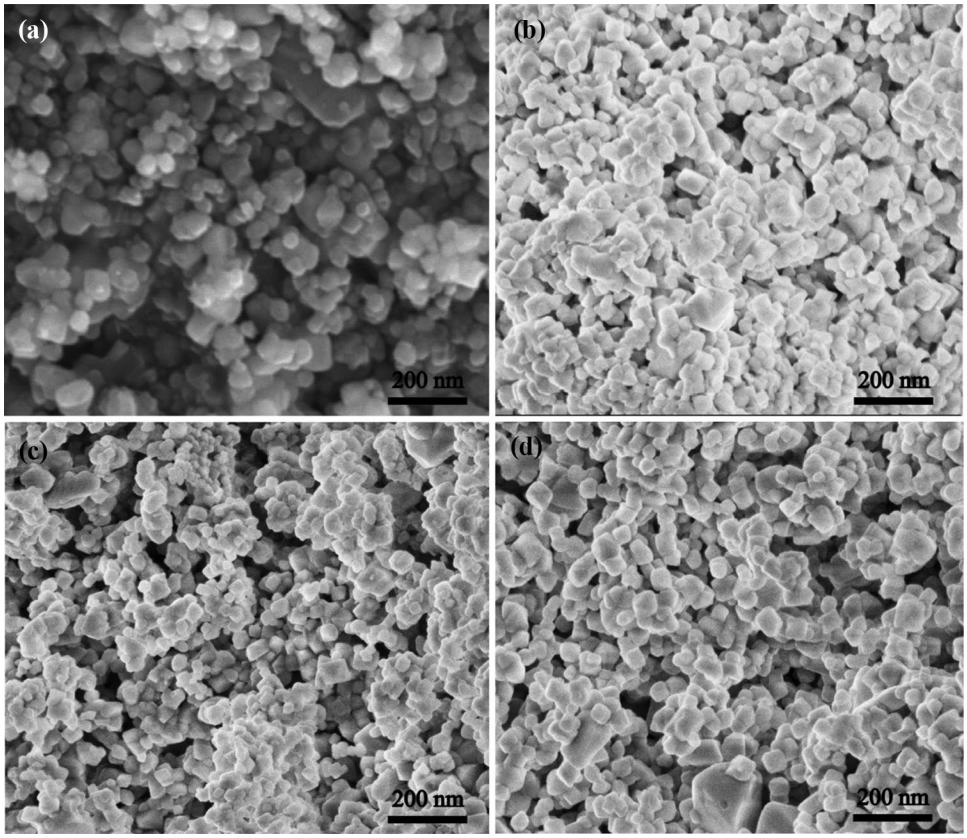


Fig. 3 **a–d** SEM images at 50X magnification for: **a**-pristine LZTO; Cr doped LZTO at **b**-$x=0.1$, **c**-$x=0.3$, and **d**-$x=0.5$ synthesized at 800 °C for 8 h in air

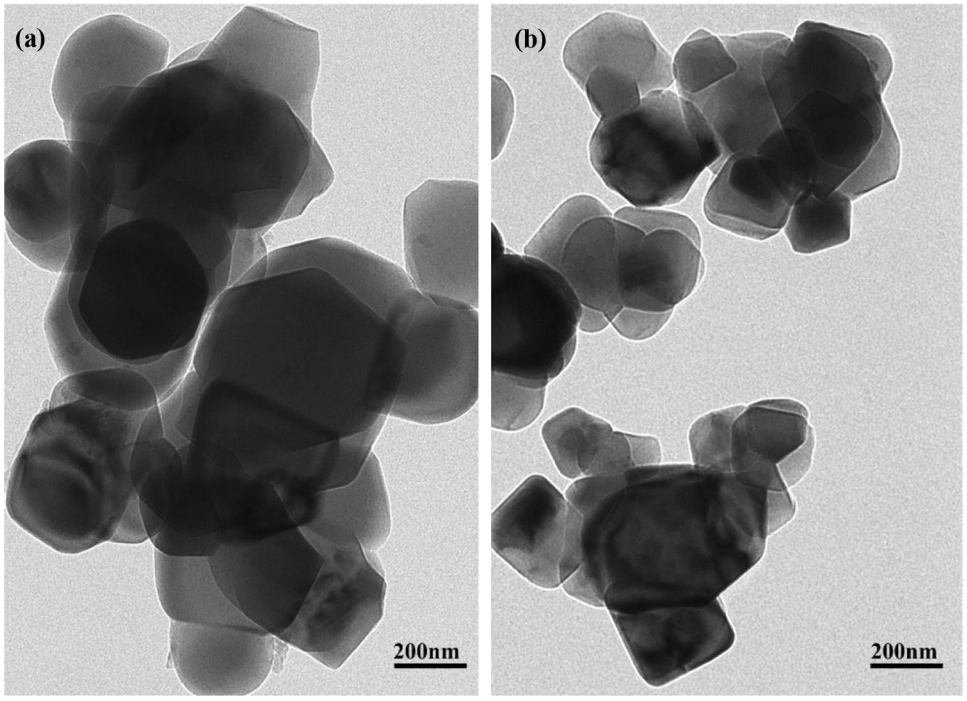


Fig. 4 TEM micrographs observed for: **a** pristine LZTO, **b** 0.3Cr doped LZTO synthesized at 800 °C for 8 h in air

Table 2 Electrical parameters calculated for $Li_2ZnTi_{3-x}Cr_xO_8$ ($x=0$, 0.1, 0.3, 0.5)

Sample	Activation energy (meV)	D.C. resistance (Ω)	Electronic conductivity (S cm^{-1})
$Li_2ZnTi_3O_8$	997.9	2.60×10^{7}	9.14×10^{-9}
$Li_2ZnCr_{0.1}Ti_{2.9}O_8$	979.2	3.16×10^{8}	6.39×10^{-10}
$Li_2ZnCr_{0.3}Ti_{2.7}O_8$	957.1	1.70×10^{6}	7.46×10^{-8}
$Li_2ZnCr_{0.5}Ti_{2.5}O_8$	991.8	5.70×10^{6}	2.18×10^{-8}

that kinetics of lithium-ion intercalation and de-intercalation improves, thereby increasing the specific capacity of the material. There is also a reduction in peak intensity of around 0.4–0.5 V that is observed for all samples. It is associated with the multiple restoration of Ti^{4+}, which is in good agreement with the previous reports [16–20].

Further, to examine the degree of polarization of the electrodes, potential difference values between the peaks of cathodic and anodic scan for all samples are tabulated in Table 3, where V_a and V_b are the potentials of anodic and cathodic peaks, respectively. It can be concluded that Cr doping provides the stability during the insertion-extraction process of Li^+, thereby improving the reversibility of electrodes. Among all, doping of 0.3Cr has the lowest potential difference (320 mV) which shows at this particular doping LZTO possesses the lowest polarization. Therefore, CV curves imply that Cr-doped LZTO improves the kinetic performance as well as has higher reversibility than the pristine LZTO electrode.

3.5.2 Electrochemical impedance spectroscopy (EIS)

To study the kinetics of lithium ions movement from electrode surface to electrolyte and vice-versa EIS has been performed on all prepared cells of pristine and Cr doped LZTO. Figure 6 shows the Nyquist plots for all the samples before the cycling. The equivalent circuit diagram used and narrow range enlarged region of Nyquist plots for better display of impedances of electrodes are shown as the inset in Fig. 6. All the plots have depressed semicircles in the high-frequency region and low-frequency region relating to the charge transfer resistance, $R_{1,}$ and the diffusion of lithium ions in the electrode material, respectively [26, 27]. The Nyquist plots are fitted based on the equivalent circuit which represents solution resistance (R_1), charge transfer resistance (R_2), constant phase element (Q_2), and element used to describe restricted diffusion (M_2). The values of resistances are computed using the line of best fit and provided in Table 3. As shown in Table 3, 0.3Cr has low solution resistance as well as low charge transfer resistance which suggests that this particular doping has low ohmic resistances between electrode/electrolyte interface and has higher electrochemical activity relative to undoped LZTO. Interestingly, for 0.1Cr

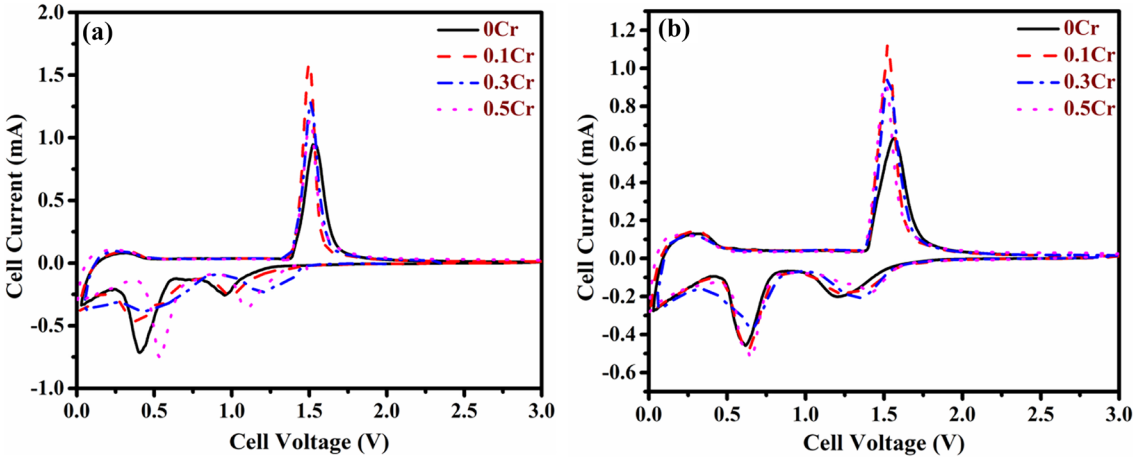


Fig. 5 Cyclic Voltammograms for **a** first cycle, **b** third cycle of synthesized $Li_2ZnTi_{3-x}Cr_xO_8$ ($x=0$, 0.1, 0.3, 0.5) as anode material

Table 3 Potentials and EIS parameters calculated for $Li_2ZnTi_{3-x}Cr_xO_8$ ($x=0$, 0.1, 0.3, 0.5)

Samples	V_a (V)	V_b (V)	V_{ab} (mV)	R_1 (ohm)	R_2 (ohm)	D_{Li} ($cm^2{\cdot}s^{-1}$)
$Li_2ZnTi_3O_8$	1.53	0.95	580	4.19	375.5	1.6×10^{-17}
$Li_2ZnCr_{0.1}Ti_{2.9}O_8$	1.49	0.99	500	6.61	315.9	9.5×10^{-17}
$Li_2ZnCr_{0.3}Ti_{2.7}O_8$	1.51	1.19	320	3.80	68.86	1.8×10^{-16}
$Li_2ZnCr_{0.5}Ti_{2.5}O_8$	1.50	1.09	410	4.60	116.0	8.6×10^{-16}

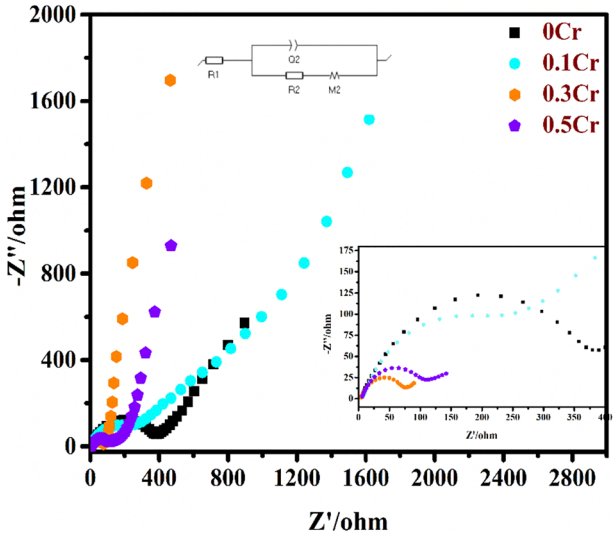


Fig. 6 Nyquist Plot measured at AC pulse signal of amplitude 5 mV of all samples of $Li_2ZnTi_{3-x}Cr_xO_8$ ($x=0$, 0.1, 0.3, 0.5) anode

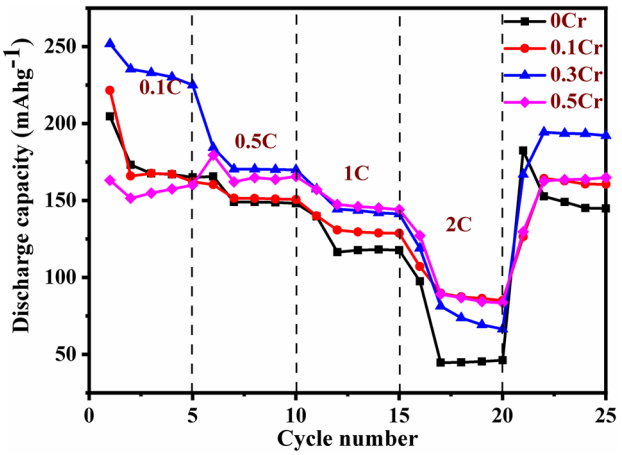


Fig. 7 Rate performance of $Li_2ZnTi_{3-x}Cr_xO_8$ ($x=0$, 0.1, 0.3, 0.5) anode recorded at different C-rate ranging from 0.1 to 2C

sample resistances values indicate that diffusion of lithium ions is poor at a lower current rate same as 0Cr that can be validated from rate capability results. However, 0.5Cr shows intermediate results between both of the doping concentrations. Therefore, it is reasonable that the optimized 0.3Cr has the best electrochemical results as compared to the pristine LZTO electrode.

Diffusion coefficient (D_{Li+}) for all the samples can be estimated based on the relation of restricted linear diffusion which depends on the thickness of the electrode (δ) and diffusion time constant (τ_d) which is described according to relation (3):

$$D_{Li^+} = \frac{\delta^2}{\tau_d} \tag{3}$$

The calculated diffusion coefficient for all the samples is tabulated in Table 3. The diffusivity of lithium ions in electrode material escalates as the content of Cr^{3+} increases thereby improving the chemical kinetics of reaction during charge/discharge cycles [13, 22].

3.5.3 Galvanostatic charge/discharge analysis (GCD)

To study the electrodes can function at high power applications, they are being subjected to varied current rates. Figure 7 shows the rate capability of $Li_2ZnCr_xTi_{3-x}O_8$ ($x=0$, 0.1, 0.3, 0.5) samples measured at varied scan rates ranging from 0.1 to 2C. There is a difference between first and second cycle discharge capacities for all samples and after that the capacities are almost the same. This loss in capacity is ascribed to the irreversible reactions that occurred on the surface of the electrodes or the formation of the solid electrolyte interface (SEI) layer thereby reducing the diffusion of lithium-ion through the material [28]. The first and second cycles discharge capacities results at different C-rates are tabulated in Table 4. Conversely, at 0.5Cr doping, the first discharge capacity is 163.18 mAh g^{-1} at 0.1C which is lower than the pristine (0Cr) LZTO sample. As majorly of lithium ions contribute to the electrochemical energy during the insertion/de-insertion process, hence it may possible that there could be the presence of Cr^{3+} at the Li^+ octahedral site rather than Ti^{4+}, which could be a reason to deduct the amount of lithium ions and thereby reduces the initial discharge capacity [29]. However, on increasing the rate of current, 0.5Cr LZTO sample capacity improved while the capacity of pristine LZTO decreases rapidly. It has been noticed from the results shown in Table 4 that Cr doping not only boosts the capacity but also stabilizes the electrochemical performance of the material. On comparing discharge capacities of all the samples of pristine and Cr doped LZTO, sample 0.3Cr displays improved specific discharge capacities of 251.73 ± 4, 184.34 ± 5, 157.42 ± 5, and 119.03 ± 4 mAh g^{-1} at 0.1C, 0.5C, 1C, and 2C, respectively. These specific discharge capacities at different rates are better as compared to the pristine (0Cr) of specific discharge capacities of 204.59 ± 4, 165.57 ± 4, 139.61 ± 3, and 97.49 ± 3 mAh g^{-1} at 0.1C, 0.5C, 1C, and 2C, respectively. Therefore, it can be observed that the doping of Cr significantly enhances the performance of LZTO material.

Cycling performances and stability are major features to study for battery applications. Figure 8a, b displays the discharge capacities of pristine (0Cr) LZTO and 0.3Cr LZTO at a constant C-rate of 0.1 C. For the first cycle, the discharge capacity is found to be 255.12 mAh g^{-1} for cell made by sample 0.3Cr, which is improved as compared to the pristine sample and discussed in detail in GCD. In both cases,

Table 4 Galvanostatic discharge of $Li_2ZnTi_{3-x}Cr_xO_8$ by changing C-rate ($x = 0, 0.1, 0.3, 0.5$) through the first and second cycles

Sample	0.1C/ (mAh g^{-1})		0.5C/ (mAh g^{-1})		1C/ (mAh g^{-1})		2C/ (mAh g^{-1})	
	First	Second	First	Second	First	Second	First	Second
$Li_2ZnTi_3O_8$	204.59 (±4)	173.10 (±4)	165.57 (±4)	148.96 (±4)	139.61 (±3)	116.45 (±3)	97.19 (±3)	44.63 (±4)
$Li_2ZnCr_{0.1}Ti_{2.9}O_8$	221.51 (±5)	165.91 (±6)	160.26 (±7)	151.45 (±3)	140.61 (±6)	130.76 (±3)	107.10 (±3)	89.78 (±7)
$Li_2ZnCr_{0.3}Ti_{2.7}O_8$	251.73 (±4)	235.34 (±4)	184.34 (±5)	170.37 (±4)	157.42 (±5)	144.41 (±4)	119.03 (±4)	81.28 (±4)
$Li_2ZnCr_{0.5}Ti_{2.5}O_8$	163.18 (±)5	151.41 (±)5	179.51 (±5)	162.13 (±6)	157.64 (±4)	147.4 (±5)	127.21 (±5)	89.21 (±5)

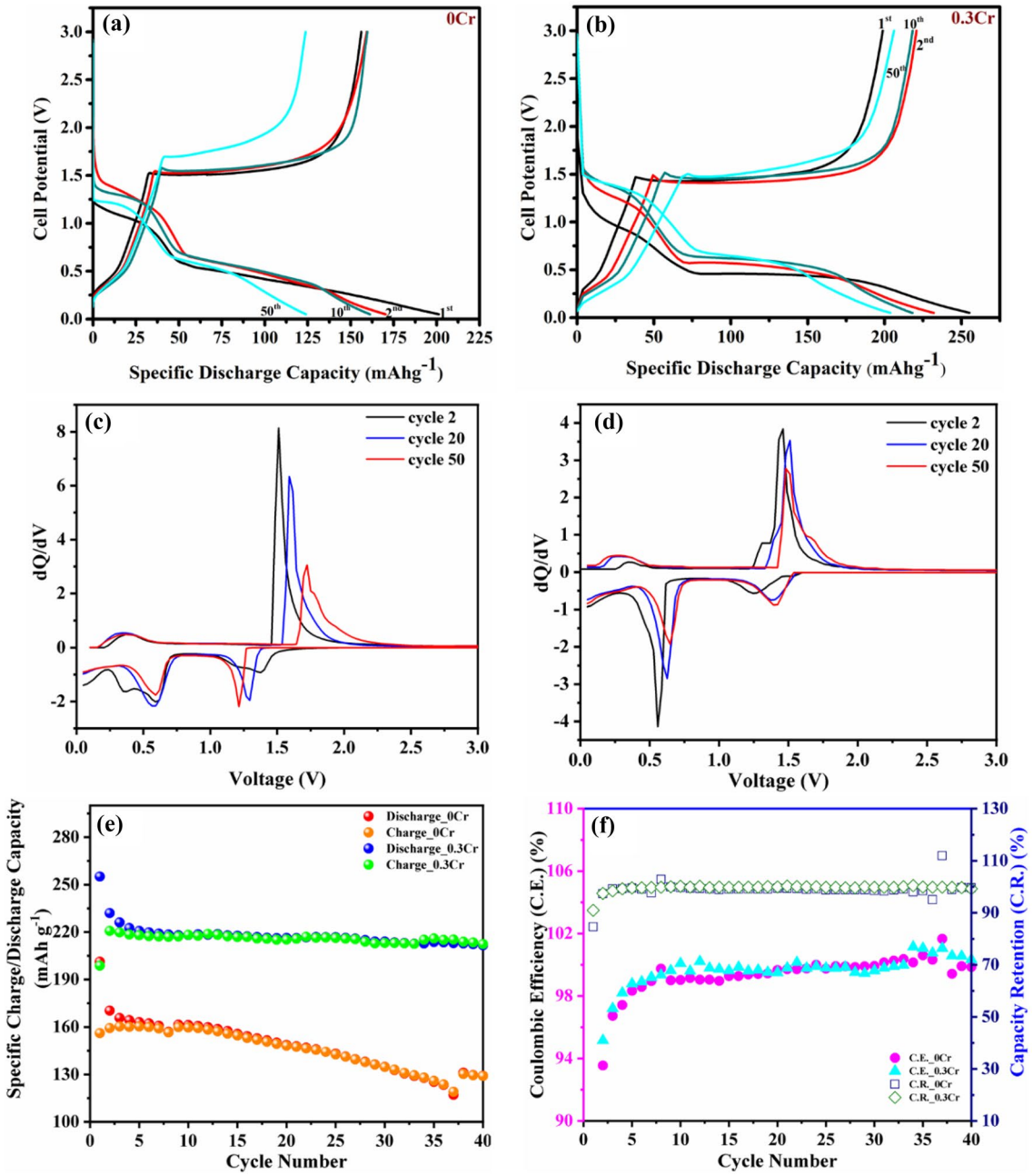


Fig. 8 First cycle charge–discharge performance of cell tested for cycling analysis: **a**, **b** Pristine (0Cr) and 0.3Cr doped LZTO samples at 0.1C, respectively, **c**, **d** Corresponding dQ/dV plots of 2nd, 20th, and 50th cycle of 0Cr and 0.3Cr doped LZTO, **e** Cycling performance during charging-discharging for 0Cr and 0.3Cr sample of LZTO, and **f** Coulombic efficiency versus cycle number and capacity retention versus cycle number of 0Cr and 0.3Cr LZTO anode

it has been observed that due to the formation of the SEI layer, there was a capacity loss after the initial cycle. To investigate the change in capacity with respect to voltage, differential capacities plots are studied. Figure 8c, d show the corresponding dQ/dV plots for the 2nd, 20th and 50th cycles. From both the curves, it can be observed that there is a shift of dQ/dV peaks (~ 1.50 V) toward higher potential (~ 1.72 V) in case of extraction of Li^+, simultaneously, dQ/dV peaks (~ 1.30 V) shifts toward lower potentials (~ 1.21 V) during insertion of Li^+. These shifts could be resulted due to the formation of the SEI layer after 2nd cycle at the electrode surface [30–32]. Moreover, there is also reduction peak around ~ 0.5 to 0.6 V which could be ascribed to multiple restoration of Ti^{4+} [32]. However, from Fig. 8d, in the case of 0.3Cr, the potential shift predominantly decreases which indicates that doping of Cr stabilizes the electrode structure. Additionally, the potential for the 20th and 50th cycles are nearly same, i.e., 1.50 V and 1.48 V, respectively, which suggests the improved electrochemical performance after the doping of Cr. Furthermore, it can be seen that as the cycle number increases, the intensity of dQ/dV peaks reduces for both cases, which may be ascribed to the fading of active material electrodes [30]. However, after the doping of Cr, high intense and sharp peaks show the improved reversible nature of the LZTO electrode. Furthermore, the polarization of LZTO material, ΔV reduces after the doping of Cr ions which could also be confirmed from CV results.

Figure 8 e shows the cycling performance versus cycle number of both the electrodes during charging and discharging, depicting that the capacity remains almost same as the number of cycles increases. However, in the case of 0Cr, it can be observed that there is a loss in capacity while 0.3Cr shows improved electrochemical performance. For the 40th cycle, $Li_2ZnTi_{2.7}Cr_{0.3}O_8$ sample still holds 211.64 mAh g^{-1} with ~ 99% capacity retention. The discharge capacity of 0.3Cr is higher and more stable in comparison to 0Cr, which may be due to the reasons: (1) reduction in particle size shorten the path-length for lithium ions and electrons migration within the active electrode, and (2) enhanced electronic conductivity and lower resistances allows the rapid Li^+ intercalation/de-intercalation. This result reveals that the Cr-doped sample exhibits a higher discharge capacity than the pristine sample.

To study the efficiency of electrodes to transfer charge during the electrochemical reaction, coulombic efficiency (C.E.) is calculated for 0Cr and 0.3Cr LZTO electrodes as shown in Fig. 8f. C.E. is represented as the ratio of discharge capacity after a full charge to the charge capacity for the same cycle. From Fig. 8 f, it can be seen that the coulombic efficiency is maintained at ~ 98 to 99% after the 3rd cycle for the 0.3Cr LZTO sample, while it varies for 0Cr sample. To study the ability of the material to retain the stored charge, capacity retention (C.R.) is calculated for both the electrodes and is shown in Fig. 8f for 0Cr and 0.3Cr electrodes. Hence, C.R. is referred as the ratio of the measured discharged capacity of cycle $n+1$ to the measured discharge capacity of cycle n.

4 Conclusions

This study concludes the synthesis of pristine and Cr-doped $Li_2ZnTi_3O_8$ samples via a high-temperature solid-state reaction route. The structural analysis by XRD results reveal that Cr doping does not affect the crystalline nature of pristine LZTO to a certain limit. Morphological analysis using SEM and TEM displays that as the content of Cr increases in LZTO, it results in spherical polyhedron-shaped facets with reduced size and agglomeration. The activation energy and electronic conductivity are also improved and observed to be 957.1 eV and 7.46×10^{-8} S cm^{-1}, respectively, for the 0.3Cr doped sample confirming the better electrical performance as compared to pristine LZTO. The electrochemical analysis reveals that $Li_2ZnCr_{0.3}Ti_{2.7}O_8$ shows improved specific capacity of 251.73 ± 4, 184.34 ± 5, 157.42 ± 5, and 119.03 ± 4 mAh g^{-1} for an optimized sample of Cr doping, 0.3 Cr at 0.1C, 0.5C, 1C, and 2C, respectively. Thus, it is reasonable to say that the Cr doping should be optimized because it drastically modifies all the parameters of electrodes.

Acknowledgements The authors are thankful for all the guidance and financial assistance via project Grant No. F. No. DTU/IRD/619/2019/2114 to the lithium-ion battery technology (LIBT) lab, Delhi Technological University (DTU), New Delhi which helped in successfully conducting this research.

Declarations

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

References

1. Y. Ding, Z.P. Cano, A. Yu, J. Lu, Z. Chen, Electrochem. Energy Rev. **2**, 1 (2019)
2. R. Schmuch, R. Wagner, G. Hörpel, T. Placke, M. Winter, Nat. Energy **3**(4), 267–278 (2018). https://doi.org/10.1038/s41560-018-0107-2
3. A. Eftekhari, Energy Storage Mater. **7**, 157 (2017)
4. G.N. Zhu, Y.G. Wang, Y.Y. Xia, Energy Environ. Sci. **5**, 6652 (2012)
5. H. Tang, T. Liu, Z. Tang, Dalton Trans. **48**, 12303 (2019)
6. S. Wang, Y. Yang, Y. Dong, Z. Zhang, Z. Tang, J. Adv. Ceram. **8**(1), 1 (2019)
7. Y. Cai, Y. Huang, W. Jia, X. Wang, Y. Guo, D. Jia, Z. Sun, W. Pang, Z. GuoJ, Mater. Chem. **A4**, 9949 (2016)

8. B.V. Babu, K.V. Babu, G.T. Aregai, L.S. Devi, B.M. Latha, M.S. Reddi, K. Samatha, V. Veeraiah, Results Phys. **9**, 284 (2018)
9. H. Yang, N. Lun, Y.X. Qi, H.L. Zhu, J.R. Liu, J.K. Feng, L. Zhao, Y.J. Bai, Electrochim. Acta **315**, 24 (2019)
10. H. Kawai, M. Tabuchi, M. Nagata, H. Tukamoto, A.R. West, J. Mater. Chem. **8**, 1273 (1998)
11. Y.R. Wu, J. Pan, S. Ren, Y. Xie, C. Yue, T.F. Yi, Ionics **25**, 373 (2019)
12. Z. Li, Y. Cui, J. Wu, C. Du, X. Zhang, Z. Tang, RSC Adv. **6**, 39209 (2016)
13. L. Wang, Z. Meng, H. Wang, X. Li, G. Zhang, Ceram. Int. **42**, 16872 (2016)
14. T. Liu, H. Tang, L. Zan, Z. Tang, J. Electroanal. Chem. **771**, 10–16 (2016). https://doi.org/10.1016/j.jelechem.2016.03.036
15. H. Tang, J. Zhu, C. Ma, Z. Tang, Electrochim. Acta **144**, 76 (2014)
16. L. Wang, B. Chen, Z. Meng, B. Luo, X. Wang, Y. Zhao, Electrochim. Acta **188**, 135 (2015)
17. H. Tang, L. Zan, W. Mao, Z. Tang, J. Electroanal. Chem. **751**, 57 (2015)
18. H. Tang, Z. Tang, C. Du, F. Qie, J. Zhu, Electrochim. Acta **120**, 187 (2014)
19. H. Tang, J. Zhu, Z. Tang, C. Ma, J. Electroanal. Chem. **731**, 60 (2014)
20. F. Qie, Z. Tang, Mater. Express **4**, 221 (2014)
21. W. Chen, Z. Zhou, R. Wang, Z. Wu, H. Liang, L. Shao, J. Shu, Z. Wang, RSC Adv. **5**, 49890 (2015)
22. C. Chen, C. Ai, X. Liu, Y. Wu, Electrochim. Acta **227**, 285 (2017)
23. T.F. Yi, T.T. Wei, Y. Li, Y.B. He, Z.B. Wang, Energy Storage Mater. **26**, 165 (2020)
24. S. Furusawa, H. Ochiai, K. Murayama, Key Eng. Mater. **497**, 26 (2012)
25. X. Li, Q. Xiao, B. Liu, H. Lin, J. Zhao, J. Power Sour. **273**, 128 (2015)
26. R. Saroha, A.K. Panwar, J. Phys. D: Appl. Phys. **50**, 255501 (2017)
27. R. Saroha, A.K. Panwar, Y. Sharma, P.K. Tyagi, S. Ghosh, Appl. Surf. Sci. **394**, 25 (2017)
28. C. Chen, C. Ai, X. Liu, Electrochim. Acta **265**, 448 (2018)
29. Q. Zhang, C. Zhang, B. Li, S. Kang, X. Li, Y. Wang, Electrochim. Acta. **98**, 146 (2013)
30. S. Liu, L. Xiong, C. He, J. Power Sources. **261**, 285 (2014)
31. H. Tang, Y. Zhou, L. Zan, N. Zhao, Z. Tang, Electrochim. Acta. **191**, 887 (2016)
32. L. Wang, Z. Han, Q. Weng, T. Liu, Z. Tang, H. Tang, Ionics. **28**, 669 (2022)

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

Check for updates

View Article Online
View Journal

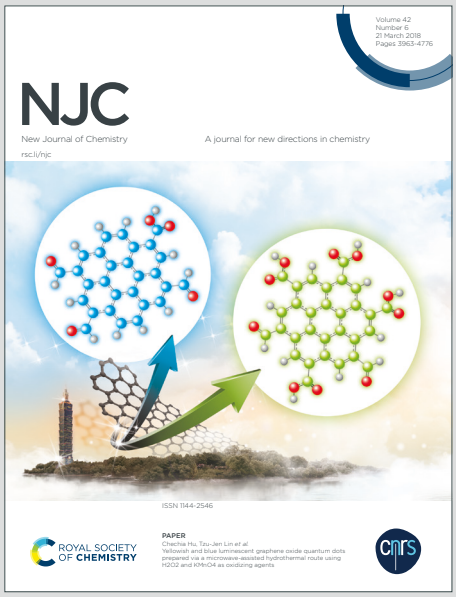
NJC

New Journal of Chemistry

A journal for new directions in chemistry

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: R. Khatri and N. K. Puri, *New J. Chem.*, 2022, DOI: 10.1039/D2NJ00540A.



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

View Article Online
DOI: 10.1039/D2NJ00540A

Electrochemical Studies of Biofunctionalized MoS_2 Matrix for Highly Stable Immobilization of antibodies and Detection of Lung Cancer Protein Biomarker

Ritika Khatri[1], Nitin K. Puri[1*]

[1]Nanomaterials Research Laboratory, Department of Applied Physics, Delhi Technological University, Delhi, India

[*]Corresponding author email id: nitinkumarpuri@dtu.ac.in

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

Abstract

To address the issue of lack of stable immobilization of antibodies on the biosensing matrix for repeated cycles of measurement, chitosan (CS) bio-functionalized MoS_2 is prepared to serve as a biosensing matrix. Electrochemical performance of CS/MoS_2 matrix towards detection of neuron specific enolase (NSE), a lung cancer biomarker is also investigated. Unlike other complex matrices involving various steps of modification, the matrix studied herein involves only 2 steps of modification without the use of any label, amplifying its appeal for biosensing applications. The fabricated immunoelectrode is found to exhibit remarkable cyclic stability, with a sensitivity of 3.5 μA /ng/mL/cm^2 and a wide linear detection range of 0.1 to 100 ng/mL. Also, the sensor is found to be fast, specific, reproducible, regenerable up to 4 cycles, and has a shelf life of 6 weeks. The stability study of fabricated immunoelectrode revealed that the cumulative effect of CS and MoS_2 ensures stable immobilization of biomolecules on the immunoelectrode for 50 cyclic voltammetry (CV) cycles. The results of this study suggest that as proposed matrix will be promising in the fabrication of devices for early monitoring of protein biomarkers.

Keywords: MoS_2, Chitosan, Electrochemistry, Biosensor, NSE, Stability

View Article Online
DOI: 10.1039/D2NJ00540A

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

1. Introduction

Electrochemical biosensors have gained momentum in the development of point-of-care devices owing to the ease of miniaturization, requirement of small sample volumes, rapid response, and low requirement for power [1, 2]. They use electron mediators to translate the events of analyte binding into a detectable electrical signal. The preparation of an electrochemical biosensor requires the fabrication of a suitable immobilization matrix that can hold the bio-recognition elements efficiently and effectively [3]. Nanostructured materials have been proven a good choice to fabricate the platform for loading biological molecules to detect even very low levels of analytes [4]. However, the chosen material should be able to escalate the charge transfer process without passivating the electrode surface to maintain its performance along with increasing its effective surface area. For the fabrication of an electrochemical biosensor, it is inevitably important to ensure that the bioactivity of the immobilized biomolecule is not lost on the continuous application of potential on it and with repeated interaction with redox media.

In recent times, molybdenum disulfide (MoS_2) has gained great attention in electrochemical applications (i.e., supercapacitors, lithium (Li)/sodium (Na) ion batteries, hydrogen evolution reaction (HER), sensors, etc [5]) as it possesses many intriguing properties like layered structure, various routes for synthesis, ease of tailoring the morphology, large surface to volume ratio, ease of dispersibility in aqueous solution, fast rate of heterogeneous rate transfer (HET), the feasibility of surface functionalization and modification and exceptional cyclic stability [5-9]. Owing to its hydrophobicity, MoS_2 can be utilized to fabricate the immobilization platform because it provides a strong affinity for the adsorption of protein on its surface [6]. Because of its layered structure and high surface/volume ratio, it enhances the total surface area for the immobilization of biomolecules. Owing to all these fascinating properties of MoS_2, it is used to fabricate the matrix for electrochemical biosensing. Further,

View Article Online
DOI: 10.1039/D2NJ00540A

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

to anchor functional groups on MoS_2 and impart biocompatibility to the matrix, Chitosan (CS) is used. CS is a non-toxic biocompatible polymer, which is present in abundance, has excellent film-forming and stabilizing properties along with polycationic nature [10, 11]. It is rich in functional groups like amino and hydroxyl, on its polysaccharide chain [12]. Thus, CS is blended with MoS_2 in the presence of ionic liquid (IL) because of its high polarity [13]**.** Due to its polar nature, IL provides a reaction atmosphere for MoS_2 and CS to interact with each other. The benefit of compositing different materials is governed by the fact that both materials contribute to the properties of the final blend [1].

In this work, the sensing behavior of bio-functionalized MoS_2 is examined by choosing protein biomarker neuron-specific enolase (NSE) as the target analyte. NSE is considered to be a specific protein biomarker for the diagnosis of small cell lung cancer (SCLC) and is found to be > 9 ng/mL in the case of 70% of patients diagnosed with SCLC [14]. Quite a few reports are available on the detection of NSE, using different matrices with outstanding performances [15-18]. However, achieving a high signal response along with maintaining the signal stability and the bioactivity of the platform, is still a significant challenge [15]. For instance, Tang et al [16] developed PPy-PEDOT-Au-GCE as a substrate material for stable immobilization of immune molecules along with the maintenance of their bioactivity for detection of NSE. They found that their platform could provide a stable environment for sensing till 30 CV cycles. Yu and co-workers designed an electrochemical immunosensor for the detection of NSE based on graphene supported by hollow carbon balls [17]. A stable trend for this immobilization matrix was found for 30 cycles of CV. Mo et al fabricated an immunosensor for simultaneous detection of NSE and CYFRA21-1 and found the stability of the immunosensor for 15 CV cycles [18]. Thus, it has been observed that after a few cycles of measurements the signal response tends to degrade. In this investigation, we explored the CS/MoS_2 for stable immobilization of anti-NSE antibodies and the detection of NSE for the first time, to the best of our knowledge. The

View Article Online
DOI: 10.1039/D2NJ00540A

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

CS/MoS_2 based matrix proved to be a stable matrix, with nearly 94% of the initial signal response still retained after continuous scanning for 50 CV cycles. The bioactivity of immobilized antibodies is ensured by the biocompatible CS, while a stable cyclic current response is kept in check by MoS_2.

The method of synthesis and fabrication of immunoelectrode discussed in this work is appealing because of its large yield, scalability, and simplicity without involving various steps of modifications which ensures remarkable electrochemical performance and stability. There are three major novel features of our work. First, a highly stable matrix is formed favorable for immobilization of biomolecules during repeated cycles of measurement. Secondly, the platform is functionalized for the attachment of biomolecules without any harsh chemicals treatment. Third, the large electroactive surface (55% of the geometric area (1×1 cm^2)), makes it a remarkable platform for electrochemical sensors. Our sensor is found to be very fast, reproducible, specific, stable, regenerable, with a wide linear detection range and remarkable shelf life. Considering the concentration of NSE found in the serum of SCLC patients, the sensitivity and lowest limit which could be detected by our sensor, is expected to meet the requirement of detection of NSE in clinical samples too.

2. Experimental Details

2.1 Chemicals and Reagents Used for Synthesis, Cleaning, and Biosensing

The monoclonal anti-NSE antibody, neuron-specific enolase (NSE) antigen, N-(3-Dimethylaminopropyl)-N'-ethyl carbodiimide hydrochloride (EDC), N-Hydroxysuccinimide (NHS), bovine serum albumin (BSA), ammonium molybdate tetrahydrate $(NH_4)_6Mo_7O_{24}.4H_2O$, and acetic acid (CH_3COOH) are purchased from Sigma-Aldrich. Thiourea (CH_4N_2S) is bought from SRL. Liquor ammonia (NH_4OH) and N, N-dimethylformamide (DMF) are bought from Fisher Scientific. Chitosan is purchased from HIMEDIA. 1-butyl-3-methyldiazolium hexafluorophosphate ($[C_4mim][PF_6]$) is procured from

View Article Online
DOI: 10.1039/D2NJ00540A

GLR innovations. Other Chemicals like hydrazine hydrate (N_2H_4), Sodium chloride (NaCl), potassium bromide (KBr) are procured from Thermo Fisher Scientific. Hydrogen peroxide (H_2O_2) is purchased from qualigens. Concentrated hydrochloric acid (HCl) is bought from RFCL limited. Chemicals including acetonitrile, potassium hexacyanoferrate(II) trihydrate/ potassium ferrocyanide ($K_4Fe(CN)_6.3H_2O$), potassium hexacyanoferrate (III)/ potassium ferricyanide ($K_3Fe(CN)_6$), disodium hydrogen diphosphate dihydrate ($Na_2HPO_4.2H_2O$), Sodium dihydrogen phosphate dihydrate ($NaH_2PO_4.2H_2O$), iso-propyl alcohol (IPA) and ethanol are procured from Merck. A Millipore Milli-Q system (Bedford, MA, USA) is used to obtain milli-Q water for the preparation of the solutions which are used for various experiments. Tris hydrochloride (Tris-HCl), magnesium sulfate ($MgSO_4$), and potassium chloride (KCl) which serve as a buffer for dilution of NSE are purchased from lobachemie (Tris- HCl) and Thermo Fisher Scientific ($MgSO_4$ and KCl).

Before usage and storing the buffer solutions, all the glassware is autoclaved and cleaned using 70% isopropyl alcohol (IPA). All the aqueous solutions are prepared in autoclaved de-ionized (dI) water.

2.2 Characterization Equipment's and Facilities

To record the x-ray diffraction (XRD) pattern of samples, a Bruker D8 Advance facility is used. Raman spectrum for the samples is recorded at a wavelength of 532 nm using a WITec Raman spectrometer. Scanning electron microscopy (SEM) is done using Jeol Japan mode. The Fourier Transform Infrared (FTIR) spectroscopy measurements are done using Perkin Elmer spectrum two. The thermal stability of samples is recorded using a Perkin Elmer 4000 thermogravimetric analyzer. To carry out the thickness measurement an Ambios XP-200 stylus profiler is used. A Metrohm Potentiostat/Galvanostat Autolab along with NOVA software is used for recording all the electrochemical measurements.

2.3 Synthesis of CS/MoS_2

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

View Article Online
DOI: 10.1039/D2NJ00540A

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

MoS_2 nanostructures constituting a large number of layers are synthesized using our previously reported work [19]. To synthesize CS/MoS_2, an ionic liquid (IL) assisted mechanical grinding method is used. The synthesis of CS/MoS_2 is done according to the method adopted by Zhang et al [20] with slight modifications for scaling up the quantity of the finally obtained product. For this, 500 mg of the above synthesized MoS_2 powder is mixed with 200 mg of CS. The mixture is ground in an agate mortar pestle for 20 minutes. Next, a gel is formed by dropping 1 mL of IL (1-butyl-3-methyldiazolium hexafluorophosphate, [C_4mim][PF_6]) in the mortar pestle. This mixture of MoS_2, CS, and IL is further grounded for 60 minutes. Then the obtained product is washed using acetone, acetic acid, and N,N-dimethylformamide (DMF) to remove excess CS and IL. Then the obtained residue is put in an oven at 100℃ for 24 hours for desolvation. Thus, dried powder of CS/MoS_2 is collected in an Eppendorf for further use.

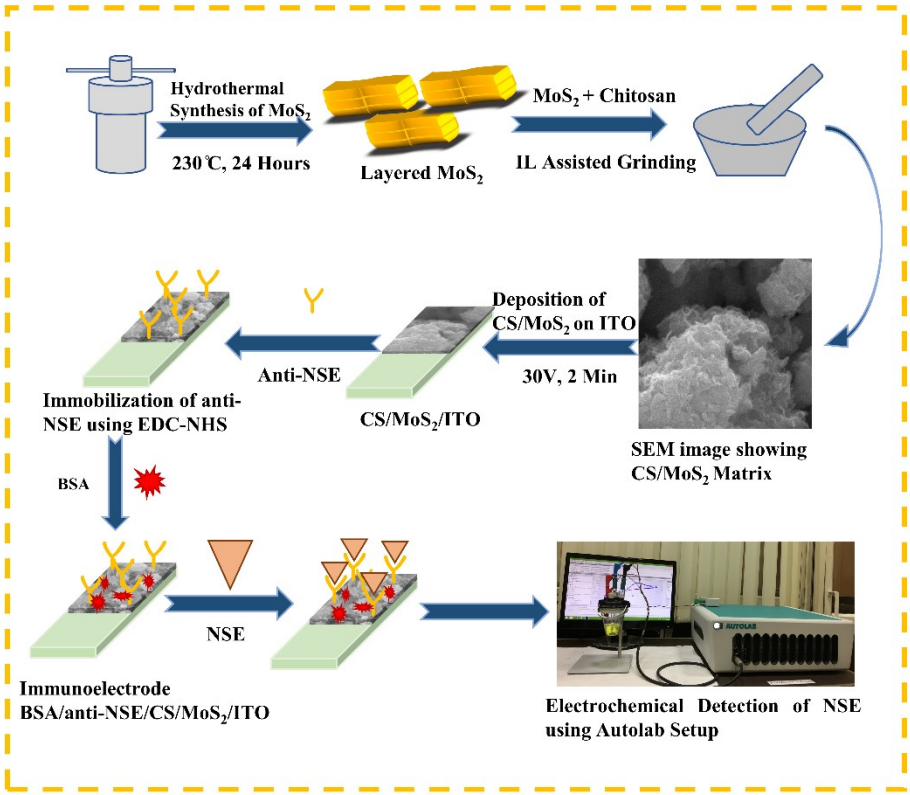
2.4 Fabrication of Immunoelectrode

To fabricate the immunoelectrode BSA/anti-NSE/CS/MoS_2/ITO, a series of steps are followed. First of all, ITO cut in a dimension of 3 × 1 cm^2 are hydrolyzed using NH_3:H_2O_2:H_2O in the ratio 1:1:5 at 80℃ for 1 hour. The hydrolyzed ITO is cleaned using water and IPA. Next, films of CS/MoS_2 are deposited on hydrolyzed ITO using electrophoretic deposition (EPD). For this CS/MoS_2 is well dispersed in acetonitrile in a concentration of 1 mg/mL. This solution is taken in an EPD cell containing a copper (Cu) wire and platinum counter electrode (CE). Hydrolyzed ITO is clipped at Cu as a working electrode (WE) and a potential of 30 V is applied for 120 seconds. CS/MoS_2 film is formed on a geometric area of 1 × 1 cm^2. The film is left for overnight air drying at room temperature. The thickness is measured at various points of the CS/MoS_2/ITO electrode and it is found to be varying in the range of 1.1 to 3 µm indicating a non-uniform surface. For the immobilization of anti-NSE on CS/MoS_2/ITO, 0.4 M EDC (used as a coupling agent), 0.1 M NHS (used as an intermediate stabilizer), and 25 µg/mL anti-NSE are aliquoted together in a ratio of 1:1:2. Then, 20 µL of this aliquoted solution is spread

View Article Online
DOI: 10.1039/D2NJ00540A

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

uniformly on CS/MoS_2/ITO electrode. The anti-NSE/CS/MoS_2/ITO electrode is kept in a humid chamber for an optimized time duration of 5 hours for proper attachment of anti-NSE on CS/MoS_2/ITO. Then, the anti-NSE/CS/MoS_2/ITO is washed in pBS 7.4 to remove unbound antibodies from the electrode surface. To mask the non-specific binding sites on anti-NSE/CS/MoS_2/ITO, 10 µL of 0.1 mg/mL of BSA is dropped cast to make the final immunoelectrode BSA/anti-NSE/CS/MoS_2/ITO. After 1 hour the immunoelectrode is washed in pBS 7.4 to remove excess BSA from the surface of the immunoelectrode. The BSA/anti-NSE/CS/MoS_2/ITO immunoelectrode is finally prepared and kept at 4℃ when not in use. The illustration of the synthesis of the CS/MoS_2 matrix and fabrication of BSA/anti-NSE/CS/MoS_2/ITO immunoelectrode for detection of NSE is shown in **Scheme 1**.



Scheme 1: Schematic diagram illustrating the synthesis of CS/MoS_2 matrix and fabrication of BSA/anti-NSE/CS/MoS_2/ITO immunoelectrode for detection of NSE.

3. Results and Discussion

View Article Online
DOI: 10.1039/D2NJ00540A

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

3.1 Crystallographic Study of CS/MoS_2 Matrix Using X-ray Diffraction (XRD)

The primary confirmation of the formation of the CS/MoS_2 matrix and the study of its crystallographic structure is done using XRD (**Fig. 1)**. The XRD spectra obtained for commercially purchased CS shows a broad diffraction peak around 20˚, without the presence of any additional impurity peaks. For comparison, the XRD of MoS_2 is also shown and both MoS_2 and CS/MoS_2, XRD spectra show the peaks corresponding to the planes (002), (100), (103), (110) of 2H-MoS_2 (JCPDS card no. – 37-1492) [19]. Besides these peaks, a halo around 20˚ in the spectrum of CS/MoS_2 confirms the formation of the composite of CS and MoS_2.

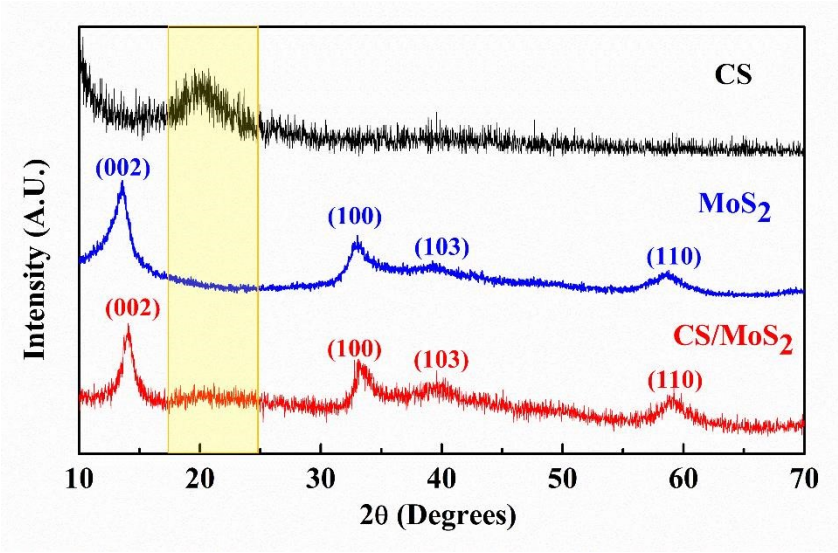


Fig. 1: XRD spectra of CS, MoS_2 and, CS/MoS_2

3.2 Study of Vibrational Modes and Qualitative Information About Layered Structure and Functionalization Using Raman Spectroscopy

Fig. 2(a) shows the Raman spectra obtained for CS as well as CS/MoS_2 in the range of 250 cm^{-1} to 3000 cm^{-1}. The appearance of the A_{1g} and E^1_{2g} vibrational modes of MoS_2 in the Raman spectrum of CS/MoS_2, **(shown in Fig. 2(a) and 2(b))** confirm the retention of the structure of MoS_2 after the addition of CS. The A_{1g} vibrational mode corresponding to monolayer MoS_2 is found to be at 402.72 cm^{-1} **(Fig. 2(b))** [21]. However, the E^1_{2g} mode is found to be red-shifted to 375.39 cm^{-1} leading to a wavenumber separation between the two modes to be 27.3 cm^{-1} (**Fig.**

View Article Online
DOI: 10.1039/D2NJ00540A

2(b)). An increase in wavenumber separation may be attributed to the stacking of layers of MoS_2 due to the incorporation of CS in the structure of MoS_2. Further, the characteristic peaks of CS are observed at 893.14 cm^{-1}, 1096 cm^{-1}, 1372.67 cm^{-1}, and 2885.37 cm^{-1} corresponding to the C-H bond of polysaccharide in CS, with OH peak at 3311.37 cm^{-1} and 3376.83 cm^{-1} (**Fig. 2(a)**) [22]. The confirmation of functionalization of MoS_2 by CS can be drawn from the Raman spectrum of CS/MoS_2 **(Fig. 2(a))**, which shows a peak at 1368.61 cm^{-1} corresponding to the δCH of the polysaccharide chain of CS. Further, a halo is observed in the range 2500-3260 cm^{-1}, which could be due to the υCH_2 peak of CS. This study confirms the successful incorporation of CS into the structure of MoS_2.

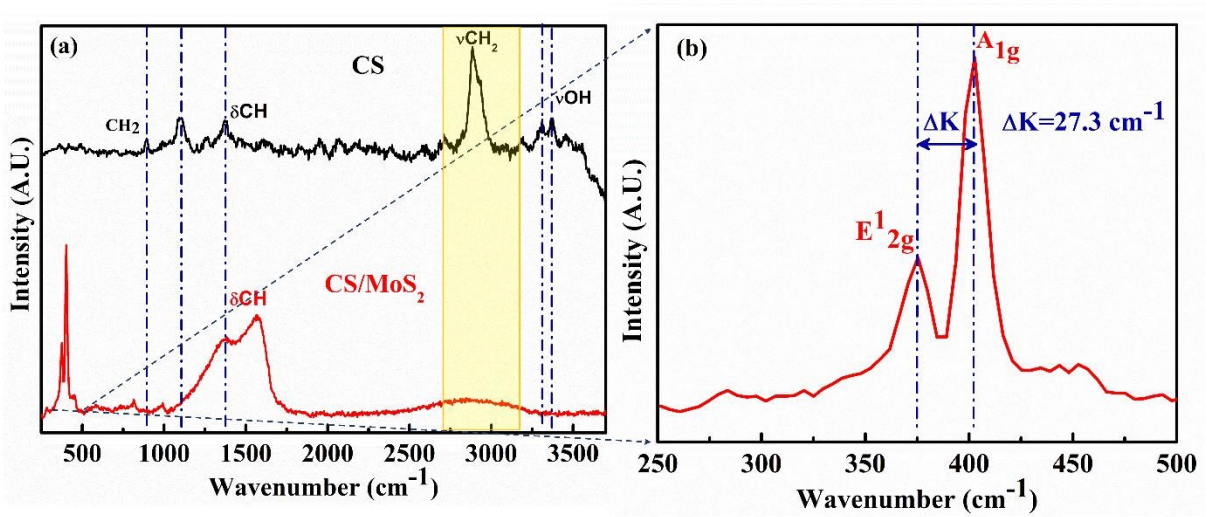


Fig. 2: (a) Raman spectra of CS and CS/MoS_2 **(b)** Raman spectrum of CS/MoS_2 in the range 250 to 500 cm^{-1} showing A_{1g} and E^1_{2g} modes of MoS_2 in CS/MoS_2

3.3 Thermogravimetric Analysis (TGA) of CS/MoS_2 Matrix

Fig. 3 shows the thermal behaviour of CS, MoS_2, and CS/MoS_2 from 30 ˚C to 850 ˚C. An initial weight loss at around 100 ˚C in all the samples could be attributed to moisture loss. The weight loss for CS starts at 280 ˚C while CS/MoS_2 starts losing some weight around 220 ˚C, which could be due to some leftovers from the process of synthesis of MoS_2 [23]. The weight loss in the case of CS and CS/MoS_2 is more in range (280 ˚C to 500 ˚C), which can be ascribed to the

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

View Article Online
DOI: 10.1039/D2NJ00540A

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

decomposition of carbon skeleton at such high temperatures [24]. After 500 ˚C, 57% and 32% weight of MoS_2 and CS/MoS_2 are retained respectively, while CS is completely lost. Thus, the matrix CS/MoS_2 is found to have more thermal stability in comparison to CS. This study also implies that the incorporation of functional groups in MoS_2 reduces its thermal stability at high temperatures. But, as long as application to biosensors is concerned, this is not worrisome because biological testings are always done at room temperatures or even lower temperatures. Further, the successful incorporation of CS in MoS_2 could be derived from TGA.

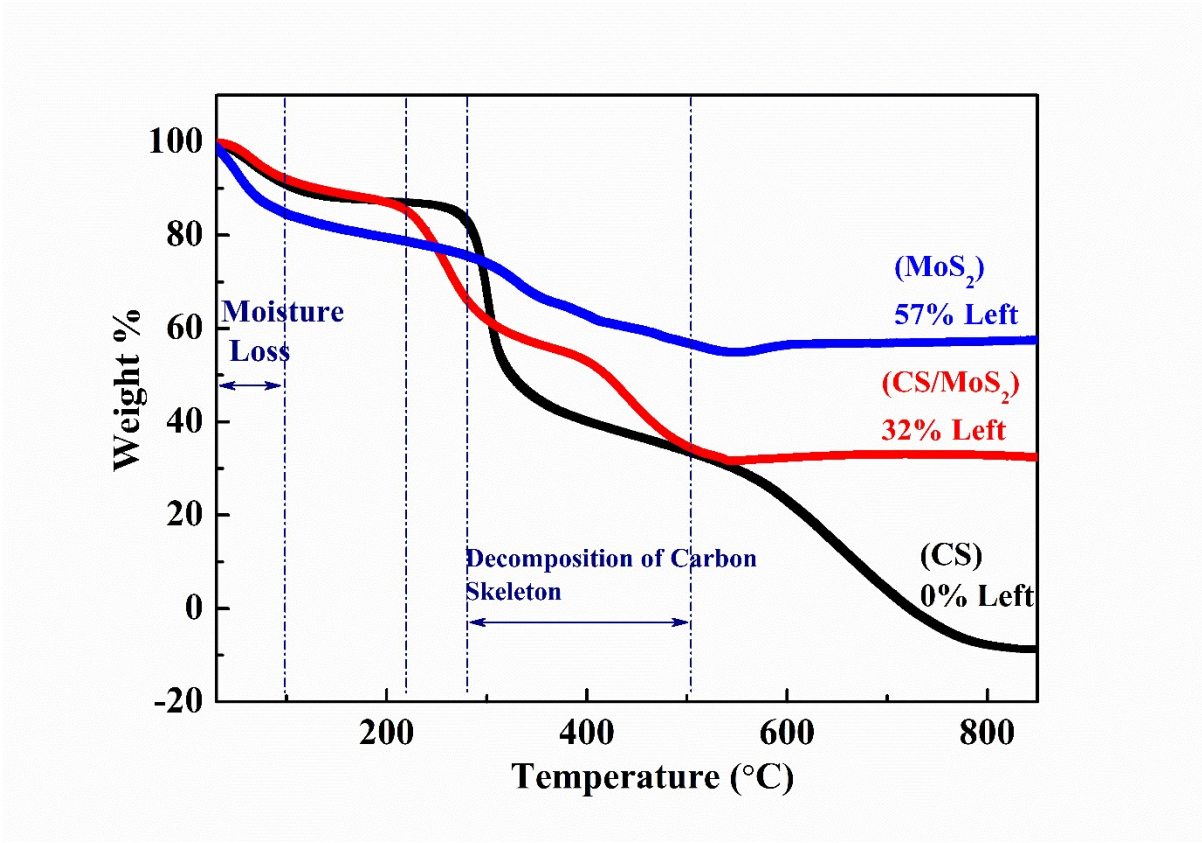


Fig. 3: TGA curves of CS, MoS_2, and CS/MoS_2

3.4 Study of Functional Groups Attached on CS/MoS_2 Matrix Using Fourier Transform Infrared Spectroscopy (FTIR)

The confirmation of successful functionalization of MoS_2 using CS is done using FTIR (**Fig. 4**). Peaks observed in CS at 1416 cm^{-1}, 1303 cm^{-1}, and 1089 cm^{-1} correspond to CH_2 bending, -NH stretching and C-O stretch of CS respectively [25]. Peak observed at 1561 cm^{-1} in CS is due

View Article Online
DOI: 10.1039/D2NJ00540A

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

to bending vibration of -NH [23]. These peaks corresponding to -NH stretching (1303 cm^{-1}) and -NH bending (1561 cm^{-1}) are missing in CS/MoS_2. The disappearance of these -NH peaks in the FTIR spectrum of CS/MoS_2 could be attributed to the polycationic nature of CS which leads to the protonation of the amine group during grinding in IL, which leads to electrostatic interaction between MoS_2 layers and CS [24]. Peaks observed in CS at 1262 cm^{-1} and 1162 cm^{-1} correspond to C-N and C-O-C bonds in CS [26]. The peak observed in the range 3440-3480 cm^{-1} in both CS and CS/MoS_2 corresponds to the hydroxyl group (OH) [24]. The peak observed in both CS and CS/MoS_2 at 1371 cm^{-1} corresponds to wagging CH_3 [27]. Further, a peak corresponding to NH_2 is observed in both CS and CS/MoS_2 at 895 cm^{-1} [28] and 1625 cm^{-1} [29], indicating functionalization of CS/MoS_2 with the amino group. The peak observed at 618 cm^{-1} corresponds to Mo-S vibration [30].

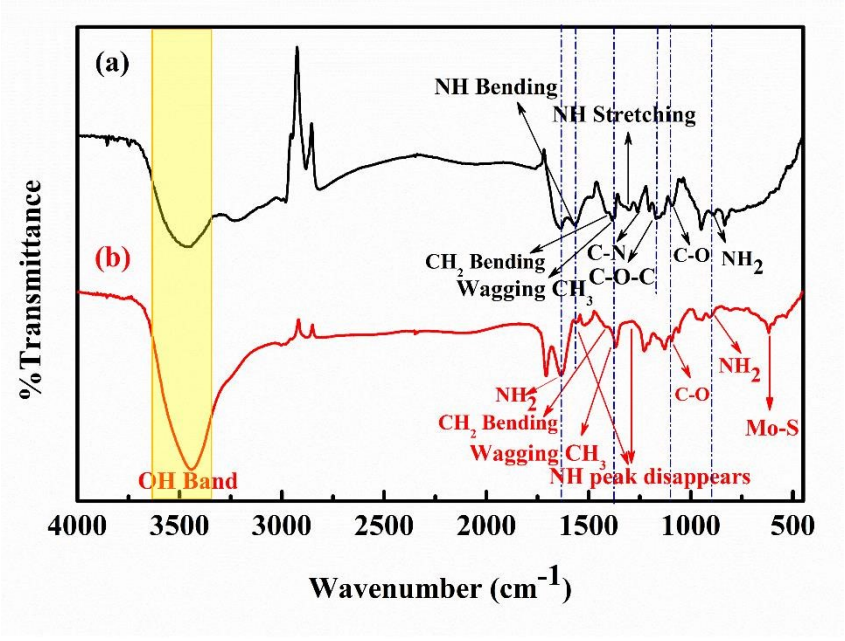


Fig. 4: FTIR curves of (a) CS and (b) CS/MoS_2

3.5 Study of Surface Morphology of CS/MoS_2 Matrix Using Scanning Electron Microscopy (SEM)

Fig. 5(a) shows nanostructures of MoS_2 of varying sizes. Larger magnification of these nanostructures can be seen constituting a large number of layers of MoS_2 in **Fig. 5(b)**. **Fig. 5(c)**

View Article Online
DOI: 10.1039/D2NJ00540A

shows a large continuous membrane of CS. The evidence of surface modification of MoS_2 nanostructures by CS is given in **Fig. 5(d)**, which shows that layers of MoS_2 are stacked together and encapsulated by CS membrane rendering a rough and uneven surface to the composite.

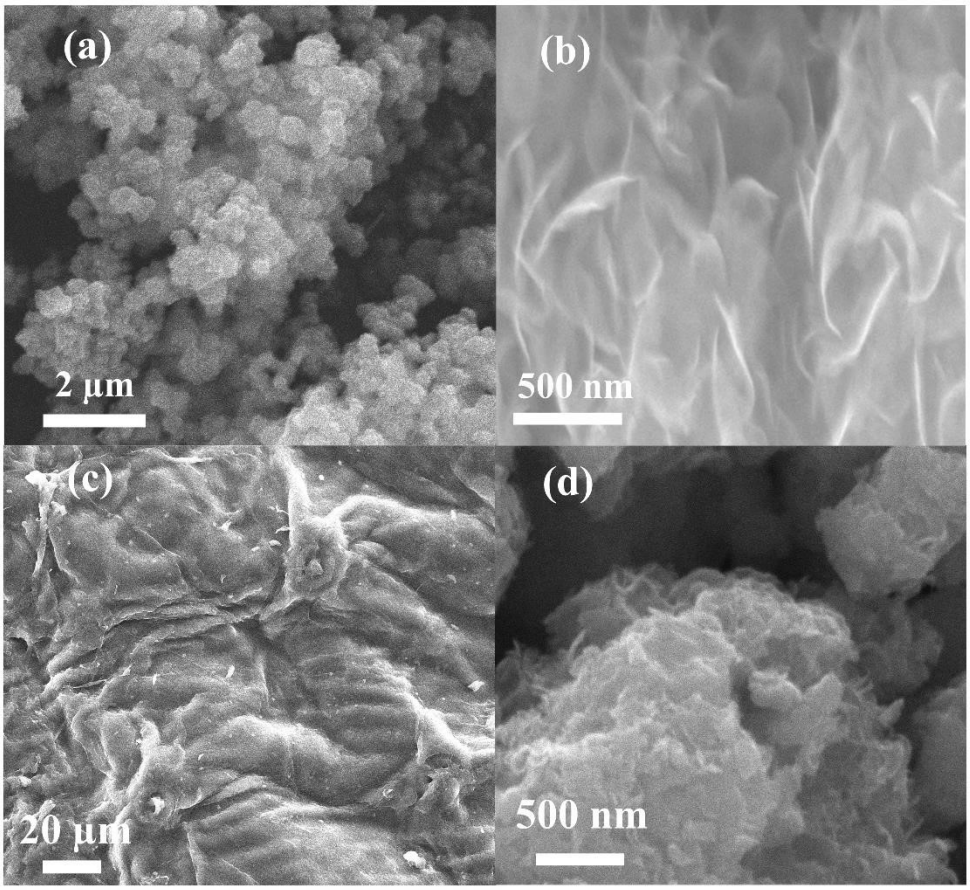


Fig. 5: SEM micrographs of (a, b) MoS_2 (c) CS (d) CS/MoS_2

4 Detailed Investigation of Electrochemical Performance of Immunoelectrode Prepared using CS/MoS_2 Matrix

4.1 Optimization of Biosensing Parameters

The concentration of anti-NSE to be attached to CS/MoS_2/ITO electrode is optimized by attaching varying concentrations of anti-NSE using EDC-NHS chemistry. **Fig. 6(a)** shows the current response along with varying concentration of anti-NSE (5 µg/mL, 10 µg/mL, 25 µg/mL, 50 µg/mL, 100 µg/mL). The response is found to be decreasing initially till 25 µg/mL, after which it becomes

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

View Article Online
DOI: 10.1039/D2NJ00540A

constant. Thus, 25 µg/mL is finalized as the concentration for attachment of anti-NSE throughout the studies.

The maximum time required for immunoreaction to take place between NSE and anti-NSE is found by incubation of BSA/anti-NSE/CS/MoS_2/ITO immunoelectrode with NSE for varying times (**Fig. 6(b)**). 20 µL of 5 ng/mL of NSE is incubated on immunoelectrode and the CV response is recorded by varying incubation times from 0, 5, 10, 15, 20, 25 minutes. Initially, the current decreases till 15 minutes, after which the current becomes nearly constant. Thus, the time required for immunoreaction to take place is found to be 15 minutes and is used for all the further studies.

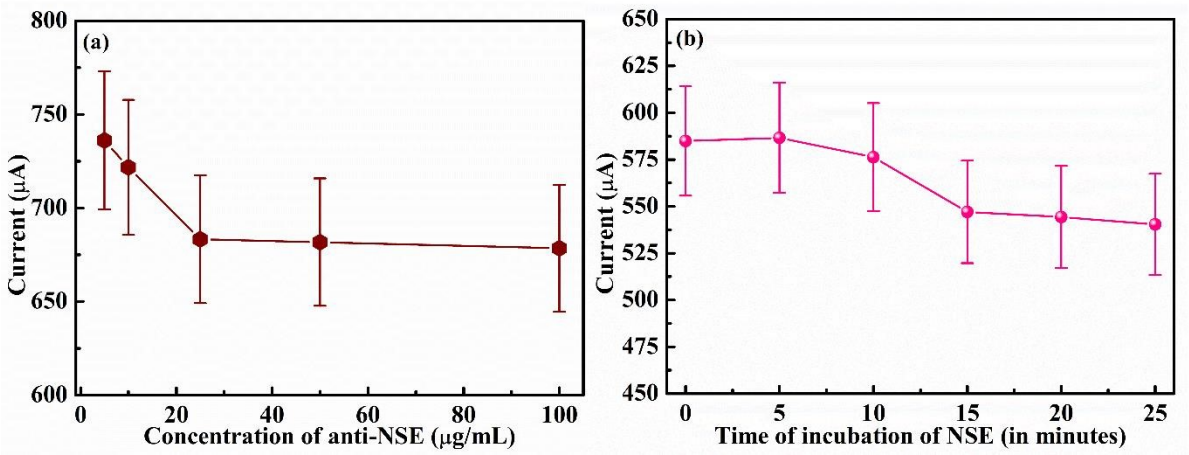


Fig. 6: (a) Current response with varying concentration of anti-NSE **(b)** Current response with variation in incubation time of NSE on BSA/anti-NSE/CS/MoS_2/ITO immunoelectrode

4.2 Electrochemical Study of the Various Steps of Fabrication of Immunoelectrode

Various steps of modification of the immunoelectrode are evaluated by recording the CV profiles at each step, in PBS 7.4 at a scan rate of 50 mV/s (**Fig. 7**). A distinct anodic and cathodic current peak (i_{pa} and i_{pc} respectively) is observed in each step. The maximum value of i_{pa} and i_{pc} is recorded to be 640.26 µA and -810.55 µA for anti-NSE/CS/MoS_2/ITO electrode followed by BSA/anti-NSE/CS/MoS_2/ITO immunoelectrode having the values as 579.35 µA and -762.63 µA. Further, the minimum current values of i_{pa} and i_{pc} are found to be 436.71 µA and -579.53 µA for CS/MoS_2/ITO. The variation in peak current values at each step confirms the modification of the electrode. The

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

View Article Online
DOI: 10.1039/D2NJ00540A

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

sequential modification at each step of fabrication of immunolelectrode is also confirmed by electrochemical impedance spectroscopy (EIS). EIS studies are done at a biasing potential of 0.01 V in the frequency range of 0.1 Hz to 100 kHz. Fig. 7(b) shows the Nyquist plots for (i) CS/MoS_2/ITO, (ii) anti-NSE/CS/MoS_2/ITO and (iii) BSA/anti-NSE/CS/MoS_2/ITO. The results are found to complement CV results with a minimum value of charge transfer resistance (R_{ct}) observed as 161.78 Ω for anti-NSE/CS/MoS_2/ITO. This is followed by an increase in R_{ct} value to 245.40 Ω for BSA/anti-NSE/CS/MoS_2/ITO and the maximum value of R_{ct} is found to be 313.38 Ω for CS/MoS_2/ITO.

An increase in the value of i_{pa} and decrease in the value of R_{ct} is observed when anti-NSE is immobilized on CS/MoS_2/ITO. This happens because of the interaction between anti-NSE and CS/MoS_2 and the spatial orientation of antibodies which provides conduction pathways to the charge carriers on the interface [31, 32]. Further, a decrease in current and increase in R_{ct} is found after attachment of BSA onto anti-NSE/CS/MoS_2/ITO, due to masking of non-specific sites for attachment, by this blocking agent.

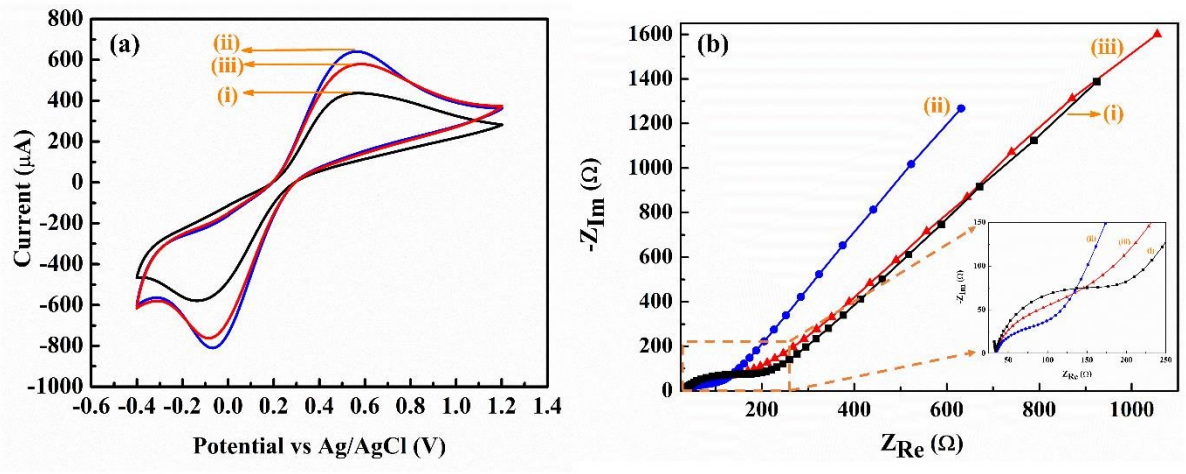


Fig. 7 (a): CV curves of (i) CS/MoS_2/ITO (ii) anti-NSE/CS/MoS_2/ITO (iii) BSA/anti-NSE/CS/MoS_2/ITO **(b)** EIS curves of (i) CS/MoS_2/ITO (ii) anti-NSE/CS/MoS_2/ITO (iii) BSA/anti-NSE/CS/MoS_2/ITO. The inset shows the magnified Nyquist plots (in the range of 25 to 250 Ω)

View Article Online
DOI: 10.1039/D2NJ00540A

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

Fig. 8(a) displays the CV response of the prepared immuno-electrode BSA/anti-NSE/CS/MoS_2/ITO with varying scan rates. The CV profile is recorded in phosphate buffer solution (PBS) of pH 7.4 having 50 mM of $[Fe(CN)6]^{3-/4-}$. An increase in i_{pa} and i_{pc} currents is observed for increasing scan rates. Further, a linear variation of i_{pa} and i_{pc} with $\sqrt{Scan\ Rate}$ **(Fig. 8(b))** shows that the process of reduction and oxidation of $[Fe(CN)_6]^{3-/4-}$ at the immuno-electrode is diffusion-controlled [33, 34], which indicates that adsorbents when gets conjugated on the electrode will inhibit the diffusion of $[Fe(CN)_6]^{3-/4-}$. The linearly fitted plots as shown in **Fig. 8(b)** are given by the following equations:

$$i_{pa} = [(60.21\mu A X \sqrt{\frac{s}{mV}}) X \sqrt{Scan\ Rate(\frac{mV}{s})}] + 119.61\ (\mu A) \ldots\ldots\ldots\ldots\ldots\ldots \textbf{(1)}$$

$R^2 = 99.92$

$$i_{pc} = [(-91.84\mu A X \sqrt{\frac{s}{mV}}) X \sqrt{Scan\ Rate(\frac{mV}{s})}] - 81.91\ (\mu A) \ldots\ldots\ldots\ldots\ldots\ldots \textbf{(2)}$$

$R^2 = 98.83$

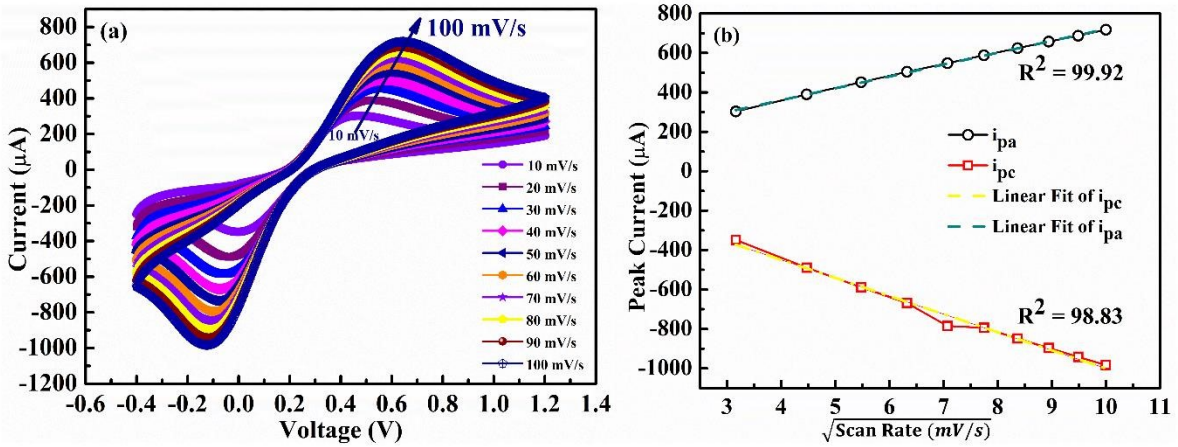


Fig. 8: (a) Scan Rate Study of immuno-electrode BSA/anti-NSE/CS/MoS_2/ITO by varying scan rate from 10 to 100 mV/s and recording CV profiles **(b)** Variation of peak anodic current (i_{pa}) and peak cathodic current (i_{pc}) with under root of scan rate.

View Article Online
DOI: 10.1039/D2NJ00540A

The evaluation of electrochemical kinetics parameters of BSA/anti-NSE/CS/MoS_2/ITO like electroactive surface area (A_e), HET rate transfer constant (K_o) is done as follows. The electroactive surface area (A_e) is determined using the Randles Sevcik equation [19, 35]

$$i_p = 2.69\ X10^5 n^{3/2} A_e C\sqrt{D_o v} \quad \text{.....(3)}$$

Where i_p is peak anodic current as obtained in CV profile of the electrode (in A), n is the number of electrons transferred during redox reaction (n=1, for ferro/ferri), A_e is electroactive surface area, C is the concentration of redox probe (C = 5 mM= 5 mol/cm^3), Do is the diffusion coefficient (D_o = 0.667 X 10^{-5} cm^2/s [19, 35]), v is scan rate. The unit of the constant $2.69\ X10^5$ is $\frac{C}{mol\sqrt{V}}$ [36]. From the plot between i_{pa} and $v^{1/2}$ (Fig. 8(b)) the value of the slope is obtained as 60.21 μA$\sqrt{\frac{s}{mV}}$ (from equation (1)). By substituting all the above values in equation (3) the electroactive surface area is determined to be 0.55 cm^2 for BSA/anti-NSE/CS/MoS_2/ITO. A high value of electroactive surface area allows the sensor to accommodate large concentrations of analyte, in turn increasing the linear detection range of the device.

The value of electron transfer constant (k^o) is determined using Nicholson's Equation [35], which is given as

$$\Psi = k^o\left(\frac{D_o}{D_R}\right)^{\frac{\alpha}{2}}\left(\frac{RT}{\pi n F D_o v}\right)^{\frac{1}{2}} \text{.....(4)}$$

Where,

D_o and D_R are the diffusion coefficient of oxidized and reduced electroactive species, which is approximately equal to 1 for symmetrical redox reaction, α is the transfer coefficient, R is the universal gas constant (8.314 JK^{-1}mol^{-1}), F is Faraday constant (96485 C/mol), v is scan rate (50 mV/s), remaining parameters are same as equation **(3)** and Ψ is a dimensionless kinetic parameter which can be calculated using ΔE_p [35] as follows

$$\Psi = \frac{-0.6288+0.0021X\Delta Ep}{1-0.017X\Delta Ep} \text{.....(5)}$$

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

View Article Online
DOI: 10.1039/D2NJ00540A

Using, equation **(4)** and **(5)**, the value of k^o is calculated to be -4.05X10^{-3} cm/s for BSA/anti-NSE/CS/MoS_2/ITO.

4.3 Stability Study of Prepared Immuno-electrode

The prepared immunoelectrode BSA/anti-NSE/CS/MoS_2/ITO is checked for the stability of immobilized antibodies on the continuous application of potential and repeated interaction with redox media ($[Fe(CN)_6]^{3-/4-}$) by carrying out CV for 50 consecutive cycles. The scans are done at a scan rate of 50 mV/s in the potential window of -0.4 V to 1.2 V in $[Fe(CN)_6]^{3-/4-}$. An essentially unchanged value of anodic and cathodic peak current is observed as shown in **Fig. 9(a).** 95% and 93.66% of initial i_{pa} are retained, after 15 and 50 CV cycles respectively while 99.45% of initial i_{pc} is retained after 50 CV cycles as shown in **Fig. 9(b).** Retention of a high percentage of initial current even after 50 CV cycles demonstrate that our immuno-electrode has high stability. MoS_2 ensures the cycling stability while CS in CS/MoS_2 imparts biocompatibility to the matrix, which helps the biomolecules to remain stably immobilized, after interaction with the redox media in continuous cycles of CV.

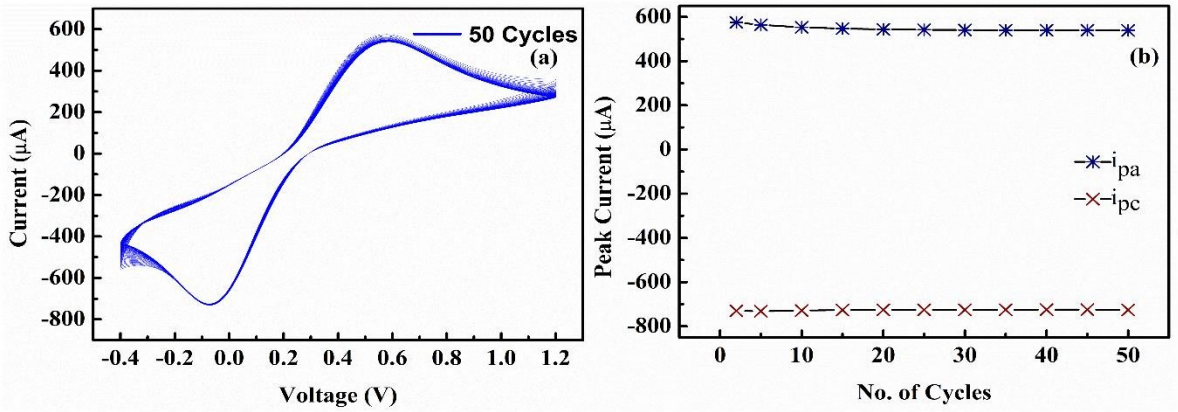


Fig. 9 (a) Consecutive scans of immunoelectrode (BSA/anti-NSE/CS/MoS_2/ITO) for 50 CV cycles **(b)** Variation of peak anodic (i_{pa}) and peak cathodic current (i_{pc}) with no. of CV cycles.

4.4 Electrochemical Performance of Immunoelectrode (BSA/anti-NSE/CS/MoS_2/ITO) and Control Electrode (CS/MoS_2/ITO) Towards Detection of NSE

Electrochemical detection of NSE is done by incubating the BSA/anti-NSE/CS/MoS_2/ITO immunoelectrode by varying concentrations of NSE from 0.1 ng/mL to 200 ng/mL. The

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

View Article Online
DOI: 10.1039/D2NJ00540A

response is recorded using CV at a scan rate of 50 mV/s. The current response is found to be inversely proportional to the concentration of NSE, which is a result of specific binding between NSE and anti-NSE. This specific binding results in the formation of an immunocomplex that impedes the diffusion of $[Fe(CN)_6]^{3-/4-}$. **Fig. 10(a)** shows the variation in CV response of BSA/anti-NSE/CS/MoS_2/ITO immunoelectrode when incubated with varying concentrations of NSE from 0.1 ng/mL to 200 ng/mL. **Fig. 10(b)** shows the variation of i_{pa} with the concentration of NSE from 0.1 ng/mL to 100 ng/mL. The variation is found to be linearly decreasing and can be fitted using the following equation with R^2 equal to 0.94.

$$y = -1.84 \ \frac{\frac{\mu A}{ng}}{mL} X(Concentration\ of\ NSE\ \left(\frac{ng}{mL}\right)) + 675.37\ \mu A \text{.....} \textbf{(6)}$$

The sensitivity of the immunoelectrode can be calculated from here to be 3.35 μA /ng/mL/cm^2 by dividing the slope of calibration curve by electroactive surface area calculated for BSA/anti-NSE/CS/MoS_2/ITO [37]. The lowest concentration which could be detected for NSE from the prepared immunoelectrode is 0.1 ng/mL.

Fig. 10(c) shows the response of control electrode (CS/MoS_2/ITO) towards varying concentrations of NSE from 0.1 ng/mL to 100 ng/mL. The response is recorded in a similar way as done for BSA/anti-NSE/CS/MoS_2/ITO. An invariant response towards NSE by the control electrode shows that response of the immunoelectrode (BSA/anti-NSE/CS/MoS_2/ITO) is due to specific interaction between NSE and anti-NSE. This interaction leads to the formation of immunocomplex leading to a decrease in the peak current of immunoelectrode with increasing concentration of NSE. Thus, this study concludes that CS/MoS_2/ITO cannot interact with NSE on its own without the conjugation of anti-NSE onto its surface.

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

View Article Online
DOI: 10.1039/D2NJ00540A

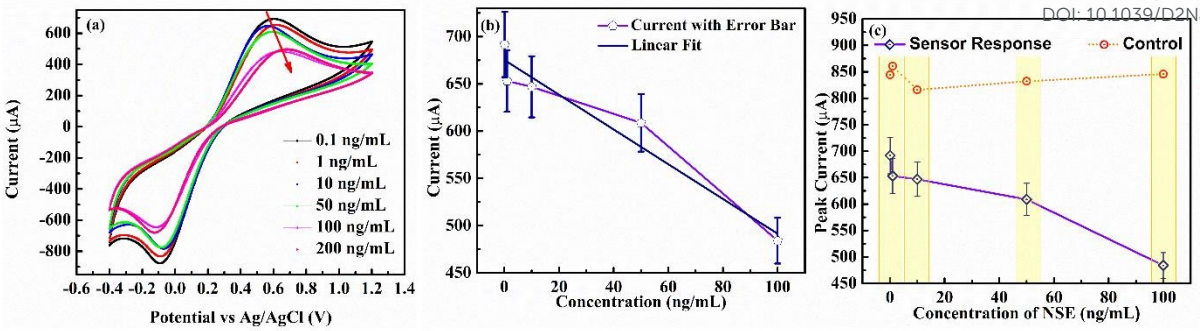


Fig. 10: (a) Electrochemical detection of NSE studied using CV with variation in concentration of NSE from 0.1 ng/mL to 200 ng/mL in PBS 7.4 containing 50 mM $[Fe(CN)_6]^{3-/4-}$ **(b)** Calibration plot between CV peak current and concentration of NSE **(c)** Response of control electrode (CS/MoS_2/ITO) towards varying concentrations of NSE from 0.1 to 100 ng/mL.

4.5 Investigation of Reproducibility and Specificity of Prepared Immunoelectrode Towards NSE

To testify the reproducibility of prepared immunoelectrode BSA/anti-NSE/CS/MoS_2/ITO, 5 different immunoelectrodes are prepared under the same conditions and following a similar protocol. The response of these immunoelectrodes is checked for 10 ng/mL of NSE. The immunoelectrodes exhibit an acceptable relative standard deviation (RSD) of 3.10% **(Fig. 11(a))**, indicating the protocol mentioned in **section 2.4** gives reproducible immunoelectrodes. To examine the specificity of the prepared immunoelectrode BSA/anti-NSE/CS/MoS_2/ITO towards NSE, various other commonly present interferents in human serum, like C-reactive protein (CRP, 10 µg/mL), Myoglobin (mB, 85 ng/mL), Cardiac Troponin I (cTnI, 0.1 ng/mL), Urea **(**0.1 mg/mL), Sodium Chloride (NaCl, 6.2 mg/mL), are taken. For recording the CV response, 20 µL of each of these interferents are incubated onto the immunoelectrode for 15 minutes. The immunoelectrode shows an RSD value in the range of 0.27% to 0.94% for other interferents whereas a relatively higher value of RSD equal to 7.8% is found for NSE, showing the specificity of immunoelectrode towards NSE. **Fig. 11(b)** displays the bar graph depicting the response of prepared immunoelectrode towards various other interferents in human serum

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

View Article Online
DOI: 10.1039/D2NJ00540A

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

including NSE (10 ng/mL). Further, to examine the performance of immunoelectrode towards the detection of NSE, when all the other common interferents are also present simultaneously, a simple control experiment is performed using the CV technique. **Fig. 11(c)** shows the performance of immunoelectrode when all the other interferents like CRP (10 µg/mL), mB (85 ng/mL), cTnI (0.1 ng/mL), Urea (0.1 mg/mL), NaCl (6.2 mg/mL) are present in their physiological concentrations at the same time without NSE (0 ng/mL, control) and along with NSE (10 ng/mL). It is observed that the immunoelectrode displays an RSD value of 0.33% in the case of control measurement which increases to 6.06% when NSE is also present along with all the other interferents. The results indicate excellent specificity of BSA/anti-NSE/CS/MoS_2/ITO towards NSE. The immobilization of anti-NSE antibodies onto the immunoelectrode confers it to the requisite specificity.

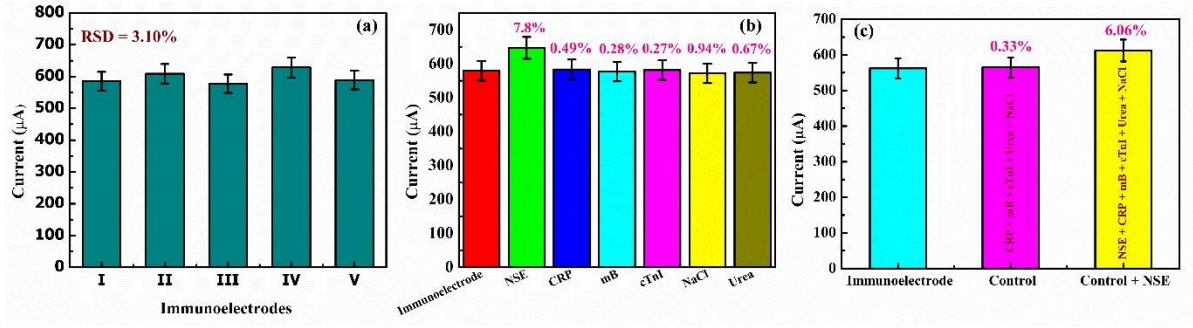


Fig. 11: (a) Reproducibility study of 5 different immunoelectrodes prepared under similar conditions **(b)** Specificity study of immunoelectrode towards various interferents commonly present in human serum (c) Specificity study of immunoelectrode towards NSE in the simultaneous presence of various other interferents.

4.6 Shelf Life and Regeneration Study of Prepared Immunoelectrode

The shelf life of BSA/anti-NSE/CS/MoS_2/ITO immunoelectrode is done by incubating it with 10 ng/mL of NSE for 15 minutes at a regular interval of 7 days. The immunoelectrode is found to retain more than 80% of its initial response till the 6^{th} week **(Fig. 12(a))**. While the response

View Article Online
DOI: 10.1039/D2NJ00540A

reduces to 70% in the 7th week. Thus, the immunoelectrode displays a shelf life of up to 6 weeks.

To check the regeneration ability of the immunoelectrode, 0.1M of glycine HCl buffer of pH 2.5 is prepared. The immunoelectrode is first tested for 10 ng/mL of NSE and the CV response is recorded. Then the immunoelectrode is washed in glycine HCl buffer solution followed by washing in pBS 7.4 buffer solution and then, it is tested for 10 ng/mL of NSE. The immunoelectrode is tested for 4 cycles and is found to retain more than 90% of its initial response till the 4th cycle (**Fig. 12(b)**).

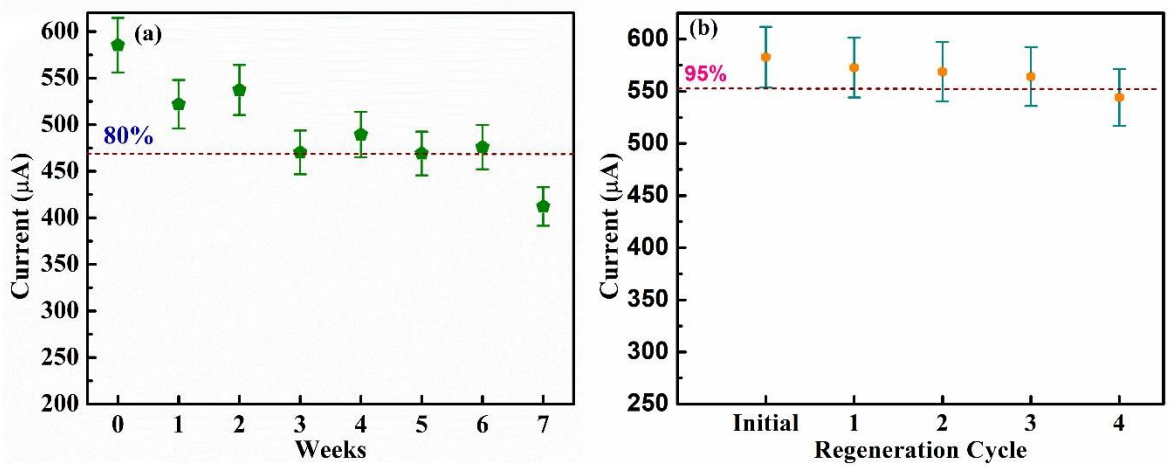


Fig. 12: (a) Shelf-life study of fabricated immunoelectrode **(b)** Regeneration study of prepared immunoelectrode.

Table 1 represents the comparison of the platform proposed in this work with other platforms reported for the detection of NSE. The proposed immunoelectrode is found to exhibit better performance in comparison to other immunoelectrodes especially in terms of the response time of immunoelectrode (incubation time of NSE). Other parameters like linear detection range, shelf life, and regenerability are also found to be appreciably good.

Table 1: Comparison of the analytical performance of immunoelectrode proposed in this study with other immunoelectrodes reported earlier for detection of NSE.

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

View Article Online
DOI: 10.1039/D2NJ00540A

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

Platform	**Linear Detection Range**	**Shelf Life (in days)**	**Incubation Time of NSE (in minutes)**	**Regenerable**	**Reference**
Au/PANI/3D-rGO/Ab	0.5 pg /mL to 10 ng/mL	30	40	Yes (5 cycles)	[3]
PPD-GR-AuNPs/Ab/SPE	1 to 1000 ng/mL	30	60	-	[38]
GCE/Ab/Ag/AuNPs-rGO/AP-anti-IgG	0.1 to 2000 ng/mL	-	40	-	[39]
BSA/anti-NSE/Polyresorcinol-Au/Pt/GCE	10 pg/mL to 100 ng/mL	30	50	-	[40]
Au-MoS_2/MOF	1 pg/mL to 100 ng/mL	30	40	-	[41]
Fc-g-Au@Pd–P(BBY) and rGO/Thi/AuPt NAs	0.0001 to 50 ng/mL	15	-	-	[42]
BSA/anti-NSE/CS/MoS_2/ITO	0.1 to 100 ng/mL	42	15	Yes (4 Cycles)	**This work**

PANI- Polyaniline, rGO- reduced graphene oxide, SPE: Screen printed electrode, PPD: poly pphenylenediamine, GR- Graphene, Au – Gold, NP- nanoparticle, Ab- antibody, GCE-Glassy

View Article Online
DOI: 10.1039/D2NJ00540A

carbon electrode, AP-alkaline phosphatase, Ag- antigen, Fc-g – ferrocene grafted, Au- gold, BBY- Bismarck Brown Y, Thi- Thionine, Pt- Platinum, NAs - Nanoassemblies

5 Conclusions

In this work, we have developed chitosan bio-functionalized MoS_2, a simple matrix based on 2 step modification process, and tested its electrochemical performance towards the detection of NSE. Compositing MoS_2 with CS provides a strategic approach to prepare a highly stable matrix favorable for immobilization of biomolecules with retention of their bioactivity along with maintenance of signal response. Also, CS provides functional groups to MoS_2 to anchor sites for the attachment of biomolecules on the matrix. Further, the immunoelectrode fabricated using CS/MoS_2 matrix exhibits a sensitivity of 3.35 μA /ng/mL/cm^2, and a wide linear detection range of 0.1 to 100 ng/mL. The immunoelectrode is found to be specific, reproducible, regenerable up to 4 cycles, and has a shelf life of 6 weeks. The sensor has an electroactive surface area of 0.55 cm^2 and a HET rate of -4.05X10^{-3} cm/s. Excellent electrochemical performance towards detection of NSE can be attributed to 1) The cumulative effect of MoS_2 and CS imparts stability and biocompatibility to the matrix 2) Introduction of functional groups which lead to the covalent attachment between antibodies and the matrix without the use of any harsh chemicals and 3) the electroactive surface area is 55% of the geometric area. The proposed platform can further be explored for the analytical detection of various other protein biomarkers using various electrochemical techniques.

Author Contributions

Ritika Khatri: Conceptualization, data curation, formal analysis, investigation, methodology, writing-original draft

Nitin K. Puri: Project administration, resources, supervision, validation, visualization, writing- review and editing.

Conflicts of Interests

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

View Article Online
DOI: 10.1039/D2NJ00540A

There are no conflicts of interest to declare.

Acknowledgments

Authors are highly obliged to Hon'ble Vice-Chancellor, Delhi Technological University, Delhi, India for promoting research and providing appropriate infrastructure and facilities. The authors also express their gratitude to Dr. Archit Dhingra for his valuable review of the manuscript. The authors would like to thank Dr. Pritam Singh (Council of Scientific and Industrial Research (CSIR)-National Physical Laboratory (NPL)) and Dr. Vikash Sharma (CSIR- NPL) for performing thickness measurement studies using a surface profilometer.

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

References

1. E. C. Welch, J. M. Powell, T. B. Clevinger, A. E. Fairman and A. J. A. F. M. Shukla, 2021, 31, 2104126. "Advances in Biosensors and Diagnostic Technologies Using Nanostructures and Nanomaterials, *Advanced Functional Materials".* https://doi.org/10.1002/adfm.202104126.

2. T. Ö. Varol, O. Hakli and U. J. N. J. o. C. Anik, 2021, 45, 1734-1742. Graphene oxide–porphyrin composite nanostructure included electrochemical sensor for catechol detection, *New Journal of Chemistry".* DOI: https://doi.org/10.1039/D0NJ05475E.

3. Q. Zhang, X. Li, C. Qian, L. Dou, F. Cui and X. J. A. b. Chen, 2018, 540, 1-8. "Label-free electrochemical immunoassay for neuron specific enolase based on 3D macroporous reduced graphene oxide/polyaniline film, *Analytical biochemistry".* https://doi.org/10.1016/j.ab.2017.10.009

4. M. Farrokhnia, G. Amoabediny, M. Ebrahimi, M. Ganjali and M. J. T. Arjmand, 2022, 238, 122947. "Ultrasensitive early detection of insulin antibody employing novel electrochemical nano-biosensor based on controllable electro-fabrication process, *Talanta".* https://doi.org/10.1016/j.talanta.2021.122947

5. I. T. Bello, A. O. Oladipo, O. Adedokun and S. M. J. M. T. C. Dhlamini, 2020, 101664. "Recent advances on the preparation and electrochemical analysis of MoS2-based materials for supercapacitor applications: A mini-review, *Materials Today Communications"* . https://doi.org/10.1016/j.mtcomm.2020.101664.

6. X. Weng and S. J. I. s. j. Neethirajan, 2018, 18, 4358-4363. "Immunosensor Based on Antibody-Functionalized MoS_2 for Rapid Detection of Avian Coronavirus on Cotton Thread, in *IEEE Sensors Journal".* DOI: 10.1109/JSEN.2018.2829084.

7. R.-M. Kong, L. Ding, Z. Wang, J. You, F. J. A. Qu and b. chemistry, 2015, 407, 369-377. "A novel aptamer-functionalized MoS 2 nanosheet fluorescent biosensor for sensitive detection of prostate specific antigen." *Analytical and bioanalytical chemistry".* https://doi.org/10.1007/s00216-014-8267-9

View Article Online
DOI: 10.1039/D1NJ00540A

8. A. Sinha, B. Tan, Y. Huang, H. Zhao, X. Dang, J. Chen and R. J. T. T. i. A. C. Jain, 2018, 102, 75-90. "MoS2 nanostructures for electrochemical sensing of multidisciplinary targets: A review, *TrAC Trends in Analytical Chemistry".* https://doi.org/10.1016/j.trac.2018.01.008.

9. N. Kaur, R. A. Mir, O. J. J. o. A. Pandey and Compounds, 2019, 782, 119-131." Electrochemical and optical studies of facile synthesized molybdenum disulphide (MoS2) nano structures, *Journal of Alloys and Compounds".* https://doi.org/10.1016/j.jallcom.2018.12.145

10. D. B. Altuntaş, F. J. M. S. Kuralay and E. B, 2021, 270, 115215. "MoS2/Chitosan/GOx-Gelatin modified graphite surface: Preparation, characterization and its use for glucose determination, *Materials Science and Engineering: B".* https://doi.org/10.1016/j.mseb.2021.115215

11. M. Chen, H. Li, X. Su, R. Wu, H. Feng, X. Shi, J. Liang, J. Chen and G. J. N. J. o. C. Li, 2021, 45, 8608-8618. "Label-free electrochemical aptasensor based on reduced graphene oxide–hemin–chitosan nanocomposite for the determination of glypican-3, *New Journal of Chemistry*". DOI: https://doi.org/10.1039/D1NJ00633A

12. Y. Wang, B. Zhang, Y. Tang, F. Zhao and B. J. M. J. Zeng, 2021, 168, 106505. "Fabrication and application of a rutin electrochemical sensor based on rose-like AuNPs-MoS2-GN composite and molecularly imprinted chitosan, *Microchemical Journal"* https://doi.org/10.1016/j.microc.2021.106505

13. M. Sabbaghan, A. S. Shahvelayati, K. J. S. A. P. A. M. Madankar and B. Spectroscopy, 2015, 135, 662-668. "CuO nanostructures: optical properties and morphology control by pyridinium-based ionic liquids, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* ". https://doi.org/10.1016/j.saa.2014.07.097

14. A. Roointan, T. A. Mir, S. I. Wani, K. K. Hussain, B. Ahmed, S. Abrahim, A. Savardashtaki, G. Gandomani, M. Gandomani, R. J. J. o. p. Chinnappan and b. analysis, 2019, 164, 93-103. "Early detection of lung cancer biomarkers through biosensor technology: A review, *Journal of pharmaceutical and biomedical analysis".* https://doi.org/10.1016/j.jpba.2018.10.017

15. R. A. Soomro, N. H. Kalwar, A. Avci, E. Pehlivan, K. R. Hallam, M. J. B. Willander and Bioelectronics, 2019, 141, 111331. "In-situ growth of NiWO4 saw-blade-like nanostructures and their application in photo-electrochemical (PEC) immunosensor system designed for the detection of neuron-specific enolase, *Biosensors and Bioelectronics".* https://doi.org/10.1016/j.bios.2019.111331

16. C. Tang, P. Wang, K. Zhou, J. Ren, S. Wang, F. Tang, Y. Li, Q. Liu, L. J. B. Xue and Bioelectronics, 2022, 197, 113779. "AgPt alloy/NGR as a dual signal amplification strategy for sensitive detection of Neuron-specific enolase, *Biosensors and Bioelectronics".* https://doi.org/10.1016/j.bios.2021.113779

17. X. Yu, Y. Li, Y. Li, S. Liu, Z. Wu, H. Dong, Z. Xu, X. Li and Q. J. T. Liu, 2022, 236, 122865. "An electrochemical amplification strategy based on the ferrocene functionalized cuprous oxide superparticles for the detection of NSE, *Talanta".* https://doi.org/10.1016/j.talanta.2021.122865

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

View Article Online
DOI: 10.1039/D2NJ00540A

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

18. G. Mo, X. He, D. Qin, X. Jiang, X. Zheng and B. J. A. Deng, 2021, 146, 1334-1339. "A potential-resolved electrochemiluminescence resonance energy transfer strategy for the simultaneous detection of neuron-specific enolase and the cytokeratin 19 fragment, *Analyst".* DOI: https://doi.org/10.1039/D0AN02106G

19. R. Khatri and N. K. J. V. Puri, 2020, 175, 109250. "Electrochemical study of hydrothermally synthesised reduced MoS2 layered nanosheets, *Vacuum".* https://doi.org/10.1016/j.vacuum.2020.109250

20. W. Zhang, Y. Wang, D. Zhang, S. Yu, W. Zhu, J. Wang, F. Zheng, S. Wang and J. J. N. Wang, 2015, 7, 10210-10217. "A one-step approach to the large-scale synthesis of functionalized MoS 2 nanosheets by ionic liquid assisted grinding, *Nanoscale".* DOI:https://doi.org/10.1039/C5NR02253C

21. H. Li, Q. Zhang, C. C. R. Yap, B. K. Tay, T. H. T. Edwin, A. Olivier and D. J. A. F. M. Baillargeat, 2012, 22, 1385-1390. "From bulk to monolayer MoS2: evolution of Raman scattering." *Advanced Functional Materials".* https://doi.org/10.1002/adfm.201102111

22. A. Zając, J. Hanuza, M. Wandas, L. J. S. A. P. A. M. Dymińska and B. Spectroscopy, 2015, 134, 114-120. "Determination of N-acetylation degree in chitosan using Raman spectroscopy, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy".* https://doi.org/10.1016/j.saa.2014.06.071

23. K. Kasinathan, B. Murugesan, N. Pandian, S. Mahalingam, B. Selvaraj and K. J. I. j. o. b. m. Marimuthu, 2020, 149, 1019-1033. "Synthesis of biogenic chitosan-functionalized 2D layered MoS2 hybrid nanocomposite and its performance in pharmaceutical applications: in-vitro antibacterial and anticancer activity, *International Journal of Biological Macromolecules".*https://doi.org/10.1016/j.ijbiomac.2020.02.003

24. J. Wang, W. Zhang, Y. Wang, W. Zhu, D. Zhang, Z. Li, J. J. P. Wang and P. S. Characterization, 2016, 33, 825-832. "Enhanced exfoliation effect of solid auxiliary agent on the synthesis of biofunctionalized MoS2 using grindstone chemistry, *Particle & Particle Systems Characterization".* https://doi.org/10.1002/ppsc.201600114

25. H. Wang, P. Cheng, J. Shi, D. Wang, H. Wang, J. Pezoldt, M. Stich, R. Chen, P. A. van Aken and W. J. G. C. Huang, 2021, 23, 3642-3648. "Efficient fabrication of MoS 2 nanocomposites by water-assisted exfoliation for nonvolatile memories, *Green Chemistry".* DOI: https://doi.org/10.1039/D1GC00162K

26. M. Zhao, Z. Huang, S. Wang and L. J. C. E. J. Zhang, 2020, 401, 126006. "Ultrahigh efficient and selective adsorption of Au (III) from water by novel Chitosan-coated MoS2 biosorbents: Performance and mechanisms, *Chemical Engineering Journal.* https://doi.org/10.1016/j.cej.2020.126006

27. Z. Feng, X. Liu, L. Tan, Z. Cui, X. Yang, Z. Li, Y. Zheng, K. W. K. Yeung and S. J. S. Wu, 2018, 14, 1704347. "Electrophoretic deposited stable chitosan@ MoS2 coating with rapid in situ bacteria-killing ability under dual-light irradiation, S*mall".* https://doi.org/10.1002/smll.201704347.

View Article Online
DOI: 10.1039/D2NJ00540A

28. O. Akakuru, H. Louis, P. Amos, O. Akakuru, E. Nosike and E. J. B. P. Ogulewe, 2018, 7, 2167-0501.1000241. "The chemistry of chitin and chitosan justifying their nanomedical utilities." *Biochem Pharmacol (Los Angel)"*. DOI: 10.4172/2167-0501.1000241.

29. D. Sandil, S. Srivastava, B. Malhotra, S. Sharma, N. K. J. J. o. A. Puri and Compounds, 2018, 763, 102-110. "Biofunctionalized tungsten trioxide-reduced graphene oxide nanocomposites for sensitive electrochemical immunosensing of cardiac biomarker, *Journal of Alloys and Compounds* ". https://doi.org/10.1016/j.jallcom.2018.04.293

30. R. Vinoth, I. M. Patil, A. Pandikumar, B. A. Kakade, N. M. Huang, D. D. Dionysios and B. J. A. O. Neppolian, 2016, 1, 971-980. "Synergistically enhanced electrocatalytic performance of an N-doped graphene quantum dot-decorated 3D MoS2–graphene nanohybrid for oxygen reduction reaction, *Acs Omega".* https://doi.org/10.1021/acsomega.6b00275

31. A. K. Singh, T. K. Dhiman, G. Lakshmi and P. R. J. B. Solanki, 2021, 137, 107684. "Dimanganese trioxide (Mn2O3) based label-free electrochemical biosensor for detection of Aflatoxin-B1, *Bioelectrochemistry".* https://doi.org/10.1016/j.bioelechem.2020.107684

32. V. K. Singh, S. Kumar, S. K. Pandey, S. Srivastava, M. Mishra, G. Gupta, B. Malhotra, R. Tiwari, A. J. B. Srivastava and Bioelectronics, 2018, 105, 173-181. "Fabrication of sensitive bioelectrode based on atomically thin CVD grown graphene for cancer biomarker detection, *Biosensors and Bioelectronics".* https://doi.org/10.1016/j.bios.2018.01.014

33. Y. Guo, Y. Shu, A. Li, B. Li, J. Pi, J. Cai, H.-h. Cai and Q. J. J. o. M. C. B. Gao, 2017, 5, 5532-5538. "Efficient electrochemical detection of cancer cells on in situ surface-functionalized MoS 2 nanosheets, *Journal of Materials Chemistry B".* DOI: https://doi.org/10.1039/C7TB01024A

34. M. Rezki, N. L. W. Septiani, M. Iqbal, S. Harimurti, P. Sambegoro, D. R. Adhika and B. J. J. o. M. C. B. Yuliarto, 2021. "Amine-functionalized Cu-MOF nanospheres towards label-free hepatitis B surface antigen electrochemical immunosensors, *Journal of Materials Chemistry B".* DOI: https://doi.org/10.1039/D1TB00222H

35. X. Xi, D. Wu, W. Ji, S. Zhang, W. Tang, Y. Su, X. Guo and R. J. A. F. M. Liu, 2020, 30, 1905361. "Manipulating the Sensitivity and Selectivity of OECT-Based Biosensors via the Surface Engineering of Carbon Cloth Gate Electrodes, *Advanced Functional Materials".* https://doi.org/10.1002/adfm.201905361

36. A. A. Aljabali, J. E. Barclay, J. N. Butt, G. P. Lomonossoff and D. J. J. D. t. Evans, 2010, 39, 7569-7574. "Redox-active ferrocene-modified Cowpea mosaic virus nanoparticles, *Dalton transactions".* DOI: https://doi.org/10.1039/C0DT00495B

37. A. Kaur, S. Rana, A. Bharti, G. R. Chaudhary and N. J. M. A. Prabhakar, 2021, 188, 1-14. "Voltammetric detection of vitamin D employing Au-MoS2 hybrid as immunosensing platform, *Microchimica Acta".* https://doi.org/10.1007/s00604-021-04862-6

38. J. Amani, M. Maleki, A. Khoshroo, A. Sobhani-Nasab and M. J. A. b. Rahimi-Nasrabadi, 2018, 548, 53-59. "An electrochemical immunosensor based on poly p-phenylenediamine and

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

View Article Online
DOI: 10.1039/D2NJ00540A

graphene nanocomposite for detection of neuron-specific enolase via electrochemically amplified detection, *Analytical biochemistry".* https://doi.org/10.1016/j.ab.2018.02.024

39. Z. Wei, J. Zhang, A. Zhang, Y. Wang and X. J. M. Cai, 2017, 22, 392. "Electrochemical detecting lung cancer-associated antigen based on graphene-gold nanocomposite, *Molecules".* https://doi.org/10.3390/molecules22030392

40. H. Wang and Z. J. M. A. Ma, 2017, 184, 3247-3253. "Amperometric immunoassay for the tumor marker neuron-specific enolase using a glassy carbon electrode modified with a nanocomposite consisting of polyresorcinol and of gold and platinum nanoparticles, *Microchimica Acta* ". https://doi.org/10.1007/s00604-017-2287-z.

41. H. Dong, S. Liu, Q. Liu, Y. Li, Y. Li, Z. J. B. Zhao and Bioelectronics, 2022, 195, 113648. "A dual-signal output electrochemical immunosensor based on Au–MoS2/MOF catalytic cycle amplification strategy for neuron-specific enolase ultrasensitive detection, *Biosensors and Bioelectronics".* https://doi.org/10.1016/j.bios.2021.113648

42. Y. Chen, X.-Y. Ge, S.-Y. Cen, A.-J. Wang, X. Luo, J.-J. J. S. Feng and A. B. Chemical, 2020, 311, 127931. "Ultrasensitive dual-signal ratiometric electrochemical aptasensor for neuron-specific enolase based on Au nanoparticles@ Pd nanoclusters-poly (bismarck brown Y) and dendritic AuPt nanoassemblies, *Sensors and Actuators B: Chemical".* https://doi.org/10.1016/j.snb.2020.127931

Published on 17 March 2022. Downloaded on 3/23/2022 7:06:37 AM.

arXiv:2203.06456v1 [cs.LG] 12 Mar 2022

ENERGY NETWORKS FOR STATE ESTIMATION WITH RANDOM SENSORS USING SPARSE LABELS

A PREPRINT

Yash Kumar
Department of Mechanical Engineering
Delhi Technological University
Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India
yashk8481@gmail.com

Souvik Chakraborty
Department of Applied Mechanics
Indian Institute of Technology Delhi
Hauz Khas - 110042, New Delhi, India
souvik@am.iitd.ac.in

March 15, 2022

ABSTRACT

State estimation is required whenever we deal with high-dimensional dynamical systems, as the complete measurement is often unavailable. It is key to gaining insight, performing control or optimizing design tasks. Most deep learning-based approaches require high-resolution labels and work with fixed sensor locations, thus being restrictive in their scope. Also, doing Proper orthogonal decomposition (POD) on sparse data is nontrivial. To tackle these problems, we propose a technique with an implicit optimization layer and a physics-based loss function that can learn from sparse labels. It works by minimizing the energy of the neural network prediction, enabling it to work with a varying number of sensors at different locations. Based on this technique we present two models for discrete and continuous prediction in space. We demonstrate the performance using two high-dimensional fluid problems of Burgers' equation and Flow Past Cylinder for discrete model and using Allen–Cahn equation and Convection-diffusion equations for continuous model. We show the models are also robust to noise in measurements.

***K*eywords** State estimation · Differentiable implicit layers · Dynamical systems

1 Introduction

State estimation is the ability to recover flow based on a few measurements. It is an inverse problem and arises in many engineering applications such as remote sensing, medical imaging, ocean dynamics, reservoir modeling, and blood flow modeling. Uses of fluid estimation include flow control [1, 2], cardiac blood flow modeling [3, 4, 5], ship wake identification [6], climate prediction [7], optimizing machine design for low-drag vehicles, efficient turbo-machines, etc. Few challenges faced in the processes are limited sensors, sparse label data, moving sensors, ill-posed problems, noisy measurements, etc. This work focuses on learning from moving sparse label data and sensor measurements with a deep learning-based model using an implicit optimization layer for network training. Sparse fluid is encountered in various situations. One reason is that storing high-resolution data generated during direct numerical simulation is challenging due to limited storage space. It makes analysis, sharing, and visualization difficult. Another important reason is that real data is hard to measure on a full scale, like cardiovascular blood flow data obtained from flow magnetic resonance imaging (MRI) [8, 9, 10].

For high dimensional state estimation problems like fluids, popular approaches include library-based approaches observer dynamical system stochastic approaches. Library-based methods use offline data, and the library consists of generic modes such as Fourier, wavelet, discrete cosine transform basis, or data specific Proper orthogonal decomposition (POD) or Dynamic mode decomposition (DMD) modes, or training data. Library-based approaches using sparse representation assume state can be expressed as the combination of library elements. In an observer dynamical system, we assume the system's dynamics to produce a full state and update it based on new measurements to reduce estimation error forming a closed feedback loop. The estimate is maintained by Kalman filtering [11, 12, 13]. Tu et al. [14] applied

dynamic mode decomposition [15, 16] as a reduced-order model to Kalman smoother estimate to identify coherent structures. Buffoni et al. [17] used a nonlinear observer-based on Galerkin projection of Navier-Stokes equation to estimate POD coefficients. Stochastic estimation was proposed by Adrian [18] for a turbulence study where the conditional mean was approximated using a power series. Extension to this method [19, 20, 21] can be found in the literature. Bonnet et al. [22] extended stochastic approach to estimate POD coefficients. A linear mapping between sensors and coefficients was assumed. These approaches allow more flexibility in sensor placements and have been applied for flow control over airfoil [23] and analyzing isotropic turbulence [24, 25].

We consider a problem with sparse label data whose position may vary with time, for which POD can not be performed. Thus it restricts us from using traditional POD-based approaches. Other deep learning-based approaches like [26, 27, 28] give accurate predictions with fewer sensors but require high-resolution labels for training. Gao et al. [29] uses physics-based loss for super-resolution using spare data and thus assumes a fixed number of sensors and their positions. In this work, we propose an Energy network for state estimation with random sensors (ENSERS), a technique to learn a model from spare training labels capable of predicting full states given a varied number of sensors at random locations. We present two models trained using this technique. The first one produces discrete high-dimensional predictions in space. Second, produce continuous prediction utilizing the information of coordinates. We demonstrate the results corresponding to four high complexity problems: 2-dimensional (2D) coupled Burgers' equation, transient flow, Allen–Cahn equation, and Convection-diffusion equation.

The remainder of the paper is organized as follows. In Section 2, details on the problem statement is provided. Details on the proposed approach are provided in Section 3.1. Section 3 and 4 give details on discrete and continuous formulation with two numerical examples in each to illustrate the performance of the proposed approach. Finally, Section 5 provides the concluding remarks.

2 Problem statement

Consider a dynamical system obtained by partial discretization of the d-dimensional governing differential equations:

$$J_t(x,t) = F(x, J(x,t)), \quad J^n = J(x,t_n), \quad x \in \Omega \tag{1}$$

The simulation time domain is discretized by L steps and the space domain is discretized by ω segments resulting in

$$Z = \{J_m^l \in \mathbb{R}^\omega \mid l = 0,...,L-1,\ m = 0,...,M-1\} \tag{2}$$

where $Z \in \mathbb{R}^{L\times M\times\omega}$, l is time step index, $M =$ number of system's state variables. e.g. $m = 0$ represents x-velocity and $m = 1$ represents y-velocity in § 3.2 of 2D coupled Burgers' equation. We consider sensor location and data location, represented by integers in discrete domain respectively as

$$S = \{\lambda_m^l \in \mathbb{Z}^p \cap [0,\omega-1]^p \mid l = 0,...,L-1,\ m = 0,...,M-1\} \tag{3}$$

$$T = \{\pi_m^l \in \mathbb{Z}^h \cap [0,\omega-1]^h \mid l = 0,...,L-1,\ m = 0,...,M-1\} \tag{4}$$

where $p =$ number of sensors, $h =$ number of data nodes, $S \in \mathbb{Z}^{L\times M\times p} \cap [0,\omega-1]^{L\times M\times p}$, $T \in \mathbb{Z}^{L\times M\times h} \cap [0,\omega-1]^{L\times M\times h}$. The corresponding sensor values and data values are

$$X = \{\sigma_m^l \in \mathbb{R}^p \mid l = 0,...,L-1,\ m = 0,...,M-1\} \tag{5}$$

$$\Phi = \{\psi_m^l \in \mathbb{R}^h \mid l = 0,...,L-1,\ m = 0,...,M-1\} \tag{6}$$

$$\sigma_m^l = \Lambda_m^l J_m^l \tag{7}$$

$$\psi_m^l = \Pi_m^l J_m^l \tag{8}$$

where $X \in \mathbb{R}^{L\times M\times p}$, $\Phi \in \mathbb{R}^{L\times M\times h}$. $\Lambda_m^l \in \mathbb{R}^{p\times\omega}$ and $\Pi_m^l \in \mathbb{R}^{h\times\omega}$ are measurement matrices composed of one-hot row vectors. Λ_m^l and Π_m^l are defined as

$$\Lambda_{mij}^l = \begin{cases} 1 & \text{if } \lambda_{mi}^l = j \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

$$\Pi^l_{mij} = \begin{cases} 1 & \text{if } \pi^l_{mi} = j \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

Note that integer senor locations λ^l_m are selected randomly and are kept fixed during training. Similarly, a different set of sensor locations S is selected for testing the network. In this work, we aim to use sensor data from a set of γ system states and produce γ high dimensional states. Thus we divided the sensor values and data into chunks, each with γ states.

$$\chi^k = \{\sigma^{k*z+i}_m \in \mathbb{R}^p \mid i = 0, ..., \gamma-1,\ m = 0, ..., M-1\} \tag{11}$$

$$\phi^k = \{\psi^{k*z+i}_m \in \mathbb{R}^h \mid i = 0, ..., \gamma-1,\ m = 0, ..., M-1\} \tag{12}$$

where $\chi^k \in \mathbb{R}^{\gamma \times M \times p}$ are sensor values for γ time steps, $\phi^k \in \mathbb{R}^{\gamma \times M \times h}$ are data values for γ time steps, $k = 0, ..., N-1$, N is number of train samples, z is time steps between first state σ^{k*z}_m of each tensor χ^k.

3 Discrete space models

3.1 Proposed approach

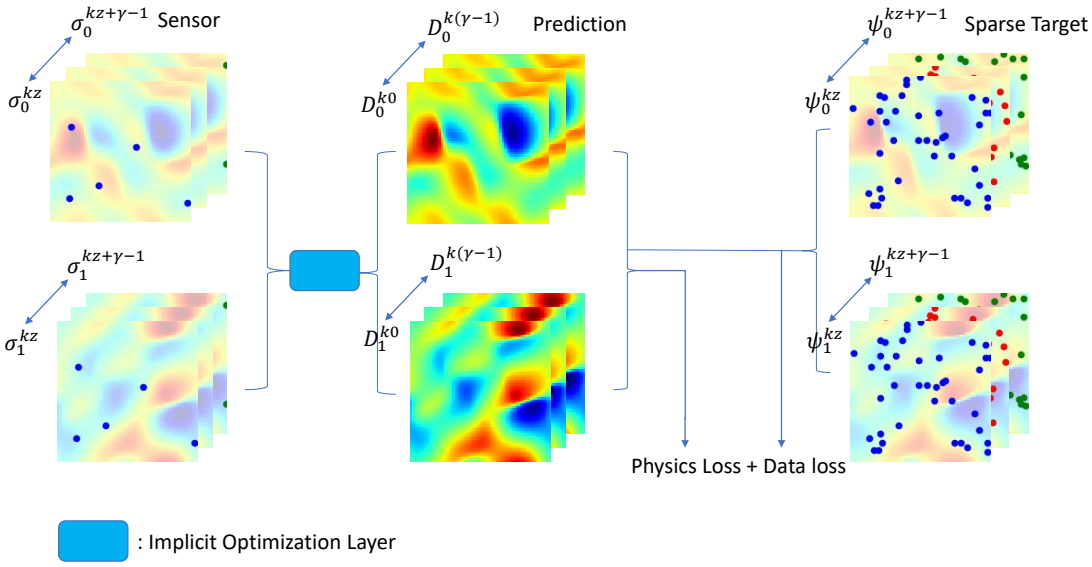


Figure 1: Network architecture of proposed Ensers model.

In this section, we propose a novel deep learning-based framework for state estimation. Prediction of γ high dimensional states is done via feed-forward neural network (FNN) Γ using optimized reduced state vector $\xi^k \in \mathbb{R}^\varsigma$:

$$\Gamma(\xi^k) = D^k = \{D^{ki}_m \in \mathbb{R}^\omega \mid i = 0, ..., \gamma-1,\ m = 0, ..., M-1\} \tag{13}$$

where $D^k \in \mathbb{R}^{\gamma \times M \times \omega}$ is a third-order tensor of predicted states. The output vector of FNN has a dimension of $\gamma * M * \omega$ which is reshaped into a tensor of shape $\gamma \times M \times \omega$. The reduced state is obtained by solving the following minimization problem using sensor data χ^k and predicted states by the neural network.

$$\xi^k = \arg\min_{\tilde{\xi}^k} \|vec(\chi^k - \rho(\tilde{\xi}^k))\|_2^2 \tag{14a}$$

$$\rho(\tilde{\xi}^k) = \{\tilde{Q}_m^{ki} \in \mathbb{R}^p \mid i = 0, ..., \gamma - 1,\ m = 0, ..., M - 1\} \tag{14b}$$

$$\tilde{Q}_m^{ki} = \Lambda_m^{kz+i} \tilde{D}_m^{ki} \tag{14c}$$

$$\Gamma(\tilde{\xi}^k) = \{\tilde{D}_m^{ki} \in \mathbb{R}^\omega \mid i = 0, ..., \gamma - 1,\ m = 0, ..., M - 1\} \tag{14d}$$

where $\tilde{Q}_m^{ki} \in \mathbb{R}^p$ are values of predicted states at sensor locations λ_m^{kz+i}, $\Lambda_m^{kz+i} \in \mathbb{R}^{p \times \omega}$ is measurement matrix. Note that in practice ξ^k is obtained by a few steps of gradient descent instead of global minimization. This inner optimization loop is implemented using the library 'higher' [30] in PyTorch. Also, dimension ς of the reduced state vector is quite small (e.g. 8 in first experiment § 3.2) therefore the time required for gradient descent steps is negligible. The network Γ is trained by minimizing data loss and physics-based loss $P(\chi^k)$ across training samples N.

$$\theta^* = \arg\min_\theta \sum_{k=0}^{N-1} \|vec(\phi^k - \Upsilon(\chi^k))\|_2^2 + P(\chi^k) \tag{15a}$$

$$\Upsilon(\chi^k) = \{Y_m^{ki} \in \mathbb{R}^h \mid i = 0, ..., \gamma - 1,\ m = 0, ..., M - 1\} \tag{15b}$$

$$Y_m^{ki} = \Pi_m^{kz+i} D_m^{ki} \tag{15c}$$

where $D_m^{ki} \in \mathbb{R}^\omega$ are predicted states by network Γ from Eq. (13), $Y_m^{ki} \in \mathbb{R}^h$ are values of predicted states at data locations π_m^{kz+i}, $\Upsilon(\chi^k)$ are predicted states composed of γ time steps. Fig. 1 shows the network architecture during training. Equations (13) and (14) together forms the implicit optimization layer shown in the Fig. 1. Training and testing procedure is shown in Algorithm 1 and 2 respectively. Note that for demonstration purpose, in all Algorithms a batch size of 1 is considered but in practice batch size is selected based on problem as mentioned in each experiment section. Also we use Huber loss function during testing as it is more robust to noise than mean squared error (MSE) loss function.

Methods based on implicit optimization layers come under the category of Optimization-based Modeling architectures [31] and are well-studied for generic classification, and structured prediction tasks [32, 33, 34, 35]. The most common way of training such models is through Unrolled Differentiation [36, 37, 38, 33]. It is done by introducing an optimization procedure such as gradient descent into the inference procedure. Other ways of training include Implicit argmin differentiation using implicit function theorem but need argmin operations to be convex. This method can be found in works of [39, 40]. In this work, we use Unrolled Differentiation technique for training.

3.1.1 Physics-based loss function

The physics-based loss function is used in the approach because of spare training labels. Training any network with just spare labels will produce garbage values on nodes without labels. Physics-based loss functions have been used to train neural networks for solving PDEs. A popular class of methods is PINNs [41]. The basic idea here is to place a neural network prior to the state variable and then estimate the neural network parameters by using a physics-informed loss function. Several improvements to the originally proposed PINN can also be found in the literature. For example, Zhu et al. [42] developed convolutional PINN for time-independent systems. Geneva and Zabaras [43] used physics constrained auto-regressive model for surrogate modeling of dynamical systems. We use Runge-Kutta methods with q stages for defining loss between γ state predictions. Let

$$V^n = D^{k0} \tag{16a}$$

$$V^{n+c_i} = D^{ki}, \quad i = 1, ..., q \tag{16b}$$

$$V^{n+1} = D^{k(\gamma-1)} \tag{16c}$$

where D^{ki} are network prediction. General form of Runge-Kutta methods with q stages applied to Eq. (1):

$$V^{n+c_i} = V^n - \Delta t \sum_{j=1}^{q} a_{ij} F(V^{n+c_j}), \quad i = 1, .., q \tag{17a}$$

Algorithm 1: Training ENSERS

Inputs: Z, S, T. {Eq. (2), Eq. (3), Eq. (4)}
Set Hyper-parameters: η_o: outer learning rate, η_{i0}: inner learning rate at epoch=0, $\hat{\eta}_i$: inner learning rate rate, β: batch size, I_o: outer iterations, I_i: inner iterations, $\hat{\zeta}$: physics penalty rate, ζ_0: physics penalty at epoch=0, N: number of train samples, z, γ.
Calculate data-set: χ, ϕ, Λ, Π
Initialize: Neural network model: $\Gamma(\cdot;\theta)$
for $\iota_o = 0$ **to** $I_o - 1$ **do** {Outer optimization Loop}
  $\eta_i = \eta_{i0} + \iota_o \hat{\eta}_i$ {Schedule inner learning rate}
  $\zeta = \zeta_0 + \iota_o \hat{\zeta}$ {Schedule physics penalty}
  for $k = 0$ **to** $N - 1$ **do**
    Initialize: $\tilde{\xi}^k$
    for $\iota_i = 0$ **to** $I_i - 1$ **do** {Inner optimization Loop}
      $\tilde{D}^k = \Gamma(\tilde{\xi}^k)$ {Eq. (14d)}
      $\tilde{Q}^k \leftarrow$ DrawValuesAtSensorLocations($\tilde{D}^k$) {Eq. (14c)}
      $\mathfrak{L} = MSE(\chi^k, \tilde{Q}^k)$
      $\frac{\partial \mathfrak{L}}{\partial \tilde{\xi}^k} \leftarrow$ Backprop($\mathfrak{L}$)
      $\tilde{\xi}^k = \tilde{\xi}^k - \eta_i \frac{\partial \mathfrak{L}}{\partial \tilde{\xi}^k}$
    end for
    $\xi^k = \tilde{\xi}^k$
    $D^k = \Gamma(\xi^k)$ {Eq. (13)}
    $Y^k \leftarrow$ DrawValuesAtDataLocations(D^k) {Eq. (15c)}
    $\mathscr{L} = MSE(\phi^k, Y^k) + \zeta P(D^k)$ {Calculate loss}
    $\frac{\partial \mathscr{L}}{\partial \theta} \leftarrow$ Backprop($\mathscr{L}$)
    $\theta \leftarrow \theta - \eta_o \frac{\partial \mathscr{L}}{\partial \theta}$ {Update weights}
  end for
end for
Output: Trained network $\Gamma(\cdot;\theta^*)$.

Algorithm 2: Testing ENSERS

Inputs: Trained network $\Gamma(\cdot;\theta^*)$, S. {Eq. (3)}
Set Hyper-parameters: η_i: inner learning rate, I_i: inner iterations, $\hat{N}$: number of test samples, z, γ.
Calculate data-set: χ, Λ
for $k = 0$ **to** $\hat{N} - 1$ **do**
  Initialize: $\tilde{\xi}^k$
  for $\iota_i = 0$ **to** $I_i - 1$ **do** {Inner optimization Loop}
    $\tilde{D}^k = \Gamma(\tilde{\xi}^k)$ {Eq. (14d)}
    $\tilde{Q}^k \leftarrow$ DrawValuesAtSensorLocations($\tilde{D}^k$) {Eq. (14c)}
    $\mathfrak{L} = MSE(\chi^k, \tilde{Q}^k)$
    $\frac{\partial \mathfrak{L}}{\partial \tilde{\xi}^k} \leftarrow$ Backprop($\mathfrak{L}$)
    $\tilde{\xi}^k = \tilde{\xi}^k - \eta_i \frac{\partial \mathfrak{L}}{\partial \tilde{\xi}^k}$
  end for
  $\xi^k = \tilde{\xi}^k$
  $D^k = \Gamma(\xi^k)$ {Eq. (13)}
end for
Output: Predicted states $\{D_m^{ki} \mid k = 0,...,N-1,\ i = 0,...,\gamma-1,\ m = 0,...,M-1\}$.

$$V^{n+1} = V^n - \Delta t \sum_{j=1}^{q} b_j F(V^{n+c_j}) \tag{17b}$$

We use the Implicit Runge-Kutta methods with q stages and thus parameters $\{a_{ij}, b_j, c_j\}$ are chosen accordingly. Now, shifting second term on right hand side (RHS) in Eq. (17) to left hand side (LHS) and replacing exact operator F with numerical gradient based operator $\hat{F}$

$$\hat{W}_i = V^{n+c_i} + \Delta t \sum_{j=1}^{q} a_{ij} \hat{F}(V^{n+c_j}), \quad i = 1, .., q \tag{18a}$$

$$\hat{W_{q+1}} = V^{n+1} + \Delta t \sum_{j=1}^{q} b_j \hat{F}(V^{n+c_j}) \tag{18b}$$

$$P(\chi^k) = \sum_{i=1}^{q+1} \|vec(\hat{W}_i - V^n)\|_2^2 \tag{18c}$$

where $\hat{W}_i$ are different estimates of V^n, $P(\chi^k)$ is the physics based loss function. For calculating loss the spatial gradients are approximated using Sobel filter 2D convolutions [44]. See § A for additional details. Note that physics-based loss function is only used during training of network.

In next section, we present two examples to show model proposed is able to learn from sparse moving data labels. We illustrate the performance of the proposed approach with plots of prediction. We show that model is robust against noisy sensor measurements by showing error corresponding to various noise level. Error used as a quantitative metric in plots in defined as

$$\varepsilon^k = \frac{\left\| D_m^{k\gamma^*} - J_m^{k*z+\gamma^*} \right\|_2}{\left\| J_m^{k*z+\gamma^*} \right\|_2} \quad k = 0, ..., \hat{N} \tag{19}$$

where $\varepsilon^k \in \mathbb{R}$ represents the error, $\hat{N} number of test samples$, $J_m^{k*z+\gamma^*} \in \mathbb{R}^\omega$ are the true state and $D_m^{k\gamma^*} \in \mathbb{R}^\omega$ are the predicted state using the proposed approach. $\|\cdot\|_2$ represents the L2 norm.

3.2 Experiment: 2D coupled Burgers' equation

As the first example, we consider the the 2D coupled Burgers' system. It has the same convective and diffusion form as the in-compressible Navier-Stokes equations. It is an important model for understanding of various physical flows and problems, such as hydrodynamic turbulence, shock wave theory, wave processes in thermo-elastic medium, vorticity transport, dispersion in porous medium. The governing equations for Burgers' equation takes the following form:

$$\boldsymbol{u}_t + \boldsymbol{u} \cdot \nabla \boldsymbol{u} - \nu \Delta \boldsymbol{u} = 0, \tag{20}$$

with periodic boundary condition

$$\begin{aligned} \boldsymbol{u}(x = 0, y, t) &= \boldsymbol{u}(x = L, y, t), \\ \boldsymbol{u}(x, y = 0, t) &= \boldsymbol{u}(x, y = L, t). \end{aligned} \tag{21}$$

Eq. (20) can be written in expanded form as

$$\begin{aligned} \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - \nu(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}) = 0 \\ \frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} - \nu(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}) = 0, \end{aligned} \tag{22}$$

where ν is viscosity, u and v are the x and y components of velocity. We consider $\{x, y\} \in [0, 1]$. The initial condition is defined using truncated Fourier series with random coefficients:

$$\boldsymbol{u}(x, y, t = 0) = \frac{2\boldsymbol{w}(x, y)}{\max_{\{x,y\}} |\boldsymbol{w}(x, y)|} + \boldsymbol{c}, \tag{23}$$

where

$$\boldsymbol{w}(x,y) = \sum_{i=-L}^{L}\sum_{j=-L}^{L} \boldsymbol{a}_{ij}\sin(2\pi(ix+jy)) + \boldsymbol{b}_{ij}\cos(2\pi(ix+jy)), \tag{24}$$

where $\boldsymbol{a}_{ij}, \boldsymbol{b}_{ij} \sim \boldsymbol{N}(0, \mathbf{I}_2)$, $L = 4$ and $\boldsymbol{c} \sim \mathscr{U}(-1,1) \in \mathbb{R}^2$.

3.2.1 Data-set and Model Parameters

We use FeNICS [45] computing platform to solve the partial differential equations (22) for generating data-set. We discretize the spatial domain with 64×64 grid and use a time-step of 0.005. Parameters related to data-set considered are displayed in Table 1. We use noisy measurements to test the approach. Noise level is measured in signal to noise ratio (SNRdB) in decibels(dB) and is represented by SNR_{dB}. Signal after adding noise r is formulated as:

$$r = s + \left(\frac{P}{2 * 10^{\frac{SNR_{dB}}{10}}}\right)^{0.5} * \mathscr{N} \tag{25a}$$

$$P = \frac{\sum s^2}{n} \tag{25b}$$

where, s is noise-free signal, $\mathscr{N}$ is random variable with standard normal distribution. Plots depicting different noise levels used for testing is shown in Fig. 2.

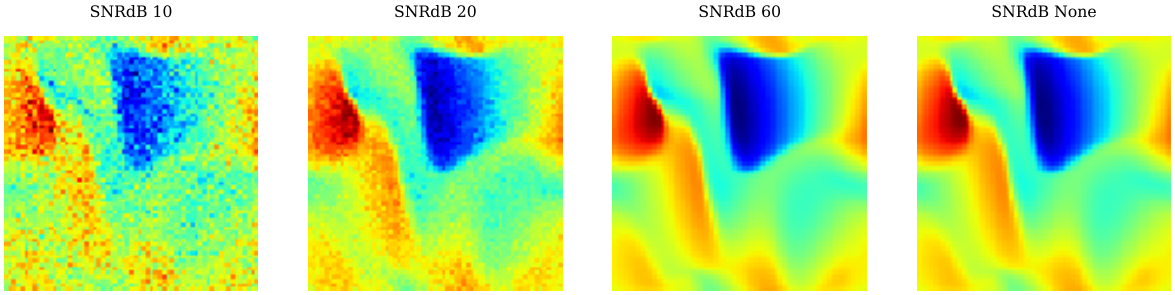


Figure 2: Plots of 2D Burgers velocity for visualizing different noise levels. SNRdB represents signal to noise ratio in decibels.

Table 1: Data-set parameters for 2D Burgers problem.

γ	M	L	z	ω	domain
5	2	50	2	3969	63×63

Table 2: Network architecture of proposed Ensers for 2D coupled Burgers' problem.

Layer	Input	Output	Activation
FC	8	64	Softplus
FC	64	64	Softplus
FC	64	$\omega * M * \gamma$	linear

Table 3: Training hyper parameters of Ensers for 2D Burgers problem.

η_o	η_{i0}	$\hat{\eta}_i$	β	I_o	I_i	ζ_0	$\hat{\zeta}$	p	h	N	ς
0.0002	0.1	0.006	11	2001	4	0.005	0.0001	32	800	22	8

Table 4: Testing hyper-parameters of Ensers for 2D Burgers problem. η_i: inner learning rate, I_i: inner iterations

η_i	I_i	p	$\hat{N}$
5	100	[4, 16]	12

FNN used in the model is a shallow network with three fully connected (FC) layers. We use Softplus [46] activation function which has smooth first derivatives. Thus it helps avoid discontinuities because unrolling the inference procedure involves computing $\nabla_\theta \nabla_\xi(\mathfrak{L})$. Network architecture is displayed in Table 2. We use 800 data nodes for training from each system variable data which is 20% of high-resolution data. Training hyper-parameters and training data are summarized in table 3. For testing, we use sensor locations different from the ones used during training.

3.2.2 Results and Discussions

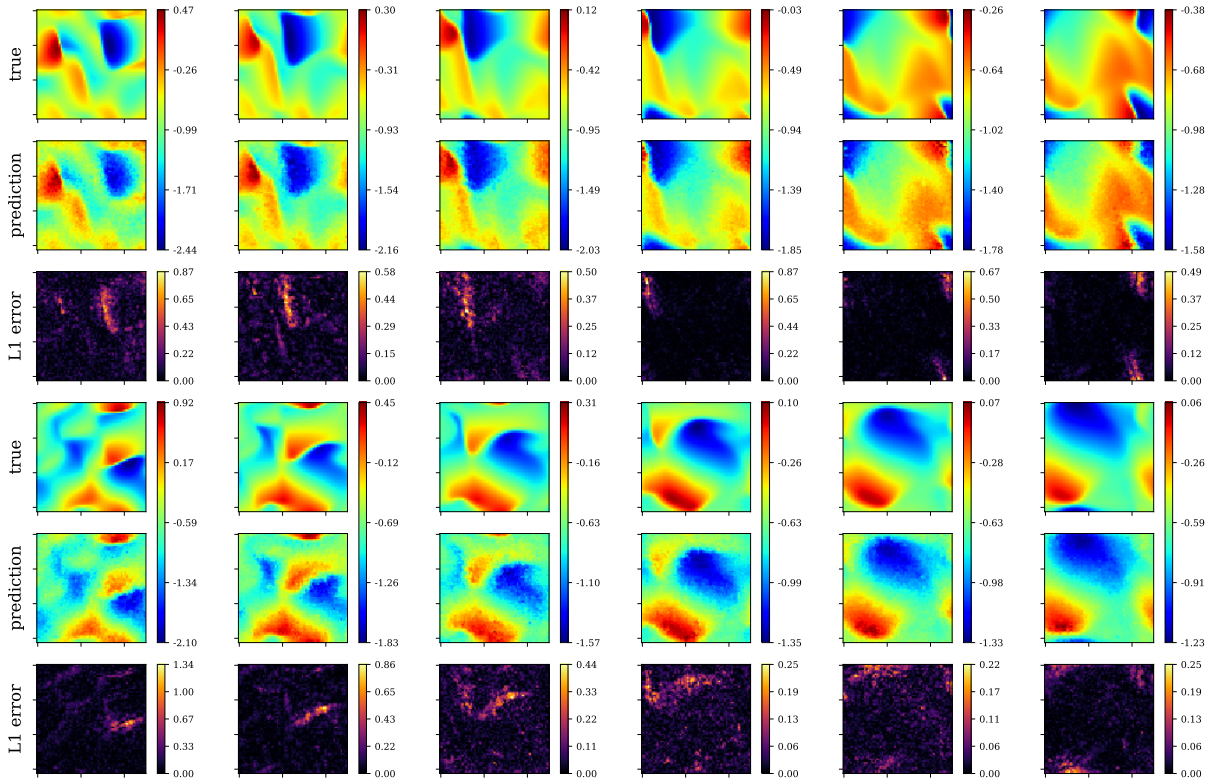


Figure 3: Ensers predictions of a 2D coupled Burgers' test case. (Top to bottom) x-velocity FEM target solution, x-velocity Ensers prediction, x-velocity L1error, y-velocity FEM target solution, y-velocity Ensers prediction and y-velocity L1error.

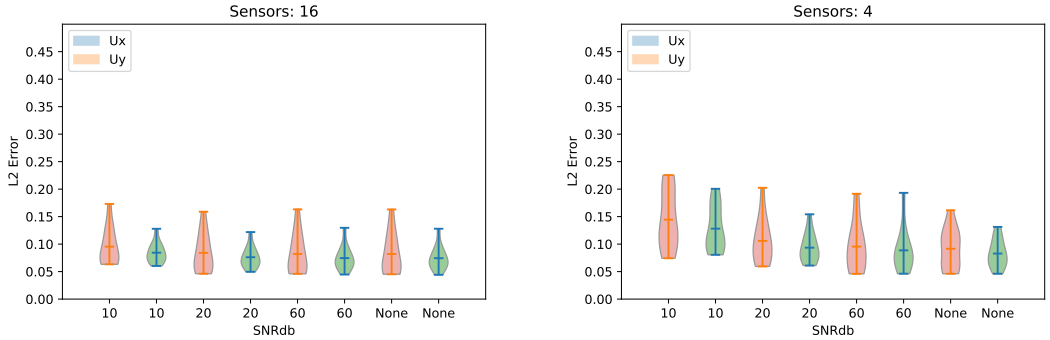
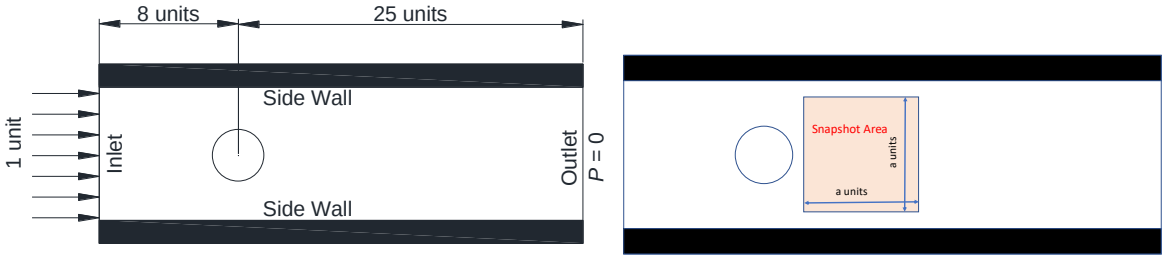


Figure 4: Violin plot representing error ε distribution in prediction vector of x and y velocity for different noise levels in sensor measurements for 2D coupled Burgers' test case.



(a) (b)

Figure 5: (a) Schematic representation of the computational domain with boundary conditions at the inlet and the outlet. The cylinder has a diameter of 1 unit. A no-slip boundary is considered at the cylinder wall. Zero pressure gradient at the inlet and zero velocity gradient at the outlet are considered. (b) Schematic of the problem domain with snapshot cutout of $a \times a$. For flow past cylinder problems, $a = 4$ units. *The schematics are not to scale.*

Fig. 3 shows the prediction of Ensers for x and y velocity components at various times in simulation with 16 sensors. We see that it produces results that accurately captures the current state of the system at most points in the domain. If a network is trained without a physics-based loss function then it will produce garbage values at points where data is absent. Fig 4 shows violin plot of error vector ε defined in Eq. (19) between target and prediction with various noise level in sensor measurement for test case. In Eq. (19), m for x velocity and y velocity are 0 and 1 respectively and $\gamma^* = 2$. The plot represents the distribution of error ε in the prediction vector. Greater spread corresponding to a point on the y-axis corresponds to more values present in the error vector ε around that value of the point on the y-axis. We see the model is robust to noise as the mean error in the plot are close for cases of no noise($SNR_{dB} = None$) and high noise($SNR_{dB} = 10$).

3.3 Experiment: Flow Past Cylinder

As the second example, we consider the flow past a cylinder problem. It is a well known canonical problem and is characterized by periodic laminar flow vortex shedding. System is governed by incompressible, laminar, Newtonian fluid equations:

$$\frac{\partial(u)}{\partial x} + \frac{\partial(v)}{\partial y} = 0 \tag{26a}$$

$$\frac{\partial(u)}{\partial t} + u\frac{\partial(u)}{\partial x} + v\frac{\partial(u)}{\partial y} = -\frac{1}{\rho}\frac{\partial(p)}{\partial x} + \nu\left(\frac{\partial^2(u)}{\partial x^2} + \frac{\partial^2(u)}{\partial y^2}\right) \tag{26b}$$

$$\frac{\partial(v)}{\partial t} + u\frac{\partial(v)}{\partial x} + v\frac{\partial(v)}{\partial y} = -\frac{1}{\rho}\frac{\partial(p)}{\partial y} + \nu\left(\frac{\partial^2(v)}{\partial x^2} + \frac{\partial^2(v)}{\partial y^2}\right) \tag{26c}$$

3.3.1 Data-set and Model Parameters

A schematic representation of the computational domain is shown in Fig. 5(a). The circular cylinder is considered to have a diameter of 1 unit. The center of the cylinder is located at a distance of 8 units from the inlet. The outlet is located at a distance of 25 units from the center of the cylinder. The sidewalls are at 4 units distance from the center of the cylinder. At the inlet boundary, a uniform velocity of 1 unit along the X-direction is applied. Pressure boundary condition with $P = 0$ is considered at the outlet. A no-slip boundary at the cylinder surface is considered. Coordinate of the snapshot cutout stretches from $[1.5, -2] \times [5.5, 2]$ which is discretized into 64×64 points in x and y directions (see Fig. 5(b)).

Table 5: Data-set parameters for Flow Past Cylinder problem.

γ	M	L	z	ω	ν	Re	domain
5	3	25	1	4096	0.005	200	64×64

The data-set is generated by using Unsteady Reynolds-averaged Navier Stokes (URANS) simulation in OpenFoam [47]. The overall problem domain is discretized into 63420 elements with finer mesh near the cylinder. Time step $\delta t = 0.02$

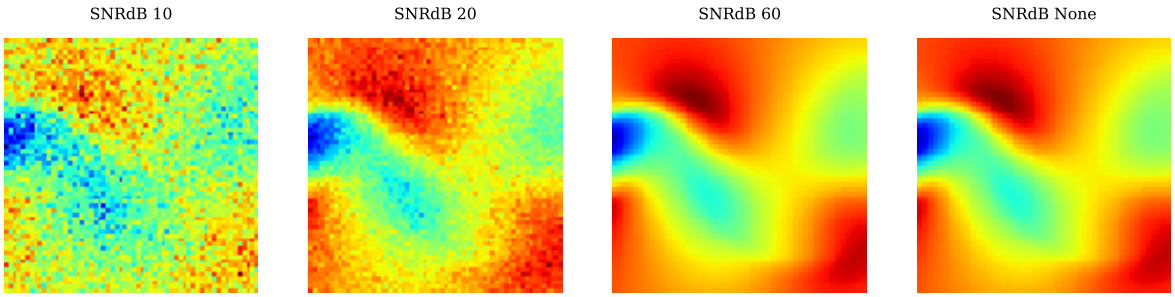


Figure 6: Plots of velocity of Flow Past Cylinder for visualizing different noise levels. SNRdB represents signal to noise ratio in decibels.

units is considered. Code for OpenFOAM simulation can be found at [48]. Parameters related to the data-set considered are displayed in Table 5.

We use a shallow network in the model with two fully connected(FC) layers. Network architecture is displayed in Table 6. We use 500 data nodes for training from each system variable data which is 12.2% of high-resolution data. Training and testing hyper-parameters are shown in table 7 and 8 respectively. Sensor locations for testing are different from the ones used during training. For this problem physics-based loss function is extended to include the continuity equation:

$$P(\chi^k) = \sum_{i=1}^{q+1} \|vec(W_i - V^n)\|_2^2 + \sum_{i=0}^{\gamma-1} \|vec(\frac{\partial(V_0^n)}{\partial x} + \frac{\partial(V_1^n)}{\partial y})\|_2^2 \tag{27}$$

Table 6: Network architecture of proposed Ensers for Flow Past Cylinder problem.

Layer	Input	Output	Activation
FC	8	64	Softplus
FC	64	$\omega * M * \gamma$	linear

Table 7: Training hyper parameters of Ensers for Flow Past Cylinder.

η_o	η_{i0}	$\hat{\eta}_i$	β	I_o	I_i	ζ_0	$\hat{\zeta}$	p	h	N	ς
0.0003	0.1	0.002	9	3001	5	0.01	0.0006	16	500	18	8

Table 8: Testing hyper parameters of Ensers for Flow Past Cylinder problem. η_i: inner learning rate, I_i: inner iterations

η_i	I_i	p	$\hat{N}$
5	100	[4, 16]	12

3.3.2 Results and Discussions

Fig. 7 shows prediction of Ensers for pressure, x and y velocity components with 16 sensors. We see that it produces results that accurately capture the current state of the system at most points in the domain. Fig. 8 shows violin plot of error defined in Eq. (19) between target and prediction with various noise levels in sensor measurement for the test case. In Eq. (19), m for x velocity, y velocity and pressure are 0, 1 and 2 respectively and $\gamma^* = 2$. The plot represents the distribution of error in prediction vectors. We see model is robust to noise as mean error in Fig 8 are close for cases of no noise($SNR_{dB} = None$) and high noise($SNR_{dB} = 10$).

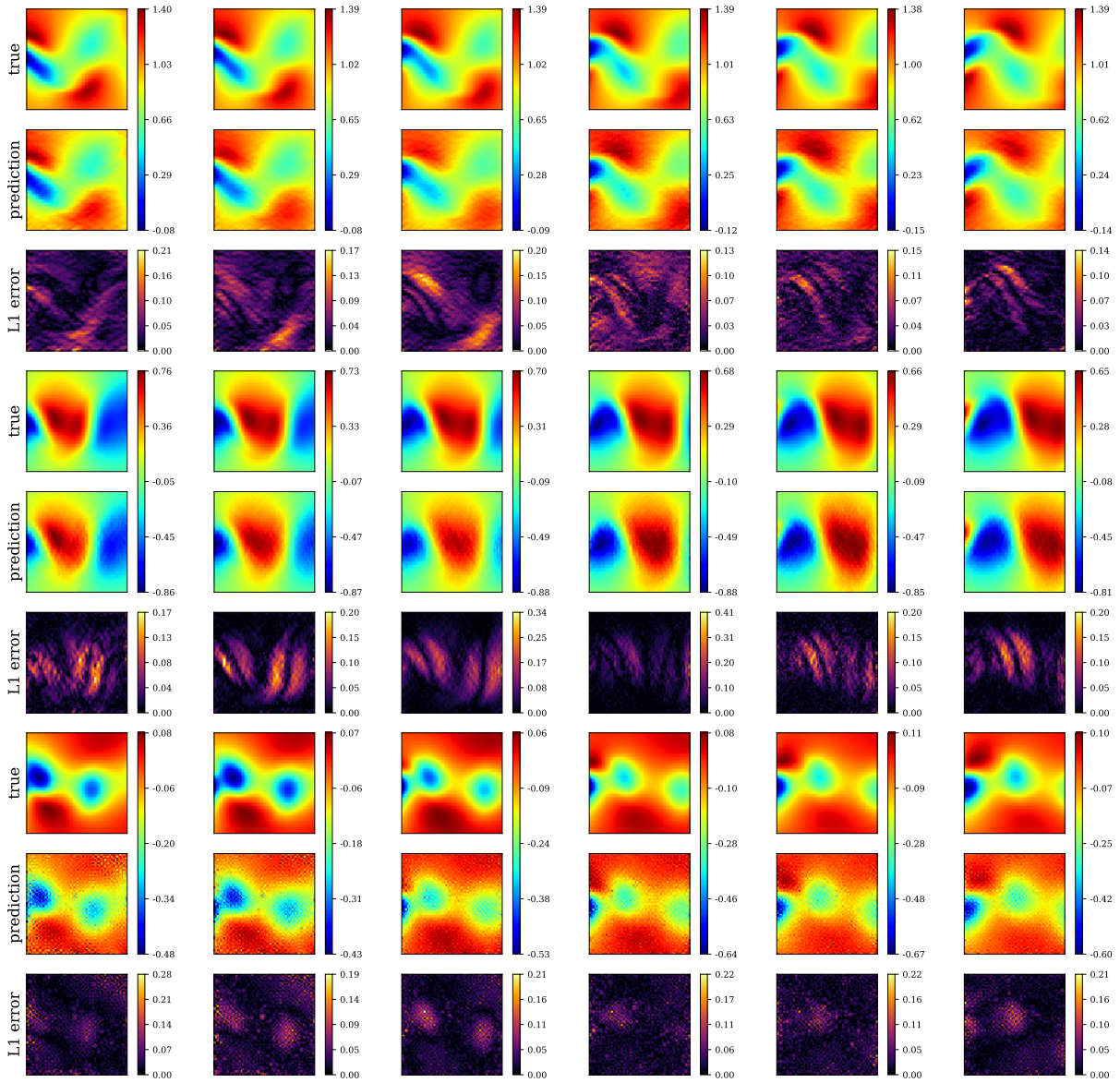


Figure 7: Ensers predictions of a Flow Past Cylinder test case. (Top to bottom) x-velocity target solution, x-velocity Ensers prediction, x-velocity L1error, y-velocity target solution, y-velocity Ensers prediction, y-velocity L1error, pressure target solution, pressure Ensers prediction and pressure L1error.

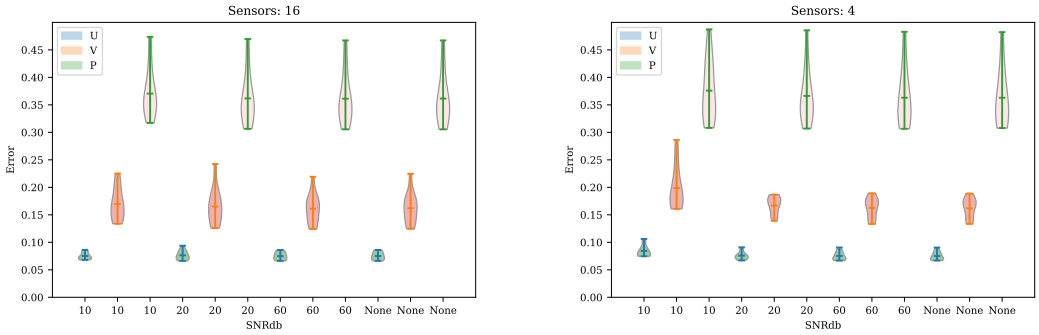
4 Continuous space models

4.1 Proposed approach

In this section, we propose a novel deep learning-based continuous framework for state estimation. Prediction of γ high dimensional states is done via multiple passes from a feed-forward neural network (FNN) Γ for every collocation point with coordinate vector X_r using optimized reduced state vectors $\xi^k \in \mathbb{R}^{\varsigma}$:

$$\Gamma(\Xi_r^k) = \{D_{mr}^{ki} \in \mathbb{R} \mid i = 0,...,\gamma-1,\ m = 0,...,M-1\}, \qquad r = 0,...,\omega \tag{28}$$

$$\Xi_r^k = \{\xi^k, X_r\}, \qquad r = 0,...,\omega \tag{29}$$



(a) (b)

Figure 8: Violin plot representing error ε distribution in prediction vector of x,y velocity and pressure for different noise levels in sensor measurements for Flow Past Cylinder test case.

where output vector of FNN Γ has dimension $\gamma * M$, D_{mr}^{ki} is predicted states value at r^{th} collocation point, X_r are coordinates of r^{th} collocation point. Output of ω FNNs are combined to form third-order tensor $D^k \in \mathbb{R}^{\gamma \times M \times \omega}$. Reduced state is obtained by solving the following minimization problem using sensor data χ^k and predicted states by the neural network.

$$\xi^k = \arg\min_{\tilde{\xi}^k} \|vec(\chi^k - \rho(\tilde{\Xi}^k))\|_2^2 \tag{30a}$$

$$\rho(\tilde{\Xi}^k) = \{\tilde{Q}_m^{ki} \in \mathbb{R}^p \mid i = 0, ..., \gamma - 1,\ m = 0, ..., M-1\} \tag{30b}$$

$$\tilde{Q}_m^{ki} = \Lambda_m^{kz+i} \tilde{D}_m^{ki} \tag{30c}$$

$$\Gamma(\tilde{\Xi}_r^k) = \{\tilde{D}_{mr}^{ki} \in \mathbb{R} \mid i = 0, ..., \gamma - 1,\ m = 0, ..., M-1\}, \quad r = 0, ..., \omega \tag{30d}$$

$$\tilde{\Xi}_r^k = \{\tilde{\xi}^k, X_r\}, \quad r = 0, ..., \omega \tag{30e}$$

where $\tilde{Q}_m^{ki} \in \mathbb{R}^p$ are values of predicted states at sensor locations λ_m^{kz+i}. Note that in practice ξ^k is obtained by a few steps of gradient descent instead of global minimization. The network is trained by minimizing data loss and physics-based loss $P(\chi^k)$ across training samples N.

$$\theta^* = \arg\min_{\theta} \sum_{k=0}^{N-1} \|vec(\phi^k - \Upsilon(\chi^k))\|_2^2 + P(\chi^k) \tag{31a}$$

$$\Upsilon(\chi^k) = \{Y_m^{ki} \in \mathbb{R}^h \mid i = 0, ..., \gamma - 1,\ m = 0, ..., M-1\} \tag{31b}$$

$$Y_m^{ki} = \Pi_m^{kz+i} D_m^{ki} \tag{31c}$$

where $D_m^{ki} \in \mathbb{R}^{\omega}$ are predicted states from Eq. (28), $Y_m^{ki} \in \mathbb{R}^h$ are values of predicted states at data locations π_m^{kz+i}, $\Upsilon(\chi^k) \in \mathbb{R}^{\gamma \times M \times p}$ are predicted states composed of γ time steps, θ are network parameters. Fig. 1 shows the network architecture during training. Equations (28) and (30) together forms the implicit optimization layer shown in the Fig. 1 for the continuous model. Training and testing procedure is shown in Algorithm 3 and 4 respectively.

Algorithm 3: Training ENSERS

1: **Inputs:** Z, S, T. {Eq. (2), Eq. (3), Eq. (4)}
2: **Set Hyper-parameters:** η_o: outer learning rate, η_{i0}: inner learning rate at epoch=0, $\hat{\eta}_i$: inner learning rate rate, β: batch size, I_o: outer iterations, I_i: inner iterations, $\hat{\zeta}$: physics penalty rate, ζ_0: physics penalty at epoch=0, N: number of train samples, z, γ.
3: **Calculate data-set:** χ, ϕ, Λ, Π
4: **Initialize:** Neural network model: $\Gamma(\cdot;\theta)$
5: **for** $\iota_o = 0$ **to** $I_o - 1$ **do** {Outer optimization Loop}
6: $\eta_i = \eta_{i0} + \iota_o \hat{\eta}_i$ {Schedule inner learning rate}
7: $\zeta = \zeta_0 + \iota_o \hat{\zeta}$ {Schedule physics penalty}
8: **for** $k = 0$ **to** $N - 1$ **do**
9: **Initialize:** $\tilde{\xi}^k$
10: **for** $\iota_i = 0$ **to** $I_i - 1$ **do** {Inner optimization Loop}
11: $\tilde{\Xi}_r^k = \{\tilde{\xi}^k, X_r\}, \quad r = 0, ..., \omega$
12: $\tilde{D}_r^k = \Gamma(\tilde{\Xi}_r^k), \quad r = 0, ..., \omega$ {Eq. (30d)}
13: $\tilde{Q}^k \leftarrow$ DrawValuesAtSensorLocations($\tilde{D}^k$) {Eq. (30c)}
14: $\mathfrak{L} = MSE(\chi^k, \tilde{Q}^k)$
15: $\frac{\partial \mathfrak{L}}{\partial \tilde{\xi}^k} \leftarrow$ Backprop($\mathfrak{L}$)
16: $\tilde{\xi}^k = \tilde{\xi}^k - \eta_i \frac{\partial \mathfrak{L}}{\partial \tilde{\xi}^k}$
17: **end for**
18: $\hat{X}_r = X_r$
19: $\Xi_r^k = \{\tilde{\xi}^k, \hat{X}_r\}, \quad r = 0, ..., \omega$
20: $D_r^k = \Gamma(\Xi_r^k), \quad r = 0, ..., \omega$ {Eq. (28)}
21: $Y^k \leftarrow$ DrawValuesAtDataLocations(D^k) {Eq. (31c)}
22: $\mathscr{L} = MSE(\phi^k, Y^k) + \zeta P(D^k)$ {Calculate loss}
23: $\frac{\partial \mathscr{L}}{\partial \theta} \leftarrow$ Backprop($\mathscr{L}$)
24: $\theta \leftarrow \theta - \eta_o \frac{\partial \mathscr{L}}{\partial \theta}$ {Update weights}
25: **end for**
26: **end for**
27: **Output:** Trained network $\Gamma(\cdot;\theta^*)$.

4.1.1 Physics-based loss function

The physics-based loss function for continuous space models differs from discrete formulation due to how RHS of Eq. (1) i.e. F is evaluated. In this case, we use automatic differentiation for calculating gradients w.r.t. coordinates used in F. For example in first case of Allen–Cahn equation $\hat{F}$ is evaluated as:

$$\hat{F}_r = 0.0001 \frac{\partial^2 u}{\partial \hat{X}_r^{\,2}} + 5u^3 - 5u, \quad r = 0, ..., \omega \tag{32}$$

where $\hat{X}_r$ is coordinate vector concatenated with optimized reduced state vector at end of inner optimization loop, see line 18 in algorithm (3). In second case of Convection-diffusion equation $\hat{F}$ is evaluated as:

$$\hat{F}_r = a(x,y) \frac{\partial u}{\partial \hat{X}_r^0} + b(x,y) \frac{\partial u}{\partial \hat{X}_r^1} + c \frac{\partial^2 u}{\partial \hat{X}_r^{0\,2}} + d \frac{\partial^2 u}{\partial \hat{X}_r^{1\,2}}, \quad r = 0, ..., \omega. \tag{33}$$

where $\hat{X}_r^0, \hat{X}_r^1$ are x and y coordinate respectively, a, b, c, d are defined in Eq. 36.

Algorithm 4: Testing ENSERS

Inputs: Trained network $\Gamma(\cdot;\theta^*)$, S. {Eq. (3)}
Set Hyper-parameters: η_i: inner learning rate, I_i: inner iterations, $\hat{N}$: number of test samples, z, γ.
Calculate data-set: χ, Λ
for $k = 0$ **to** $\hat{N} - 1$ **do**
  Initialize: $\tilde{\xi}^k$
  for $\imath_i = 0$ **to** $I_i - 1$ **do** {Inner optimization Loop}
    $\tilde{\Xi}^k = \{\tilde{\xi}^k, X_r\}, \quad r = 0, ..., \omega$
    $\tilde{D}_r^k = \Gamma(\tilde{\Xi}_r^k), \quad r = 0, ..., \omega$ {Eq. (30d)}
    $\tilde{Q}^k \leftarrow$ DrawValuesAtSensorLocations($\tilde{D}^k$) {Eq. (30c)}
    $\mathfrak{L} = MSE(\chi^k, \tilde{Q}^k)$
    $\frac{\partial \mathfrak{L}}{\partial \tilde{\xi}^k} \leftarrow$ Backprop($\mathfrak{L}$)
    $\tilde{\xi}^k = \tilde{\xi}^k - \eta_i \frac{\partial \mathfrak{L}}{\partial \tilde{\xi}^k}$
  end for
  $\Xi_r^k = \{\tilde{\xi}^k, X_r\}, \quad r = 0, ..., \omega$
  $D_r^k = \Gamma(\Xi_r^k), \quad r = 0, ..., \omega$ {Eq. (28)}
end for
Output: Predicted states $\{D_m^{ki} \mid k = 0, ..., N-1,\ i = 0, ..., \gamma - 1,\ m = 0, ..., M-1\}$.

4.2 Experiment: Allen–Cahn equation

We consider the Allen–Cahn equation along with periodic boundary conditions. The Allen–Cahn equation is a well-known equation from the area of reaction-diffusion systems. It describes the process of phase separation in multicomponent alloy systems, including order-disorder transitions.

$$\begin{aligned} u_t - 0.0001u_{xx} + 5u^3 - 5u = 0, \quad & x \in [-1,1], t \in [0,1], \\ u(0,x) &= x^2 cos(\pi x), \\ u(t,-1) &= u(t,1), \\ u_x(t,-1) &= u_x(t,1). \end{aligned} \tag{34}$$

4.2.1 Data-set and Model Parameters

Data-set is generated by simulating the Allen–Cahn equation (34) using conventional spectral methods. Starting from an initial condition $u(0,x) = x^2 cos(\pi x)$ and assuming periodic boundary conditions $u(t,-1) = u(t,1)$ and $u_x(t,-1) = u_x(t,1)$, we integrated Eq. (34) up to a final time $t = 1.0$ using the Chebfun package [49] with a spectral Fourier discretization with 512 modes and a fourth-order explicit Runge–Kutta temporal integrator with time-step $\Delta t = 10^{-5}$. For more details on the data-set see [41]. Plots depicting different noise levels used for sensor measurement during testing are shown in Fig. 9.

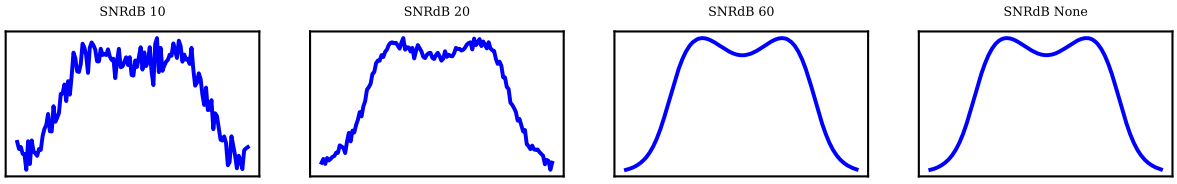


Figure 9: Data plots of Allen–Cahn equation for visualizing different noise levels. SNRdB represents signal to noise ratio in decibels.

The network considered in the continuous formulation has significantly fewer weights compared to the discrete one. This is because the output of the network is predicted at a single point. Network architecture is shown in table 10. The input size of the network is equal to the sum of the size of the reduced state ς and the dimension of coordinates. For training we use 10 data nodes p which is 7.81% of 128 total nodes ω. Similar to previous cases inner loop learning rate

Table 9: Data-set parameters for Allen–Cahn equation.

γ	M	L	z	ω	domain
5	1	50	1	128	128

η_i and physics loss penalty ζ are increased linearly with each epoch. Other training and testing hyperparameters are shown in table 11 and 12 respectively.

Table 10: Network architecture of proposed Ensers for Allen–Cahn equation.

Layer	Input	Output	Activation
FC	7	128	Tanh
FC	128	128	Tanh
FC	128	128	Tanh
FC	128	128	Tanh
FC	128	$M*\gamma$	linear

Table 11: Training hyper parameters of Ensers for Allen–Cahn equation.

η_o	η_{i0}	$\hat{\eta}_i$	β	I_o	I_i	ζ_0	$\hat{\zeta}$	p	h	N	ς
0.001	0.005	6e−5	10	1201	8	0.01	2e−5	10	10	40	6

Table 12: Testing hyper parameters of Ensers for Allen–Cahn equation. η_i: inner learning rate, I_i: inner iterations

η_i	I_i	p	$\hat{N}$
0.1	50	[6, 16]	12

4.2.2 Results and Discussions

Fig. 10 shows the prediction of Ensers with 16 sensors. A noticeable benefit of the continuous formulation is that prediction is smooth compared to discrete cases. Fig. 11 shows violin plot of L2 error defined in Eq. (19) between target and prediction with various noise levels in sensor measurement for the test case. In Eq. (19), $m = 0$ and $\gamma^* = 2$. Similar to discrete cases, the model is robust to noise in measurements. Fig. 11 shows a marginal increase in error with noise in the case of 16 sensors. Also, error bars with 6 sensors are similar to 16 sensors with low noise and increase slowly with noise.

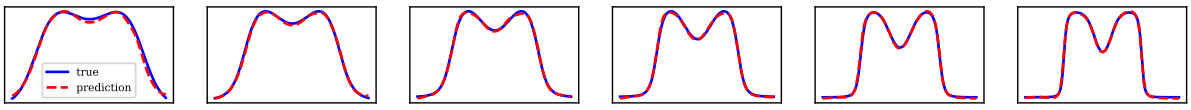


Figure 10: Ensers predictions of a Allen–Cahn equation test case with 16 sensors and no noise.

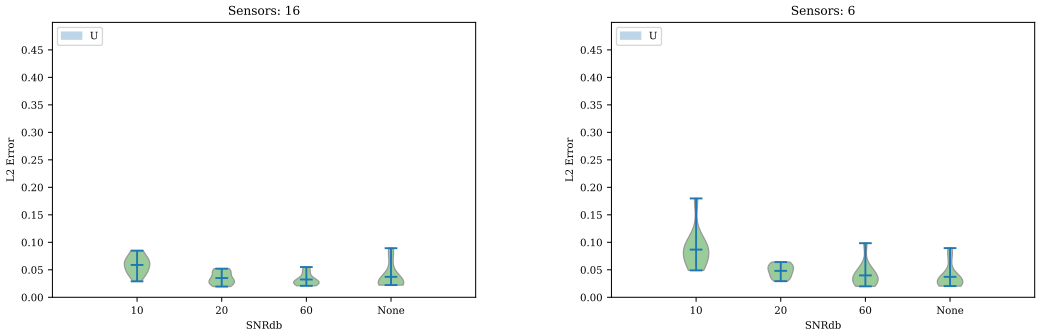
4.3 Experiment: Convection-diffusion equations

Next we consider a 2-dimensional linear variable-coefficient convection-diffusion equation on $\omega = [0, 2\pi] \times [0, 2\pi]$,

$$u_t = a(x,y)u_x + b(x,y)u_y + cu_{xx} + du_{yy} \ (t,x,y) \in [0, 0.2] \times \omega \tag{35}$$

$$\begin{aligned} a(x,y) &= 0.5(cos(y) + x(2\pi - x)sin(x)) + 0.6 \\ b(x,y) &= 2(cos(y) + sin(x)) + 0.8 \\ c &= 0.2, d = 0.3 \end{aligned} \tag{36}$$

Convection-diffusion equations are classical PDEs that are used to describe physical phenomena where particles, energy, or other physical quantities are transferred inside a physical system due to two processes namely diffusion



(a) (b)

Figure 11: Violin plot representing error ε distribution in prediction vector of data, for different noise levels in sensor measurements for Allen–Cahn equation test case.

and convection. These equations are widely applied in many scientific areas and industrial fields, such as pollutants dispersion in rivers or atmosphere, solute transferring in a porous medium, and oil reservoir simulation. We consider variables convection and diffusion coefficients Eq. (36) in this experiment.

4.3.1 Data-set and Model Parameters

Data is generated by solving the problem (35) using a high precision numerical scheme with a pseudo-spectral method for spatial discretization and 4th order Runge-Kutta for temporal discretization (with time step size $\delta t = 0.01$). We assume periodic boundary conditions and the initial value

$$u(x,y,0) = \sum_{|k|,|l| \leq N} \lambda_{k,l} cos(kx+ly) + \gamma_{k,l} sin(kx+ly) \tag{37}$$

where $N = 9, \lambda_{k,l}, \gamma_{k,l} \sim \mathcal{N}(0, 0.02)$, and k and l are chosen randomly. For more details on the data-set see [50]. Noisy data used for sensor measurement during testing is shown in Fig. 12 with various noise levels at a particular time. We use a simulation of length 40 as shown in table 13 along with other data parameters.

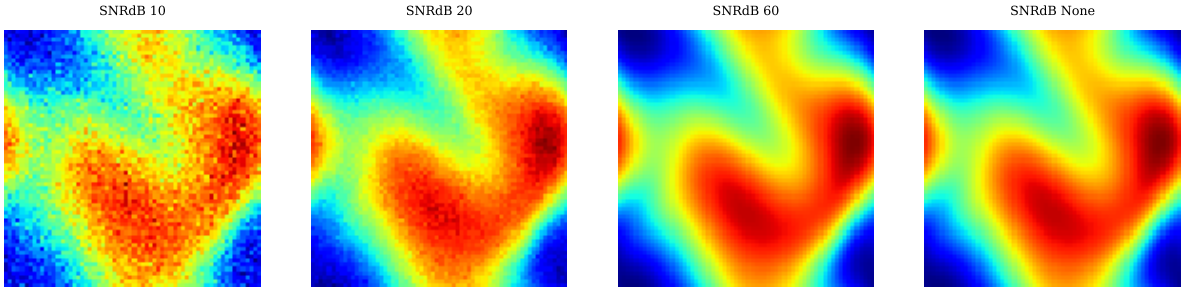


Figure 12: Data plots of Convection-diffusion equation for visualizing different noise levels. SNRdB represents signal to noise ratio in decibels.

Table 13: Data-set parameters for Convection-diffusion equation.

γ	M	L	z	ω	domain
5	1	40	1	4096	64×64

Network architecture is shown in table 14. Similar to the previous case a deep network with 5 layers and a small output size is used. The activation function is Tanh, which also has a continuous first derivative. For training, we use $p = 1024$

Table 14: Network architecture of proposed Ensers for Convection-diffusion equation.

Layer	Input	Output	Activation
FC	8	128	Tanh
FC	128	256	Tanh
FC	256	256	Tanh
FC	256	128	Tanh
FC	128	$M * \gamma$	linear

Table 15: Training hyper parameters of Ensers for Convection-diffusion equation.

η_o	η_{i0}	$\hat{\eta}_i$	β	I_o	I_i	ζ_0	$\hat{\zeta}$	p	h	N	ς
0.0005	0.005	5e−5	11	1001	10	0.005	1e−5	32	1024	33	6

Table 16: Testing hyper parameters of Ensers for Convection-diffusion equation. η_i: inner learning rate, I_i: inner iterations

η_i	I_i	p	$\hat{N}$
0.2	50	[4, 16]	12

training nodes i.e. 25% of total nodes. Training and testing hyperparameters are shown in tables k and ll respectively. Inner loop learning rate η_i and physics loss penalty ζ are increased linearly with each epoch.

4.3.2 Results and Discussions

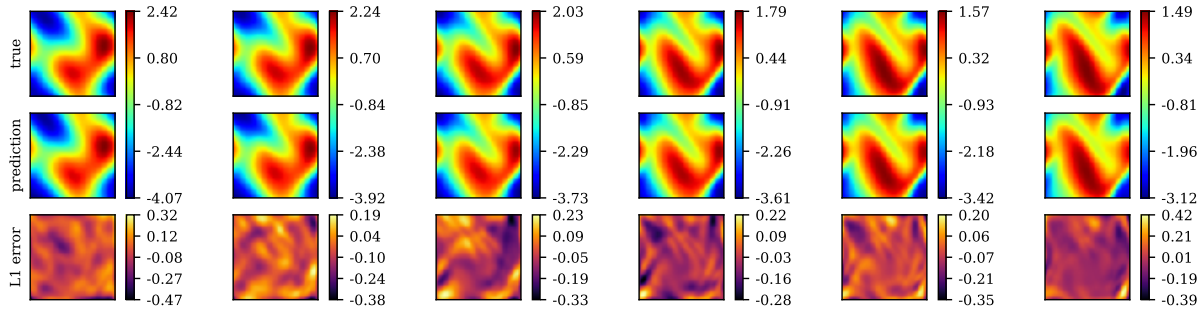
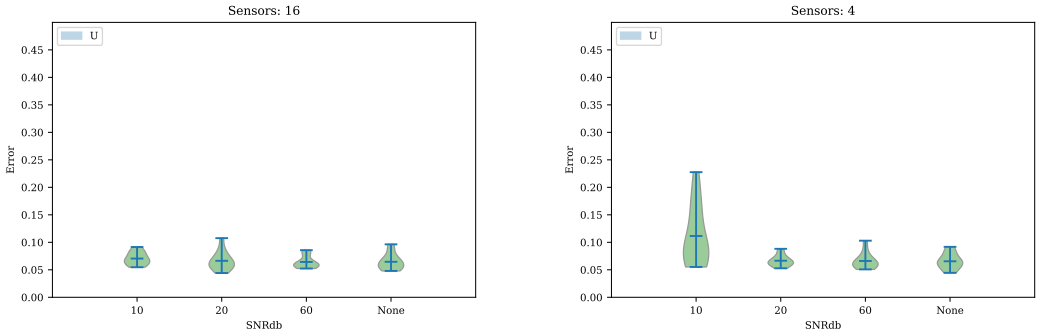


Figure 13: Ensers predictions of a Convection-diffusion equation test case. (Top to bottom) Target solution, Ensers prediction, L1error.

Fig. 13 shows the prediction of the Ensers with 16 sensors. The model is able to predict the state accurately and with little distortion. Fig. 14 shows violin plot of L2 error defined in Eq. (19) between target and prediction with various noise levels in sensor measurement for the test case. In Eq. (19), $m = 0$ and $\gamma^* = 2$ i.e. we use middle prediction out γ time steps. We see the model is robust to noise as the mean error in the plots are close for cases of no noise($SNR_{dB} = None$) and high noise($SNR_{dB} = 10$) for 16 sensors case and increases slightly noise for 4 sensor cases. Training continuous model for two dimensions takes more time than discrete because of the increased number of collocation points and multiple automatic differentiation required for calculating loss function.

5 Conclusions

In this work, we develop a novel technique to learn a deep learning model from spare moving training labels for which model reduction is nontrivial. The method uses an implicit optimization layer for minimizing the energy of the solution implemented through the technique of unrolled differentiation. We proposed two formulations based on this technique for discrete and continuous prediction in space. For learning from spare training labels we included a physics-based loss function calculated via convolutional filters in the discrete formulation and via automatic differentiation in the continuous formulation. Where most deep learning-based methods assume fixed sensors Ensers is capable of predicting full states given a varied number of sensors at random locations.



(a) (b)

Figure 14: Violin plot representing error ϵ distribution in prediction vector of x,y velocity and pressure for different noise levels in sensor measurements for Convection-diffusion equation test case.

We demonstrate the model performance using two-fluid problems of 2D coupled Burgers' equation and Flow Past Cylinder for the discrete case. For the continuous case, we used two problems namely Allen–Cahn equation and the Convection-diffusion equation. Model is shown to be robust against noisy sensor measurements. Future work can be aimed at quantifying uncertainty in such networks. Another direction can be to train networks for future states only using initial and boundary conditions.

Acknowledgements

SC acknowledges the financial support received from Science and Engineering Research Board (SERB) via project no. SRG/2021/000467 and Indian Institute Of Technology–Delhi in form of seed grant.

Reproducibility

The codes associated with the paper will be released on acceptance.

A Convolution operators for gradient and laplacian terms

Sobel Filter used to estimate 1st-order gradient is:

$$E = \begin{bmatrix} 1 & -8 & 0 & 8 & -1 \\ 2 & -16 & 0 & 16 & -2 \\ 3 & -24 & 0 & 24 & -3 \\ 2 & -16 & 0 & 16 & -2 \\ 1 & -8 & 0 & 8 & -1 \end{bmatrix} \tag{38a}$$

$$\frac{\partial}{\partial x} = E \times \frac{1}{9 * 12\delta x} \tag{38b}$$

Filter used to estimate laplacian is:

$$\frac{\partial}{\partial y} = E^T \times \frac{1}{9 * 12\delta y} \tag{38c}$$

$$\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} filter = \begin{bmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 16 & 0 & 0 \\ -1 & 16 & -60 & 16 & -1 \\ 0 & 0 & 16 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix} \times \frac{1}{12\delta x \delta y} \tag{39}$$

References

[1] L. Cordier, Springer Vienna, 2011. doi:10.1007/978-3-7091-0758-4_1.

[2] R. Semaan, M. Y. El Sayed, R. Radespiel, Springer International Publishing, Cham, 2019.

[3] S. Sankaran, M. Esmaily Moghadam, A. M. Kahn, E. E. Tseng, J. M. Guccione, A. L. Marsden, Patient-specific multiscale modeling of blood flow for coronary artery bypass graft surgery, Ann Biomed Eng 40 (2012) 2228–2242.

[4] A. Yakhot, T. Anor, G. E. Karniadakis, A reconstruction method for gappy and noisy arterial flow data, IEEE Trans Med Imaging 26 (2007) 1681–1697.

[5] G. Kissas, Y. Yang, E. Hwuang, W. R. Witschey, J. A. Detre, P. Perdikaris, Machine learning in cardiovascular flows modeling: Predicting arterial blood pressure from non-invasive 4d flow mri data using physics-informed neural networks, Computer Methods in Applied Mechanics and Engineering 358 (2020) 112623. URL: https://www.sciencedirect.com/science/article/pii/S0045782519305055. doi:https://doi.org/10.1016/j.cma.2019.112623.

[6] M. D. Graziano, M. D'Errico, G. Rufino, Ship heading and velocity analysis by wake detection in sar images, Acta Astronautica 128 (2016) 72–82. URL: https://www.sciencedirect.com/science/article/pii/S0094576516300303. doi:https://doi.org/10.1016/j.actaastro.2016.07.001.

[7] E. Kalnay, Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, 2002. doi:10.1017/CBO9780511802270.

[8] F. Ong, M. Uecker, U. Tariq, A. Hsiao, M. T. Alley, S. S. Vasanawala, M. Lustig, Robust 4d flow denoising using divergence-free wavelet transform, Magnetic Resonance in Medicine 73 (2015) 828–842. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.25176. doi:https://doi.org/10.1002/mrm.25176. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.25176.

[9] F. M. Callaghan, S. M. Grieve, Spatial resolution and velocity field improvement of 4D-flow MRI, Magn Reson Med 78 (2017) 1959–1968.

[10] M. F. Fathi, A. Bakhshinejad, A. Baghaie, D. Saloner, R. H. Sacho, V. L. Rayz, R. M. D'Souza, Denoising and spatial resolution enhancement of 4D flow MRI using proper orthogonal decomposition and lasso regularization, Comput Med Imaging Graph 70 (2018) 165–172.

[11] C. M. Bishop, Pattern recognition and machine learning, springer, 2006.

[12] K. Reif, S. Gunther, E. Yaz, R. Unbehauen, Stochastic stability of the discrete-time extended kalman filter, IEEE Transactions on Automatic control 44 (1999) 714–728.

[13] E. A. Wan, R. Van Der Merwe, S. Haykin, The unscented kalman filter, Kalman filtering and neural networks 5 (2001) 221–280.

[14] J. Tu, J. Griffin, A. Hart, C. Rowley, L. Cattafesta, L. Ukeiley, Integration of non-time-resolved piv and time-resolved velocity point sensors for dynamic estimation of velocity fields, Experiments in Fluids 54 (2012). doi:10.1007/s00348-012-1429-7.

[15] P. Schmid, J. Sesterhenn, Dynamic Mode Decomposition of numerical and experimental data, in: APS Division of Fluid Dynamics Meeting Abstracts, volume 61 of *APS Meeting Abstracts*, 2008, p. MR.007.

[16] C. W. ROWLEY, I. MEZIĆ, S. BAGHERI, P. SCHLATTER, D. S. HENNINGSON, Spectral analysis of nonlinear flows, Journal of Fluid Mechanics 641 (2009) 115–127. doi:10.1017/S0022112009992059.

[17] M. Buffoni, S. Camarri, A. Iollo, E. Lombardi, M. Salvetti, A non-linear observer for unsteady three-dimensional flows, Journal of Computational Physics 227 (2008) 2626 – 2643. URL: http://www.sciencedirect.com/science/article/pii/S0021999107004858. doi:https://doi.org/10.1016/j.jcp.2007.11.005.

[18] R. Adrian, On the role of conditional averages in turbulence theory., 1975.

[19] Y. G. Guezennec, Stochastic estimation of coherent structures in turbulent boundary layers, Physics of Fluids A: Fluid Dynamics 1 (1989) 1054–1060. URL: https://doi.org/10.1063/1.857396. doi:10.1063/1.857396. arXiv:https://doi.org/10.1063/1.857396.

[20] D. Ewing, J. H. Citriniti, Examination of a lse/pod complementary technique using single and multi-time information in the axisymmetric shear layer, in: J. N. Sørensen, E. J. Hopfinger, N. Aubry (Eds.), IUTAM Symposium on Simulation and Identification of Organized Structures in Flows, Springer Netherlands, Dordrecht, 1999, pp. 375–384.

[21] A. M. Naguib, C. E. Wark, O. Juckenhöfel, Stochastic estimation and flow sources associated with surface pressure events in a turbulent boundary layer, Physics of Fluids 13 (2001) 2611–2626. URL: https://doi.org/10.1063/1.1389284. doi:10.1063/1.1389284. arXiv:https://doi.org/10.1063/1.1389284.

[22] J. P. Bonnet, D. R. Cole, J. Delville, M. N. Glauser, L. S. Ukeiley, Stochastic estimation and proper orthogonal decomposition: Complementary techniques for identifying structure, Experiments in Fluids 17 (1994) 307–314. URL: https://doi.org/10.1007/BF01874409. doi:10.1007/BF01874409.

[23] J. Pinier, J. Ausseur, M. Glauser, H. Higuchi, Proportional closed-loop feedback control of flow separation, Aiaa Journal - AIAA J 45 (2007) 181–190. doi:10.2514/1.23465.

[24] R. Adrian, Conditional eddies in isotropic turbulence, Physics of Fluids 22 (1979). doi:10.1063/1.862515.

[25] T. C. Tung, R. J. Adrian, Higher-order estimates of conditional eddies in isotropic turbulence, The Physics of Fluids 23 (1980) 1469–1470. URL: https://aip.scitation.org/doi/abs/10.1063/1.863130. doi:10.1063/1.863130. arXiv:https://aip.scitation.org/doi/pdf/10.1063/1.863130.

[26] Y. Kumar, P. Bahl, S. Chakraborty, State estimation with limited sensors – a deep learning based approach, 2021. arXiv:2101.11513.

[27] N. B. Erichson, L. Mathelin, Z. Yao, S. L. Brunton, M. W. Mahoney, J. N. Kutz, Shallow learning for fluid flow reconstruction with limited sensors and limited data, 2019. arXiv:1902.07358.

[28] N. J. Nair, A. Goza, Leveraging reduced-order models for state estimation using deep learning, Journal of Fluid Mechanics 897 (2020) R1. doi:10.1017/jfm.2020.409.

[29] H. Gao, L. Sun, J.-X. Wang, Super-resolution and denoising of fluid flow using physics-informed convolutional neural networks without high-resolution labels, Physics of Fluids 33 (2021) 073603. URL: https://doi.org/10.1063/5.0054312. doi:10.1063/5.0054312. arXiv:https://doi.org/10.1063/5.0054312.

[30] E. Grefenstette, B. Amos, D. Yarats, P. M. Htut, A. Molchanov, F. Meier, D. Kiela, K. Cho, S. Chintala, Generalized inner loop meta-learning, arXiv preprint arXiv:1910.01727 (2019).

[31] B. Amos, Differentiable Optimization-Based Modeling for Machine Learning, Ph.D. thesis, Carnegie Mellon University, 2019.

[32] I. Goodfellow, M. Mirza, A. Courville, Y. Bengio, Multi-prediction deep boltzmann machines, Advances in Neural Information Processing Systems (2013).

[33] V. Stoyanov, A. Ropson, J. Eisner, Empirical risk minimization of graphical model parameters given approximate inference, decoding, and model structure, in: G. Gordon, D. Dunson, M. Dudík (Eds.), Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, volume 15 of *Proceedings of Machine Learning Research*, PMLR, Fort Lauderdale, FL, USA, 2011, pp. 725–733. URL: https://proceedings.mlr.press/v15/stoyanov11a.html.

[34] P. Brakel, D. Stroob, t, B. Schrauwen, Training energy-based models for time-series imputation, Journal of Machine Learning Research 14 (2013) 2771–2797. URL: http://jmlr.org/papers/v14/brakel13a.html.

[35] Y. Lecun, S. Chopra, R. Hadsell, A tutorial on energy-based learning, 2006.

[36] D. Utama, A. N., M. Iqbal, An optimal generic model for multi-parameters and big data optimizing: a laboratory experimental study, Journal of Physics: Conference Series 978 (2018) 012045. doi:10.1088/1742-6596/978/1/012045.

[37] D. Belanger, B. Yang, A. McCallum, End-to-end learning for structured prediction energy networks, 2017. arXiv:1703.05667.

[38] L. Metz, B. Poole, D. Pfau, J. Sohl-Dickstein, Unrolled generative adversarial networks (2016).

[39] M. J. Johnson, D. K. Duvenaud, A. Wiltschko, R. P. Adams, S. R. Datta, Composing graphical models with neural networks for structured representations and fast inference, in: D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, R. Garnett (Eds.), Advances in Neural Information Processing Systems, volume 29, Curran Associates, Inc., 2016. URL: https://proceedings.neurips.cc/paper/2016/file/7d6044e95a16761171b130dcb476a43e-Paper.pdf.

[40] C. Jordan-Squire, Convex optimization over probability measures, 2015.

[41] M. Raissi, P. Perdikaris, G. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational Physics 378 (2019) 686–707. doi:https://doi.org/10.1016/j.jcp.2018.10.045.

[42] Y. Zhu, N. Zabaras, P.-S. Koutsourelakis, P. Perdikaris, Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data, Journal of Computational Physics 394 (2019) 56–81. URL: https://www.sciencedirect.com/science/article/pii/S0021999119303559. doi:https://doi.org/10.1016/j.jcp.2019.05.024.

[43] N. Geneva, N. Zabaras, Modeling the dynamics of pde systems with physics-constrained deep auto-regressive networks, Journal of Computational Physics 403 (2020) 109056. URL: https://www.sciencedirect.com/science/article/pii/S0021999119307612. doi:https://doi.org/10.1016/j.jcp.2019.109056.

[44] I. Sobel, G. Feldman, A 3×3 isotropic gradient operator for image processing, Pattern Classification and Scene Analysis (1973) 271–272.

[45] A. Logg, K.-A. Mardal, G. N. Wells, et al., Automated Solution of Differential Equations by the Finite Element Method, Springer, 2012. doi:10.1007/978-3-642-23099-8.

[46] X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier neural networks, in: G. Gordon, D. Dunson, M. Dudík (Eds.), Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, volume 15 of *Proceedings of Machine Learning Research*, PMLR, Fort Lauderdale, FL, USA, 2011, pp. 315–323. URL: https://proceedings.mlr.press/v15/glorot11a.html.

[47] H. Jasak, Openfoam: open source cfd in research and industry, International Journal of Naval Architecture and Ocean Engineering 1 (2009) 89–94.

[48] R. S. Nithin Adidela, Y. Sudhakar, Laminar flow over a circular cylinder simulation with openfoam v7, https://github.com/nithinadidela/circular-cylinder, ????

[49] T. A. Driscoll, N. Hale, L. N. Trefethen, Chebfun guide, 2014.

[50] Z. Long, Y. Lu, X. Ma, B. Dong, PDE-net: Learning PDEs from data, in: J. Dy, A. Krause (Eds.), Proceedings of the 35th International Conference on Machine Learning, volume 80 of *Proceedings of Machine Learning Research*, PMLR, 2018, pp. 3208–3216. URL: https://proceedings.mlr.press/v80/long18a.html.

arXiv:2203.12943v2 [quant-ph] 26 Mar 2022

EULER-RODRIGUES PARAMETERS: A QUANTUM CIRCUIT TO CALCULATE RIGID-BODY ROTATIONS

TECHNICAL REPORT

Emilio Peláez
The University of Chicago
Chicago, Illinois, 60637, USA
epelaez@uchicago.edu

Anuranan Das
Department of Electrical Engineering
Indian Institute of Technology Bombay
Powai, Mumbai, 400076, India
anuranan@ee.iitb.ac.in

Parmeet Singh Chani
Department of Applied Physics
Delhi Technological University
Bawana, Delhi, 110042, India
parmeetsinghchani_2k19ep066@dtu.ac.in

Daniel Sierra-Sosa
Department of Computer Science & IT
Hood College
401 Rosemont Ave, Frederick, MD 21701, USA
sierra-sosa@hood.edu

March 29, 2022

ABSTRACT

The use of vectorial parameterization to create geometrical representations in computational models has a large number of applications. One particular application is the calculation of the 3D rotational motion of rigid bodies, that could be used for the spatial location estimation from objects. Provided the algebraic nature of this problem, it could benefit from Quantum Computing, in particular several vectors could be superposed to be transformed with a single operation, providing a quantum processing advantage. In this article, we propose an implementation of a Quantum Computing algorithm to compute Euler-Rodrigues Parameters to model rigid body rotations to transform arbitrary functions, rotating multiple vectors in superposition. We developed this algorithm using Qiskit, taking into account the limitations imposed by the current Noisy Intermediate Scale Quantum (NISQ) devices, such as the reduced number of qubits available and the limited coherence time.

Keywords Quantum Computing · Linear Algebra Transformation · 3D Rotation

1 Introduction

Quantum mechanics postulates are of algebraic nature, which means there exists an intrinsic relation between quantum computation and algebraic operations [1]. Multiple advances in the field of quantum information processing have provided promising prospects relying on that advantage. Therefore, it has been proven that quantum computing could lead to exponential speed-up in different fields, including Principal Component Analysis (PCA) [2], k-means clustering [3] and regularized Support Vector Machines (SVM) [4]. Advances in NISQ devices imply the development of more diverse and meaningful applications, increasing the relevance of conducting research in this area.

One of the applications that could profit from quantum computing advantages is the vector parametrization, which relies on algebraic operations to model real-world phenomena. These parameters are commonly used in computer vision, enabling applications on engineering and applied science such as the rotational motion of rigid bodies and the estimation of 3-D rigid body transformations. Some special cases of the vector parameterization are the mathematical models including Rodrigues rotation, Cayley transform, Euler-Rodrigues equations, Gibbs, Wiener, and Milenkovic parameters [5]. All of these are characterized by a set of three parameters which behave as the Cartesian components of a geometric vector in a three-dimensional space.

Euler-Rodrigues Parameters (ERP) have received less attention in the literature than their isomorphs, quaternions, because ERP in general require more computational resources to be calculated. Taking aside the computational burden, ERP provide a simpler model interpretation and allows to obtain an invariant rotation matrix. In an invariant matrix the quantities change according to a well-known model, a transformation in linear algebra. Quantum computing could enable the calculation of these parameters given the exponential speed-up of some linear algebra operations. This work aims to study the feasibility of using quantum computing circuits to compute ERP in particular rigid body rotations, considering the limitations imposed by the current NISQ devices such as the reduced number of qubits available and the limited coherence time.

2 Background and notation

In this section, we will introduce the necessary basic concepts about Euler-Rodrigues parameters, rotations, and quantum computation.

2.1 Rodrigues' rotation matrix

The matrix of a proper rotation R by angle θ around the axis $u = (u_x, u_y, u_z)$, a vector with $u_x^2 + u_y^2 + u_z^2 = 1$, is given by:

$$R = \begin{bmatrix} \cos\theta + u_x^2(1-\cos\theta) & u_x u_y(1-\cos\theta) - u_z\sin\theta & u_x u_z(1-\cos\theta) + u_y\sin\theta \\ u_y u_x(1-\cos\theta) + u_z\sin\theta & \cos\theta + u_y^2(1-\cos\theta) & u_y u_z(1-\cos\theta) - u_x\sin\theta \\ u_z u_x(1-\cos\theta) - u_y\sin\theta & u_z u_y(1-\cos\theta) + u_x\sin\theta & \cos\theta + u_z^2(1-\cos\theta) \end{bmatrix} \tag{1}$$

The basic idea to derive this matrix is dividing the problem into a few known simple steps:

1. First rotate the given axis and the point such that the axis lies in one of the coordinate planes (xy, yz or zx).
2. Then rotate the given axis and the point such that the axis is aligned with one of the two coordinate axes for that particular coordinate plane (x, y or z).
3. Use one of the fundamental rotation matrices (about the x-, y-, or z-axis) to rotate the point depending on the coordinate axis with which the rotation axis is aligned.
4. Reverse rotate the axis-point pair such that it attains the final configuration as that was in the second step, in other words, undo the second step.
5. Reverse rotate the axis-point pair which was done in the first step; undo step one.

Mathematically, the above procedure can be written as:

$$R = (\cos\theta)\, I + (\sin\theta)\, [\mathbf{u}]_\times + (1-\cos\theta)\,(\mathbf{u}\otimes\mathbf{u}) \tag{2}$$

where

$$\mathbf{u}\otimes\mathbf{u} = \mathbf{u}\mathbf{u}^\mathsf{T} = \begin{bmatrix} u_x^2 & u_x u_y & u_x u_z \\ u_x u_y & u_y^2 & u_y u_z \\ u_x u_z & u_y u_z & u_z^2 \end{bmatrix}, \qquad [\mathbf{u}]_\times = \begin{bmatrix} 0 & -u_z & u_y \\ u_z & 0 & -u_x \\ -u_y & u_x & 0 \end{bmatrix} \tag{3}$$

In $\mathbb{R}^3$ the rotation of a vector $\vec{x}$ around the axis $\mathbf{u}$ by an angle θ can be written as:

$$R_{\mathbf{u}}(\theta)\vec{x} = \mathbf{u}(\mathbf{u}\cdot\vec{x}) + \cos(\theta)(\mathbf{u}\times\vec{x})\times\mathbf{u} + \sin(\theta)(\mathbf{u}\times\vec{x}) \tag{4}$$

2.2 Euler-Rodrigues parameters

Three-dimensional rotations can be represented by a variety of parametrizations each suitable for applications in different domains [6]. The Euler-Rodrigues parametrization provides a way of performing a rotation as a Cartesian transformation. A central rotation about a rotation axis for a point at x' is represented by four real numbers a, b, c, d as

$$\vec{x}' = \begin{pmatrix} a^2+b^2-c^2-d^2 & 2(bc-ad) & 2(bd+ac) \\ 2(bc+ad) & a^2+c^2-b^2-d^2 & 2(cd-ab) \\ 2(bd-ac) & 2(cd+ab) & a^2+d^2-b^2-c^2 \end{pmatrix}\vec{x}, \tag{5}$$

subject to the constraint $a^2+b^2+c^2+d^2 = 1$. The rotation can be uniquely determined by the rotation angle θ_r and axis of rotation (n_x, n_y, n_z) by the following equations for a, b, c and d.

$$a = \cos(\theta_r/2) \quad b = n_x\sin(\theta_r/2) \quad c = n_y\sin(\theta_r/2) \quad d = n_z\sin(\theta_r/2) \tag{6}$$

2.3 Basic quantum operations

A qubit is the fundamental unit of information in quantum computation. This abstract object can take on states in a Hilbert space spanned by two basis vectors $|0\rangle$ and $|1\rangle$ which correspond to the column vectors $(1 \quad 0)^\mathsf{T}$ and $(0 \quad 1)^\mathsf{T}$, respectively. Thus, the general state of a qubit is $\alpha|0\rangle + \beta|1\rangle$ with $\alpha, \beta \in \mathbb{C}$ such that $|\alpha|^2 + |\beta|^2 = 1$. When measured, qubits collapse to one of the basis states, which are referred to as the computational basis states. The probabilities of collapsing to $|0\rangle$ and $|1\rangle$ are $|\alpha|^2$ and $|\beta|^2$, respectively. Hence the need for the constraint that these two quantities add up to unity.

To act on qubits, we apply operators which can be represented as unitary matrices. Some of the most common single-qubit gates are the rotation gates, which can be obtained by taking the exponentials of the Pauli matrices. The Pauli matrices are defined as shown below.

$$X = \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{7}$$

In some cases, the identity matrix is considered to be the zeroth Pauli matrices, defined as

$$I = \sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{8}$$

Applying these gates to a qubit works as matrix multiplication. Given a general state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, we can apply the X gate to it by doing $X|\psi\rangle$; which is simply matrix-vector multiplication. We can clearly see that $X(\alpha|0\rangle + \beta|1\rangle) = \beta|0\rangle + \alpha|1\rangle$ thanks to the columns of X. Note that the X gate flips the amplitudes of the two basis states, so it is commonly referred to as the bit-flip gate. The effect of a gate will not always be this easy to figure out, but it always boils down to matrix-vector multiplication.

Say you have two unitary matrices U_0 and U_1. If you want to apply U_0 followed by U_1 to the state $|\psi\rangle$, you would compute $U_1U_0|\psi\rangle$. Then, you could proceed by doing two matrix-vector multiplications or one matrix-matrix multiplication and one matrix-vector multiplication. Basically, the unitary matrices and statevectors follow the usual rules of linear algebra.

With the Pauli matrices, we can now define the rotation gates. These represent a rotation of θ radians about one of the three axes in the Bloch sphere, which helps visualize one qubit quantum states. These gates are the following.

$$R_x(\theta) = \exp\left(-iX\frac{\theta}{2}\right) = \begin{pmatrix} \cos(\theta/2) & -i\sin(\theta/2) \\ -i\sin(\theta/2) & \cos(\theta/2) \end{pmatrix} \tag{9}$$

$$R_y(\theta) = \exp\left(-iY\frac{\theta}{2}\right) = \begin{pmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{pmatrix} \tag{10}$$

$$R_z(\theta) = \exp\left(-iZ\frac{\theta}{2}\right) = \begin{pmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{pmatrix} \tag{11}$$

These gates are fundamental to quantum computation since they are the easily implementable in current hardware. An important property about these gates is that their inverse is defined by a rotation about the same axis but with a negative angle. This is to say that $R_i^{-1}(\theta) = R_i^\dagger(\theta) = R_i(-\theta)$. Also, having two or more consecutive rotations about the same axis is equivalent to a single rotation about that axis with the angle being the total sum of the angles of the individual rotations. It should be noted that the Pauli matrices follow the equality:

$$X^2 = Y^2 = Z^2 = -iXYZ = I \tag{12}$$

Thanks to the fact that each Pauli matrix squared is equal to the identity, we can expand the exponential in the equations for the rotation operators as $\exp(-i\sigma_i\theta/2) = \cos(\theta/2)I - i\sin(\theta/2)\sigma_i$. Equation (12) also gives us a blueprint to calculate the product of any pair of Pauli matrices. To do this take the last identity ($-iXYZ = I$) and do the necessary left and right multiplications until the right hand side of the equation is the pair (or triplet) you want to compute. Additionally, a general single-qubit unitary U can be expressed as $e^{i\delta}R_z(\alpha)R_y(\theta)R_z(\beta)$ [7], where the first term represents a global phase that becomes relevant when working with multi-qubit systems.

3 Rotation around an arbitrary axis

3.1 Rotation scheme

A general single-qubit state can be defined as $|\psi\rangle = e^{i\delta}\left(e^{-i\phi/2}\cos(\theta/2)|0\rangle + e^{i\phi/2}\sin(\theta/2)|1\rangle\right)$, where $e^{i\delta}$ is an undetectable global phase and we can interpret the other two angles as spherical coordinates of the form $(1, \theta, \phi)$ [8].

Which in turn represents the point in the surface of the Bloch sphere corresponding to the state $|\psi\rangle$. Thinking about this state in geometrical terms, we know that it is θ radians away from the unit vector in the positive z-axis direction and ϕ radians away from the x-axis unit vector.

Therefore, if we want to transform the $|\psi\rangle$ state into $|0\rangle$, which has spherical coordinates $(1, 0, 0)$, we can do so by applying $R_z(-\phi)$ followed by $R_y(-\theta)$. This procedure can be visually represented with rotations in the Bloch sphere. First, as shown in Figure 1, a rotation around the z-axis places the statevector $|\psi\rangle$ in the xz plane.

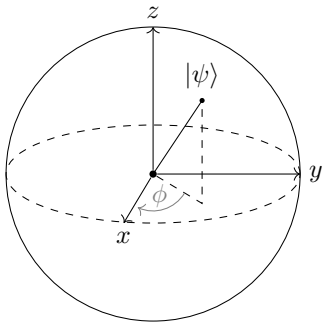


Figure 1: Rotating $-\phi$ radians around the z-axis

After this rotation, the new state is $|\psi'\rangle = e^{i\delta}\left(\cos(\theta/2)|0\rangle + \sin(\theta/2)|1\rangle\right)$ which has spherical coordinates $(1, \theta, 0)$. Then, the $R_y(-\theta)$ gate rotates $|\psi'\rangle$ about the y-axis to finally place it in the north pole of the Bloch sphere.

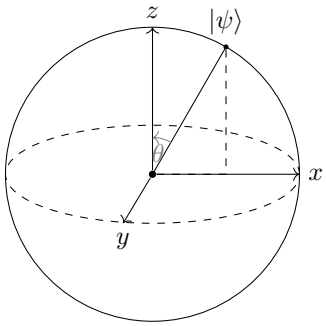


Figure 2: Rotating $-\theta$ radians around the y-axis (x and y axis exchanged for clarity)

Thus, the final state after these two rotations is $|\psi''\rangle = e^{i\delta}|0\rangle$, which is the usual ground state up to a negligible global phase. In mathematical terms, the procedure goes as follows.

$$
\begin{aligned}
&R_y(-\theta)R_z(-\phi)|\psi\rangle \\
&\quad = R_y(-\theta)R_z(-\phi)\left[e^{i\delta}\left(e^{-i\phi/2}\cos\left(\theta/2\right)|0\rangle + e^{i\phi/2}\sin\left(\theta/2\right)|1\rangle\right)\right] \\
&\quad = e^{i\delta}R_y(-\theta)\left(e^{i\phi/2}e^{-i\phi/2}\cos\left(\theta/2\right)|0\rangle + e^{-i\phi/2}e^{i\phi/2}\sin\left(\theta/2\right)|1\rangle\right) \\
&\quad = e^{i\delta}R_y(-\theta)\left(\cos\left(\theta/2\right)|0\rangle + \sin\left(\theta/2\right)|1\rangle\right) \\
&\quad = e^{i\delta}\left[\cos\left(\theta/2\right)\left(\cos\left(-\theta/2\right)|0\rangle + \sin\left(-\theta/2\right)|1\rangle\right) + \sin\left(\theta/2\right)\left(\cos\left(-\theta/2\right)|1\rangle - \sin\left(-\theta/2\right)|0\rangle\right)\right] \\
&\quad = e^{i\delta}\left[\cos(\theta/2)\cos\left(-\theta/2\right)|0\rangle - \sin\left(\theta/2\right)\sin\left(-\theta/2\right)|0\rangle\right] \\
&\quad = e^{i\delta}\cdot 1/2\cdot\left[(\cos(\theta)+1)|0\rangle - (\cos(\theta)-1)|0\rangle\right] \\
&\quad = e^{i\delta}|0\rangle
\end{aligned}
\tag{13}
$$

This allows us to define a way of rotating an arbitrary quantum state around an arbitrary axis $\vec{n} = (n_x, n_y, n_z)$ with rotating angle θ_r, thinking in terms of the Bloch sphere. First, of all, we are going to assume the vector $\vec{n}$ is normalized and thus we can write it in spherical coordinates as $(1, \theta_n, \phi_n)$. Given this, we can define a scheme for rotation around $\vec{n}$ as follows:

1. Perform the necessary rotations so $(1, \theta_n, \phi_n)$ is now at $(1, 0, 0)$. This is done by applying $R_z(-\phi_n)$ followed by $R_y(-\theta_n)$.
2. Now that $\vec{n}$ is at $|0\rangle$, we can perform our desired rotation by doing $R_z(\theta_r)$. This works since the original rotation axis is now aligned with the z-axis.
3. Perform rotations to get everything to its original spot, in other words, do the inverse of step 1. This is done by applying $R_y(\theta_n)$ followed by $R_z(\phi_n)$.

In geometric terms, this can be interpreted as translating the rotation axis into the z-axis, performing the desired rotation about the z-axis, and finally bringing the original rotation axis back to its original spot. The complete operation for this procedure is $R_z(\phi_n)R_y(\theta_n)R_z(\theta_r)R_y(-\theta_n)R_z(-\phi_n)$.

3.2 Euler-Rodrigues parameters from circuit unitary

The unitary operation corresponding to the scheme described above can also be written as $R_{\vec{n}}(\theta_r) = \exp(-i \cdot \theta_r/2 \cdot \vec{n} \cdot \vec{\sigma})$, where σ is a vector of Pauli matrices. Expanding the exponent gives us the following expression for the unitary:

$$\begin{aligned} R_{\vec{n}}(\theta_r) = \exp\left(-i \cdot \frac{\theta_r}{2} \cdot \vec{n} \cdot \vec{\sigma}\right) &= \cos\left(\frac{\theta_r}{2}\right) I - i \sin\left(\frac{\theta_r}{2}\right) \vec{n} \cdot \vec{\sigma} \\ &= \cos\left(\frac{\theta_r}{2}\right) I - i \sin\left(\frac{\theta_r}{2}\right) (n_x\sigma_1 + n_y\sigma_2 + n_z\sigma_3) \end{aligned} \tag{14}$$

We can show that equation (14) is equal to the five gate sequence $R_z(\phi_n)R_y(\theta_n)R_z(\theta_r)R_y(-\theta_n)R_z(-\phi_n)$ by using the identities introduced in equations (9), (10) and (11). We start by expanding the expression for the unitary in the middle, which gives us:

$$R_z(\phi_n)R_y(\theta_n)\left(\cos\left(\frac{\theta_r}{2}\right) I - \sin\left(\frac{\theta_r}{2}\right) Z\right) R_y(-\theta_n)R_z(-\phi_n) \tag{15}$$

Now, we are going to briefly show that $R_y(\theta)ZR_y(-\theta) = \cos(\theta)Z + \sin(\theta)X$ by using Pauli products that can be derived from equation (12).

$$\begin{aligned} R_y(\theta)ZR_y(-\theta) &= \left(\cos\left(\frac{\theta}{2}\right) I - i \sin\left(\frac{\theta}{2}\right) Y\right) Z \left(\cos\left(\frac{\theta}{2}\right) I + i \sin\left(\frac{\theta}{2}\right) Y\right) \\ &= \cos^2\left(\frac{\theta}{2}\right) Z + i \cos\left(\frac{\theta}{2}\right) \sin\left(\frac{\theta}{2}\right) ZY - i \cos\left(\frac{\theta}{2}\right) \sin\left(\frac{\theta}{2}\right) YZ + \sin^2\left(\frac{\theta}{2}\right) YZY \\ &= \left(\cos^2\left(\frac{\theta}{2}\right) - \sin^2\left(\frac{\theta}{2}\right)\right) Z + 2\cos\left(\frac{\theta}{2}\right) \sin\left(\frac{\theta}{2}\right) X \\ &= \cos(\theta)Z + \sin(\theta)X \end{aligned} \tag{16}$$

The first step was to expand the rotation matrices using the generalized Euler formula. Then, we carried matrix multiplication to get rid of the parentheses. After that, we used the Pauli pair identities to factorize the expression and finally we were able to use regular trigonometric identities to get the result we wanted. Using the same outline, we can prove that $R_z(\theta)ZR_z(-\theta) = Z$ and $R_z(\theta)XR_z(-\theta) = \cos(\theta)X + \sin(\theta)Y$. And remember that $R_i(\theta)R_i(\phi) = R_i(\theta + \phi)$. We can now continue with our original goal. Using the identities stated above, we can continue from equation (15) as follows:

$$\begin{aligned} &R_z(\phi_n)R_y(\theta_n)\left(\cos\left(\frac{\theta_r}{2}\right) I - \sin\left(\frac{\theta_r}{2}\right) Z\right) R_y(-\theta_n)R_z(-\phi_n) \\ &\quad = R_z(\phi_n)\left[\cos\left(\frac{\theta_r}{2}\right) I - i \sin\left(\frac{\theta_r}{2}\right) (\cos(\theta_n)Z + \sin(\theta_n)X)\right] R_z(-\phi_n) \\ &\quad = \cos\left(\frac{\theta_r}{2}\right) I - i \sin\left(\frac{\theta_r}{2}\right) [\cos(\theta_n)Z + \sin(\theta_n)(\cos(\phi_n)X + \sin(\phi_n)Y)] \\ &\quad = \cos\left(\frac{\theta_r}{2}\right) I - i \sin\left(\frac{\theta_r}{2}\right) (\sin(\theta_n)\cos(\phi_n)X + \sin(\theta_n)\sin(\phi_n)Y + \cos(\theta_n)Z) \end{aligned} \tag{17}$$

This expression already looks very similar to (14), and in fact is equivalent. The coefficients in front of the Pauli matrices are equal to the equations for converting spherical coordinates to Cartesian coordinates, with the radius r set to

1 (which is our case thanks to the normalization of $\vec{n}$). Thus, we have proved that the five-qubit sequence we use to rotate in the Bloch sphere is equal to the rotation operator obtained by exponentiation.

From the background, recall that a rotation of θ radians about an axis $\vec{n}$ can be defined in terms of the Euler-Rodrigues parameters a, b, c and d as defined in [6]. We can extract these from the unitary that corresponds to our circuit. To do this, first remember that the Pauli matrices (including the identity) form a basis for the 2×2 Hermitian matrices. So, any unitary operation can be written as follows:

$$U = \alpha\sigma_1 + \beta\sigma_2 + \gamma\sigma_3 + \delta I \tag{18}$$

Remember that the trace of a square matrix is defined as the sum of elements in its main diagonal. Thus, we have the following properties:

$$\mathrm{Tr}(U\sigma_1)/2 = \alpha \quad \mathrm{Tr}(U\sigma_2)/2 = \beta \quad \mathrm{Tr}(U\sigma_3)/2 = \gamma \quad \mathrm{Tr}(U)/2 = \delta \tag{19}$$

From expanding the parenthesis in (14), we can easily see that α, β, γ and δ for the rotation we defined correspond directly to the parameters a, b, c and d that we are looking to extract. Applying the properties in (19) to the unitary of $R_{\vec{n}}(\theta_r)$, we get the following relations:

$$\begin{aligned} \delta = \cos(\theta_r/2) = a \quad & \alpha = -in_x \sin(\theta_r/2) = -ib \\ \beta = -in_y \sin(\theta_r/2) = -ic \quad & \gamma = -in_z \sin(\theta_r/2) = -id \end{aligned} \tag{20}$$

Therefore, the information that describes a rotation using the Euler-Rodrigues formula can be extracted from the proposed circuit with minimal post-processing. However, for this to be completely accurate we rely on simulators or assume a noise-free quantum device.

3.3 Tomography for parameter extraction

The procedure described on the previous subsection to extract the rotation parameters assumes complete access to the circuit unitary, which is unrealistic when running experiments on a real device. When working on real devices, we can perform process tomography [9] to characterize our circuit's process using the Choi-matrix representation. Then, we obtain the Kraus operators through the eigenvalue decomposition of the Choi-matrix $C = USV^\dagger$, where U and V are unitary matrices and S is a diagonal matrix. For the one qubit case, which we consider here, C is a 4×4 matrix. If the process is completely-positive, as in this case, the left and right Kraus operators are identical. So, we compute the Kraus operators as

$$K_i = \sqrt{s_{ii}} \begin{bmatrix} u_{1i} & u_{3i} \\ u_{2i} & u_{4i} \end{bmatrix}. \tag{21}$$

If all the values of S are zero except for one, then the only non-zero Kraus operator satisfies $K_i^\dagger K_i = K_i K_i^\dagger = I$ and therefore is unitary. In this case, that single operator is also the unitary operator we are looking to reconstruct. To obtain the data needed for this reconstruction we create and run $4^n 3^n$ circuits a large amount of times each. That amount of circuits is needed because they consist of preparing each of the four minimal Pauli basis eigenstates ($|0\rangle, |1\rangle, |+\rangle, |+i\rangle$), sending them through the circuit and measuring in each of the three Pauli basis.

For testing purposes, we generated twenty random rotations and ran process tomography on the corresponding rotation circuits with twenty logarithmically spaced numbers of shots between two hundred and twenty thousand. This was run in a simulator and two real quantum devices. After reconstructing the unitary through tomography, we calculate the gate fidelity between that and the exact unitary and plot the results in Figure 3.

4 Rotating a vector

4.1 Initializing and rotating arbitrary vectors

Using the procedure described on the last section combined with an encoding procedure, we can rotate any normalized vector. Suppose we have an input vector with spherical coordinates $(1, \theta_i, \phi_i)$ and we want to rotate θ_r radians about $\vec{n}$. Doing this is pretty simple: first we need to initialize a single qubit to the quantum state corresponding to $(1, \theta_i, \phi_i)$ which can be done by applying two rotation gates, and then we apply the rotation scheme described before. What we are left with is the quantum state with spherical coordinates $(1, \theta_f, \phi_f)$, up to a global phase.

In the last section, we looked into the unitary matrix corresponding to the quantum circuit in order to extract useful information about the Euler-Rodrigues parameters. In this case, we want to gain information about the rotated vector and not the rotation matrix.

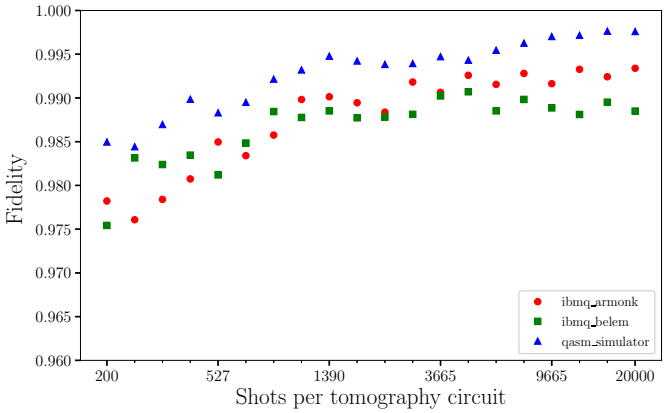
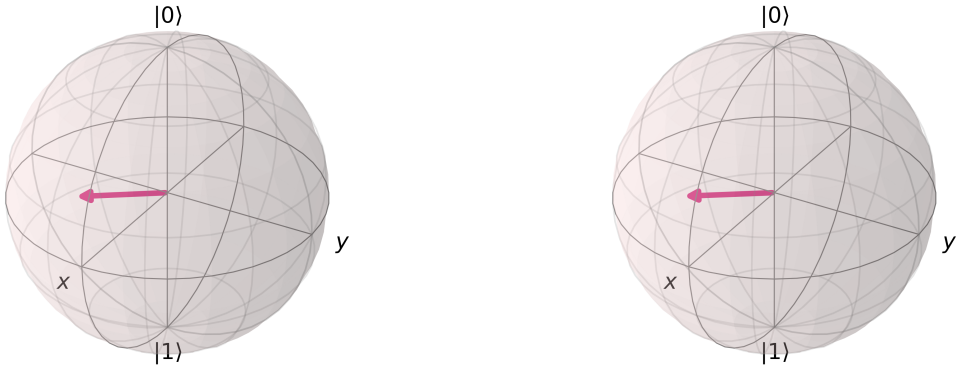


Figure 3: Fidelity of reconstructed unitary against shots used in process tomography



(a) Classical calculation of the Euler-Rodrigues matrix.

(b) Quantum rotation circuit and tomography simulation.

Figure 4: Result of a $\pi/4$ radians rotation of $[1, 0, 1]$ about $[1, 0.5, 0.5]$ (both vectors normalized to unit lenght). Figures generated with Qiskit [11].

4.2 Extracting rotated vector

To extract the rotated vector, we use a version of the quantum state tomography procedure introduced in [10]. The procedure given in this paper is quite simple and for a single qubit needs only three measurement basis: X, Y, and Z. To obtain high fidelity results, we run the circuit a large amount of times for each of these three measurement basis. The measurement results allow us to apply the algorithm introduced by Smolin to get the maximum likelihood density matrix. We then compute the eigenvalues λ_i and eigenvectors $|\lambda_i\rangle$ of this density matrix so we can get the statevector $|\psi\rangle = \sum_i \lambda_i |\lambda_i\rangle$. Finally, we can plot this state in the Bloch sphere and compare it to the vector we obtain rotating classically to confirm they are equivalent as showed in Figure 4. Visually, this is a really good indicator that the vector obtained through tomography is the desired rotated vector.

A more precise metric we can use to confirm the accuracy of our method is the angle between the classical and quantum rotated vector, which becomes smaller as the number of shots in the tomography is increased. After converting both vectors into rectangular coordinates, we can use the well-known formula for the angle between two vectors:

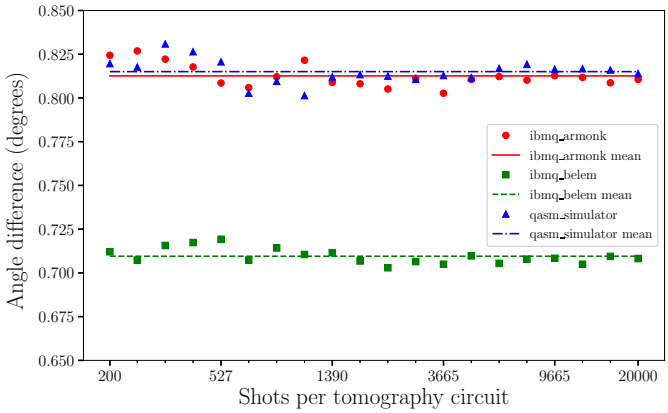


Figure 5: Angle difference between classical and quantum rotation against shots used in tomography

$\cos(\theta) = (\vec{x} \cdot \vec{y})/(|\vec{x}||\vec{y}|)$. We use the same randomly generated rotations and devices as in Section 3.3 to test state tomography on our rotation scheme, using the angle between the extracted vector and the classically calculated vector as a metric. These results are showed in Figure 5.

5 Multi-qubit rotations

5.1 Encoding initial vectors

To apply the same rotation to multiple vectors, we resort to using $n+1$ qubits, which will allow us to rotate up to 2^n vectors at once. We use n qubits as control qubits and the remaining qubit will act as a data qubit in which we will store all the vectors before and after rotating them. To illustrate our procedure, we can describe the case where $n = 2$.

First, we are going to put the n control qubits in an equal superposition by applying a Hadamard to each qubit. This is equivalent to applying $H^{\otimes n}$ and for $n = 2$ gives us the state below.

$$(H^{\otimes 2}|00\rangle)|0\rangle = (|00\rangle + |01\rangle + |10\rangle + |11\rangle)|0\rangle \tag{22}$$

As you can see, the data qubit is left undisturbed in this first step and we have an equal superposition of the control qubits.

Then, we take the n-controlled versions of the R_y and R_z gates. As explained in the single-qubit case, these rotation gates are used to encode a vector and then rotate it. The controlled versions are going to help us encode a vector for each basis state of the n control qubits. Since we have one vector for each basis state of a n-qubit register, we are able to encode 2^n vectors.

To go ahead with the encoding, first we calculate the rotation angles $\{\theta_i, \phi_i\}$ for each i-th vector where $1 \leq i \leq 2^n$ just as described in the single-qubit case. Then, we apply the controlled gates with the control qubits surrounded by X gates depending on the value of i and with the data qubit as target. When $i = 1$, no X gates are used. When $i = 2$, an X gate in the first qubit is used. And the pattern continues: the X gates are used to encode the binary representation of $i - 1$. A circuit for our example would look as follows:

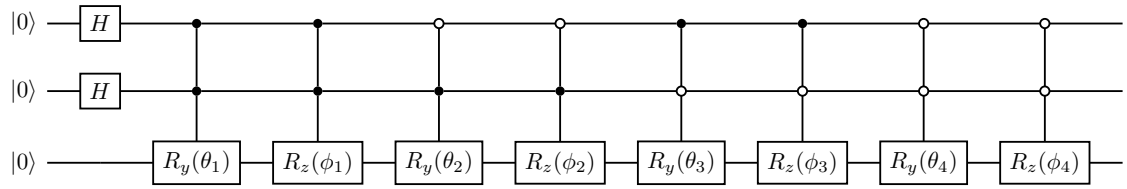


Figure 6: Encoding 4 vectors using 2 control qubits

In this figure, white controls are normal controls surrounded by X gates, i.e., they are triggered by $|0\rangle$ instead of $|1\rangle$. The state of the circuit above evolves as

$$\begin{aligned}
|00\rangle|0\rangle &\xrightarrow{H^{\otimes 2}} (|00\rangle + |01\rangle + |10\rangle + |11\rangle)|0\rangle = |00\rangle|0\rangle + |01\rangle|0\rangle + |10\rangle|0\rangle + |11\rangle|0\rangle \\
&\xrightarrow{C_{11}R_zC_{11}R_y} |00\rangle|0\rangle + |01\rangle|0\rangle + |10\rangle|0\rangle + |11\rangle|\vec{v}_4\rangle \\
&\xrightarrow{C_{10}R_zC_{10}R_y} |00\rangle|0\rangle + |01\rangle|0\rangle + |10\rangle|\vec{v}_3\rangle + |11\rangle|\vec{v}_4\rangle \\
&\xrightarrow{C_{01}R_zC_{01}R_y} |00\rangle|0\rangle + |01\rangle|\vec{v}_2\rangle + |10\rangle|\vec{v}_3\rangle + |11\rangle|\vec{v}_4\rangle \\
&\xrightarrow{C_{11}R_zC_{11}R_y} |00\rangle|\vec{v}_1\rangle + |01\rangle|\vec{v}_2\rangle + |10\rangle|\vec{v}_3\rangle + |11\rangle|\vec{v}_4\rangle
\end{aligned} \tag{23}$$

where $|\vec{v}_i\rangle$ is the i-th vector encoded as $R_z(\phi_i)R_y(\theta_i)|0\rangle$. For general n, the encoding procedure would look as follows.

$$|0\rangle^{\otimes n}|0\rangle \xrightarrow{H^{\otimes n}} \sum_{i=1}^{2^n} |i-1\rangle|0\rangle \tag{24}$$

$$\xrightarrow{C_iR_zC_iR_y} \sum_{i=1}^{2^n} |i-1\rangle|\vec{v}_i\rangle \tag{25}$$

5.2 Rotating the superposition

Now, we can apply the usual single-qubit rotation scheme to the data qubit such that it rotates all of the qubits we encoded. Remember that this scheme corresponds to applying the gate sequence $R_z(\phi_n)R_y(\theta_n)R_z(\theta_r)R_y(-\theta_n)R_z(-\phi_n)$, where we are rotating θ_r radians about an axis with spherical coordinates $(1, \theta_n, \phi_n)$. For simplicity, we can compact these gates into a single operator which we will call $R(\theta_n, \phi_n, \theta_r)$. Applying this operator to the state we left off with in equation (23), we get the state

$$\begin{aligned}
(I \otimes I \otimes R)(|00\rangle|\vec{v}_1\rangle + |01\rangle|\vec{v}_2\rangle + |10\rangle|\vec{v}_3\rangle + |11\rangle|\vec{v}_4\rangle) &= |00\rangle R|\vec{v}_1\rangle + |01\rangle R|\vec{v}_2\rangle + |10\rangle R|\vec{v}_3\rangle + |11\rangle R|\vec{v}_4\rangle \\
&= |00\rangle|\vec{r}_1\rangle + |01\rangle|\vec{r}_2\rangle + |10\rangle|\vec{r}_3\rangle + |11\rangle|\vec{r}_4\rangle
\end{aligned} \tag{26}$$

where we use $\vec{r}_i$ to denote the rotated vectors. For general n, the transformation above would look as follows.

$$(I^{\otimes n} \otimes R) \sum_{i=1}^{2^n} |i-1\rangle|\vec{v}_i\rangle = \sum_{i=1}^{2^n} |i-1\rangle|\vec{r}_i\rangle \tag{27}$$

If we want to rotate the encoded vectors by different angles and/or about different vectors, we would need to divide R into two parts: $R_t(\theta_n, \phi_n)$ which corresponds to the initial and final translation (hence the subindex t), and the regular rotation about the z-axis $R_z(\theta_r)$. This way, we could control these operations on the basis states of the control qubits to perform different rotations to the initial vectors.

Either way, if we perform the same rotation to all vectors or have any number of different rotations, at the end we end up with a state of the form showed in equation (26). The general structure of the data we encoded first doesn't change, but the encoded data is what gets modified. This is convenient as it allows us to extract a desired rotated vector following a very similar procedure to the single-qubit case.

5.3 Post-selecting control qubits

To extract a specific vector, we need to collapse the control qubits to the state corresponding to its index. Remember that the i-th vector corresponds to the binary representation of $i-1$ in the control qubits. If we measure the control

qubits as they stand in equation (26) and post-select, we would only get our desired basis state with $1/2^n$ probability, which is clearly not ideal.

To increase the probability of measuring the state we want and therefore carry on a successful post-selection, we could apply a standard amplitude amplification procedure. More specifically, we would like to mark those states that have our desired index in the control qubits. This could be easily done with a phase oracle consisting of only Z gates. Although the oracle is inexpensive, we would need to apply the initialization procedure with each iteration of the oracle, i.e., $\left\lfloor \frac{\pi}{4} 2^{n/2} \right\rfloor$ times [12] with n being the number of control qubits. Moreover, the initialization circuit uses 2^{n+1} n-controlled rotation gates, which are expensive to decompose.

Sequences of consequent controlled rotation gates with different rotation axis, just as in the circuit in figure 6, require $O(2^n)$ nearest-neighbor CNOTs when decomposed [13]. If all the controlled rotations about the $y-$axis are done before the ones about the z-axis (or vice-versa), there is a linear improvement in this decomposition, but the leading term remains exponential. Therefore, an amplitude amplification procedure to increase the success probability of post-selection is expensive with large n.

Knowing this, a first instinct would be trying to perform amplitude amplification only on the control qubits, so that the initialization operation would simply be a layer of Hadamard gates. However, entanglement between the control qubits and data qubit forbids such procedure. To see this, suppose Alice has the control qubits and Bob the data qubit and these two parties are physically apart from each other. If amplitude amplification acting only on the control qubits to achieve our goal was possible, Alice could use this to force the data qubit into a state that conveys a message to Bob, allowing faster than light communication.

Thus, we are left with either an expensive amplitude amplification algorithm or having to deal with low success probability. A possible alternative to this problem is to store the rotated vectors in a QRAM structure [14] that would allow easier access to the states stored, at the expense of having to use more spacial resources. Because of the resource constraint, this option may be more well suited for the post-NISQ era. Regardless of the post-selection strategy chosen to collapse the control qubits into the desired index, state or process tomography has to be applied to the state of the data qubit as described in earlier sections of this paper.

5.4 Application

This multi-vector rotation scheme could be used to perform rotations of three-dimensional rigid bodies. To do this, follow the encoding and rotation circuit introduced earlier this section using $n = \lceil \log_2 p \rceil$ where p is the number of points in the body. The resulting state would contain the rotated points in the data qubit, which would need to be extracted. Due to the probabilistic nature of quantum circuits and measurement techniques, the resulting points would be an approximation of the exact rotation. The accuracy of this approximation depends on the number of shots used for the tomography.

Since we want to extract all rotated vectors in order to reconstruct the rotated body and not just some specific vectors, the considerations discussed previously about post-selection are not as important in this case. Say we decide to use k shots for the tomography of each vector in order to achieve some arbitrary accuracy. In total, we will need $k \cdot p$ shots of the circuit. This quantity can be upper bounded by $k \cdot 2^n$, which happens when p is a power of 2.

Instead of needing to run an amplitude amplification procedure in order to increase the probability of a successful post-selection, we can simply run the circuit $k \cdot 2^n$ times, collapsing the control qubits without any further manipulation. Since the control qubits are in a uniform superposition, i.e., we measure each with probability $1/2^n$, we will get each state approximately $(k \cdot 2^n) \cdot (1/2^n) = k$ times, just as we need. Although we don't get exactly k samples of each state, small deviations in a large number of shots k do not affect the fidelity too much, as shown in figures 5 and 3.

It is a good time to discuss a small nuance when we talk about shots in a tomography procedure. Throughout this paper, the number of shots stated for some tomography, refer to the shots in a single circuit of said tomography. This is done since process and state tomography require different number of circuits: $3^n 4^n$ and 3^n where n is the number of qubits, respectively. So, for the whole tomography, the total number of shots is obtained by multiplying by these factors. Since we only perform tomography on single-qubit states, this only constitutes a linear increase in the total number of shots, therefore the asymptotic behavior remains the same.

Another important point to address is that rotating and extracting multiple points using this multi-qubit scheme does not present an advantage over using the single-qubit procedure for each point. However, it may present itself useful when we do not wish to extract each single point, but learn something about their structure without having to know the exact state of each point.

6 Application Example of Single Qubit Rotation

In this section, we set out to demonstrate an example of the single qubit rotation scheme as described above.

6.1 Vector Considerations

Let the axis of rotation be represented by the line passing through origin and another point $\vec{b}$ as $\vec{r} = \vec{0} + \mathrm{t}.(\vec{b} - \vec{0}) = t.\vec{b}$. In this example we consider $\vec{b} = (2, 1, 1)$. Hence the axis when represented by a unit vector will be given as $\vec{n} = \hat{b} = \frac{1}{\sqrt{6}}(2, 1, 1)$ Let the unit vector to be rotated be chosen arbitrarily as $\vec{x} = \frac{1}{\sqrt{3}}(1, 1, 1)$ and the angle of rotation be considered as $90°$. Our goal is to find the rotated vector $\vec{x}'$ with the help of a quantum circuit. In sync with the notation in other sections, the considerations is summarised as follows:

$$\vec{n} = \frac{1}{\sqrt{6}}(2, 1, 1) \qquad \vec{x} = \frac{1}{\sqrt{3}}(1, 1, 1) \qquad \theta_r = 90° \tag{28}$$

6.2 Initialisation of Circuit to $\vec{x}$ on Bloch Sphere

We initialise the circuit parameters to represent $\vec{x}$ on the bloch sphere. Here the angles θ and ϕ described in Section 3.1 are 0.9553 and $\pi/4$ radians respectively. Thus the illustration in Figure 7 depicts the circuit for initialization.

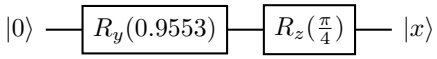


Figure 7: Initialization Circuit as per equation 28

6.3 Creating the Rotation Unitary

The θ and ϕ values for our chosen axis are 1.15 and 0.4636 radians respectively. Thus as discussed in earlier sections, to align this axis to the Z axis we need to rotate by $-\theta$ and $-\phi$. Now that our axis is aligned to the Z-axis, we rotate the vector by $\theta_r = 90°$ as parameter to a R_z gate to arrive at the rotated vector in this transformed vector space. To get the original rotated vector, we undo the earlier rotation by restoring the axis to earlier position. Fig. 8 depicts the circuit to carry out these three stages.

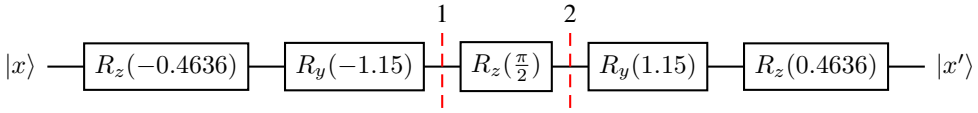


Figure 8: Illustration of transformation of initiated $|x\rangle$ to rotated $|x'\rangle$ by the quantum circuit scheme

It is also noteworthy that we can find the Euler-Rodrigues parameters solely from the unitary created here by using equation (19). The relevant values of a, b, c and d found for this particular case are

$$a = 0.707; \qquad b = 0.577; \qquad c = 0.289; \qquad d = 0.289; \tag{29}$$

6.4 Reconstructing the Rotated Vector

To reconstruct the rotated vector, we follow the method as described earlier. For running in real remote quantum devices, calibration matrix calculations were carried out in standard fashion. Finally results of all the run were fitted with either least square measure or by convex optimization. The Reconstructed vector error from the system can be intrepreted in terms of the fidelity of actual and expected density matrix. Also the angle between analytically solved and reconstructed vector can be a good metric of correctness. Figure 9 shows the relevent results for both real and simulator backends.

7 Conclusion

In this paper, we have drawn a relationship between the Euler-Rodrigues parameters for rotations in three-dimensional space and rotations on a single qubit by thinking of the gates in terms of their effect on the Bloch sphere. Then, we used

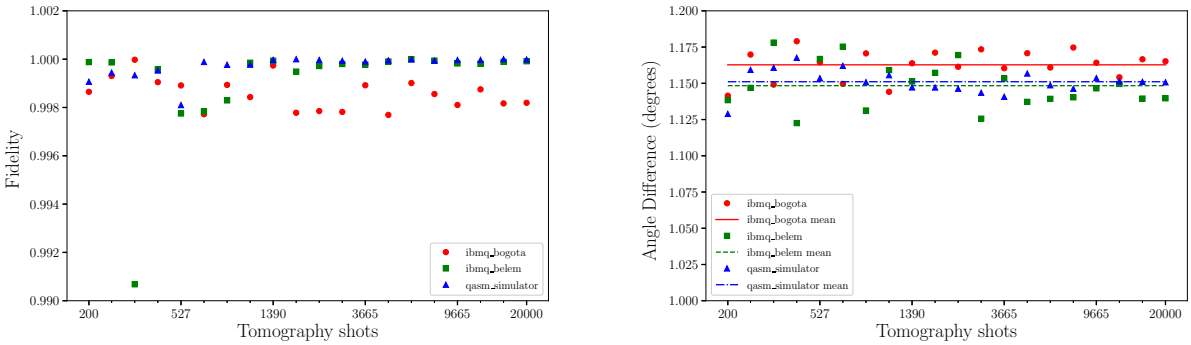


Figure 9: Fidelity and Angle difference for our specific case with $\vec{n} = \frac{1}{\sqrt{6}}(2,1,1)$, $\vec{x} = \frac{1}{\sqrt{3}}(1,1,1)$ & $\theta_r = 90°$

process tomography in the resulting circuit to extract the Euler-Rodrigues parameters from the quantum circuit up to some desired accuracy which depends on the number of shots used in the tomography. This gives us an algorithm that takes as the rotation axis and angle of rotation in radians and outputs the Euler-Rodrigues parameters as desired.

Using this characterization of the rotation parameters in a quantum circuit, we were able to rotate a vector by adding a vector initialization routine to our rotation single-qubit circuit. Then, we used state tomography (which requires less total circuits than its process counterpart) to extract the rotated vector. Using this as a basis, we proposed a multi-qubit rotation scheme to rotate 2^n vectors using $n + 1$ qubits. This scheme allows us to encode the points that make up a rigid body and rotate it a given angle about some vector. A drawback to this approach is that the points need to be of unit length.

Tests using Qiskit's [11] QASM simulator and three IBM quantum devices show that the techniques introduced in this paper yield good results when using the fidelity as a metric for the process tomography and the angle between the extracted vector and the exact rotated vector for state tomography. As expected, the larger the number of shots, the better the results. Fidelity ≥ 0.99 starts showing around just three thousand shots for the real devices and before a thousand shots for the simulator. The angle difference is always less than a degree and remains quite stable for most number of shots, with IBMQ Belem showing the lowest difference at ≈ 0.72 degrees.

References

[1] John Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2:79, Aug 2018.

[2] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum principal component analysis. *Nature Physics*, 10(9):631–633, Jul 2014.

[3] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum algorithms for supervised and unsupervised machine learning, 2013.

[4] Patrick Rebentrost, Masoud Mohseni, and Seth Lloyd. Quantum support vector machine for big data classification. *Physical Review Letters*, 113(13), Sep 2014.

[5] Wei Li, Toby Howison, and Jorge Angeles. On the use of the dual euler–rodrigues parameters in the numerical solution of the inverse-displacement problem. *Mechanism and Machine Theory*, 125, 12 2017.

[6] David Siminovitch. Rotations in nmr: Part i. euler–rodrigues parameters and quaternions. *Concepts in Magnetic Resonance*, 9:149 – 171, 01 1997.

[7] Adriano Barenco, Charles H. Bennett, Richard Cleve, David P. DiVincenzo, Norman Margolus, Peter Shor, Tycho Sleator, John A. Smolin, and Harald Weinfurter. Elementary gates for quantum computation. *Physical Review A*, 52(5):3457–3467, Nov 1995.

[8] V.V. Shende, S.S. Bullock, and I.L. Markov. Synthesis of quantum-logic circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 25(6):1000–1010, Jun 2006.

[9] Isaac L. Chuang and M. A. Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *Journal of Modern Optics*, 44(11-12):2455–2467, Nov 1997.

[10] John A. Smolin, Jay M. Gambetta, and Graeme Smith. Efficient method for computing the maximum-likelihood quantum state from measurements with additive gaussian noise. *Physical Review Letters*, 108(7), Feb 2012.

[11] MD SAJID ANIS, Héctor Abraham, AduOffei, et al. Qiskit: An open-source framework for quantum computing, 2021.

[12] Gilles Brassard, Peter Høyer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. *Quantum Computation and Information*, page 53–74, 2002.

[13] Mikko Möttönen and Juha Vartiainen. Decompositions of general quantum gates. *Frontiers in Artificial Intelligence and Applications*, 05 2005.

[14] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum random access memory. *Physical Review Letters*, 100(16), Apr 2008.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/359267071

textile dye effluent degradation by xylanase enzyme

Article · March 2022

CITATIONS
0

READS
3

2 authors, including:



Aiyakani Yogeeswaran
S T HINDU COLLEGE
3 PUBLICATIONS **50** CITATIONS

SEE PROFILE

All content following this page was uploaded by Aiyakani Yogeeswaran on 16 March 2022.

The user has requested enhancement of the downloaded file.

IJEP 42 (2) : 142-148 (2022) (ISSN : 0253-7141)

Halophilic Alkaline Protease Enzyme on Biodegradation of Contaminants in Raw Textile Effluent

A. Yogeeswaran[1], R. Raja Jeya Sekar[1] *, S. Uma Maheswari[2], S.M. Vijila[3] and T. Vijaya Kumar[4]

1. *S.T. Hindu College, P.G. and Research Department of Zoology, Nagercoil - 629 002, Tamil Nadu, India*
2. *Manonmaniam Sundaranar University, Department of Biotechnology, Tirunelveli - 627 012, Tamil Nadu, India*
3. *Pioneer Kumaraswamy College, Department of Zoology, Nagercoil - 629 003, Tamil Nadu, India*
4. *Delhi Technological University, Department of Environmental Engineering, New Delhi, India*

**Corresponding author, Email : sekar.rrj@gmail.com ; yogeemsc@gmail.com*

Textile effluent is composed of pollutants, such as dyes, dissolved solids, suspended solids and toxic materials. If this effluent is discharged without any process of treatment either into the water bodies or into the surface of the land will lead to the alteration of physico-chemical parameters of water and the soil. The present investigation revealed how the alkaline protease enzyme secreted from the halophilic bacterial strain *Bacillus cereus* RRJS2 was used to reduce the contaminants in textile wastewater. The bacterial strains were optimized at different pH, temperatures, NaCl concentrations and incubation times to produce alkaline protease enzyme. The extracted alkaline protease enzyme was used to treat the raw textile effluent. The physico-chemical parameters of both the raw textile effluent and treated effluent were compared. The results indicated that alkaline protease treated effluent recorded the reduction of physico-chemical parameters, such as pH 9.2%, temperature 4.69%, BOD 43%, COD 44%, TDS 66%, TH 44% and chlorides 46% in 5 days of degradation study. The FTIR analysis of raw and alkaline protease treated textile effluent revealed the disappearance of amine, sulphate and alkene groups of compounds and the appearance of hydroxyl, amine and aromatic groups of compounds. Thus, the alkaline protease showed a promising effect on the degradation of contaminants in the textile effluent.

KEYWORDS

Alkaline protease, *Bacillus cereus* RRJS2, Textile effluent

1. INTRODUCTION

The coloured effluent was released by two major industries; one is the dye manufacturing industries and another one is dye utilizing industries [1]. Nearly, about 10-15% of dyes used as colouring agents in the industries are finally discharged either into the barren land or water habitat [2,3]. Among the dye utilizing industries, the textile industry consumes about 300,000 tonnes of synthetic dyes for various dyeing purposes and its wastage is continuously let out into the water which causes severe water pollution [4]. The removal of contaminants in the textile effluent is a severe task and the pollutants cause a hazardous impact on the environment if it is not degraded properly [5]. The presence of dye in the textile effluent can alter the water quality, transparency, gas solubility and other physico-chemical properties of rivers, lakes and ponds [6]. To overcome such a critical condition, the biodegradation of textile effluent is an effective solution with durable, low-cost and eco-friendly methods which could not produce any secondary pollutants [7]. Therefore, in recent year's biological treatment system especially with microbial enzymatic method has become a more popular and relatively efficient system for the treatment of wastewater [8]. Both halophilic bacteria, non-halophilic bacteria and their extracellular hydrolytic enzymes were involved in the decolourization and degradation of industrial effluent [9-11]. The halophilic micro-organisms which can be able to survive in extreme conditions are proved as a suitable candidate for the production of various extracellular halotolerant enzymes, such as amylase, protease, cellulase, xylanase, etc. [12,13]. Among these, proteases are a group of enzymes that hydrolyze the peptide bonds of protein and the alkaline proteases constitute 60-65% of the global industrial enzyme market [14,15]. These bacterial enzymes are involved in the decolourization and degradation of wastewater as an alternative method to conventional physico-chemical treatment [16,17]. An enzyme, such as tyrosinase, laccases and proteases are mainly utilized in the degradation of textile dyes [18]. The enzy-

© 2022 - Kalpana Corporation

matic treatment is the most effective method to reduce the highest concentration of contaminants in textile wastewater [19,20].

The purpose of the present investigation is used to perform optimization studies with physico-chemical parameters, such as pH, temperature, NaCl concentration and incubation time on the production of alkaline protease enzyme secreted by halophilic *B. cereus* RRJS2 strain. In addition, the effect of the identified alkaline protease on the degradation of textile wastewater is also discussed.

2. MATERIAL AND METHOD

2.1 Collection, isolation, identification and screening of bacterial strain for protease activity

The sediment sample was collected from the crystallizer ponds of solar salt pans situated at Kovalam, Tamil Nadu. The collected sediment sample was serially diluted as per the standard procedure explained by Bergey [21]. From these, 1 mL of the sample (dilution factor 10^4) was inoculated into the halophilic agar medium and incubated for 48 hr at 37°C. The isolated bacterial strains were subcultured and preserved. The bacterial strain was identified by morphological and biochemical tests and 16S rRNA sequence. Further, the isolated colonies with bacterial density 23×10^3 CFU/mL were screened for protease activity by using skim milk agar medium [22].

2.2 Protease enzyme assay

The protease activity of the bacterial strain *B. cereus* RRJS2 was measured by using Anson's method [23]. Casein was used as substrate and amino acid tyrosine liberated on incubation with protease was assayed by colour development using Folin's reagent. A test sample containing 5 mL of 0.65 % (w/v) casein in 50 mM potassium dihydrogen orthophosphate buffer, pH 7.5 with 1 mL enzyme sample and the control with 1 mL of the enzyme were incubated at 37°C for 10 min. The enzyme-substrate reaction was terminated by the addition of 5 mL of 110 mM trichloroacetic acid. In control, the substrate was added after the addition of TCA. All tubes were incubated at 37°C for 30 min. The reaction mixture was centrifuged at 4500 rpm for 10 min at 4°C. 2 mL of supernatant was mixed with 5 mL of 500 mM sodium carbonate and 1 mL of Folin's reagent and the absorbance was measured at 660 nm in a spectrophotometer. One enzyme unit was defined as the amount of enzyme required to be liberated 1 μmole tyrosine per minute under assay conditions.

2.3 Optimization of bacterial strains with physico-chemical parameters

The bacterial strain *B. cereus* RRJS2 (accession no: KT347340) was optimized with various physico-chemical parameters for maximum protease production. The effect of pH on protease production was studied in pH ranging from 6 to 10 (1 pH interval) under different temperatures ranging from 30-50°C (5°C intervals) using casein as the substrate. By using the same substrate, sodium chloride (NaCl) concentration between 1-5% (1% interval) with optimum pH 8 and temperature 35°C, maximum protease production was recorded. Similarly, at 12 hr intervals, incubation time ranged from 12-60 hr, maximum protease production was observed with optimum pH 8, temperature 35°C and 1% NaCl concentration using casein as substrate.

2.4 Biomass determination

The bacterial biomass was determined with the method followed by Bakri [24]. The dry weight of bacterial biomass was determined by filtering 1 mL of cell medium through pre-weighed Whatman filter paper number 44. Then, it was dried at 80°C for an overnight in a hot air oven and reweighed. The difference in weight showed the biomass of bacteria and it was denoted as dry weight per millilitre.

2.5 Textile effluent collection

The samples were collected from the textile industry located at Erode. Samples were collected in sterilized glass bottles aseptically and transported to the laboratory in an ice pack condition. The collected samples were preserved at 4°C in the refrigerator for further analysis.

2.6 Physico-chemical characterization of raw and treated textile effluent

The physico-chemical parameters, such as pH, temperature, biological oxygen demand (BOD), chemical oxygen demand (COD), total dissolved solids (TDS), total hardness (TH) and chloride were determined in the raw textile effluent. Simultaneously, 5 mL of alkaline protease was mixed with 100 mL of raw textile effluent in a conical flask. Then, the setup was kept for 5 days in an incubator at 35°C. After the incubation period, the physico-chemical parameters of alkaline protease treated textile effluent were determined. The samples were analyzed as per the standard methods given by APHA [25].

2.7 Percentage of biodegradation

The percentage of biodegradation of textile effluent

© 2022 - Kalpana Corporation

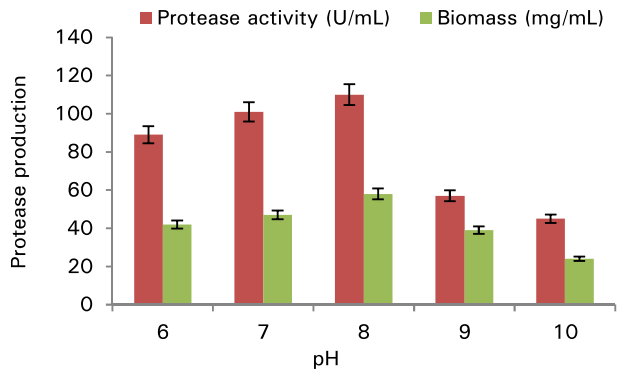


Figure 1. *Effect of pH on the alkaline protease production in B. cereus RRJS2*

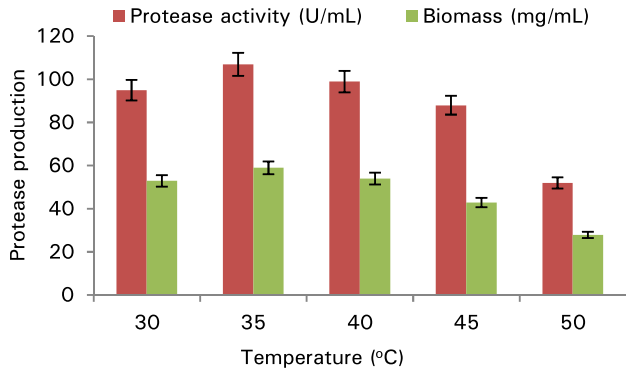


Figure 2. *Effect of temperature on the alkaline protease production in B. cereus RRJS2*

was determined by using the following formula.

$$\text{Degradation (\%)} = \frac{\text{Initial value} - \text{Final value}}{\text{Initial value}} \times 100$$

Where initial value denote physico-chemical parameters of raw textile effluent (control), final value denote physico-chemical parameters of alkaline protease treated textile effluent (experimental).

2.8 FTIR analysis

Fourier transform infrared spectroscopy (FTIR) analysis was used for determining the changes in surface functional groups of the samples, before and after microbial degradation. For FTIR analysis, raw textile effluent residue and alkaline protease treated textile effluent residue were obtained separately after evaporation of ethyl acetate extract. Then, each residue was mixed with pure KBr and the analysis was carried out individually in the mid-infrared region of 400-4000/cm with 16 scan speed by using 8400S spectrophotometer [26].

3. RESULT AND DISCUSSION

3.1 Isolation and screening of bacteria with protease activity

The morphological, biochemical properties and 16S rRNA gene sequence studies revealed that the halophilic bacterial isolate was *Bacillus cereus* strain RRJS2 (accession number: KT347340). The bacterial isolate showed a zone of protease activity around the colonies when cultured in skim milk agar plates. Jayachandran and Moreno revealed that halophilic bacteria as a source of novel extracellular hydrolytic enzymes and the isolated bacterial strain *B. cereus* RRJS2 also belongs to the same group that showed alkaline protease activity [27,28]. The isolated bacterial strain *B. cereus* RRJS2 was optimized by various physico-chemical parameters. Hence, maximum alkaline protease and biomass were obtained. This enzyme was used for the biodegradation of raw textile effluent.

3.2 Effect of pH and temperature

The extracellular protease produced from the halophilic *B. cereus* RRJS2 strain was found to be active between pH values 6 and 10. Based on the observation, the maximum extracellular protease secreted by the bacterial strain was found to be alkaline as shown in figure 1. At the alkaline pH 10, minimum enzyme activity and biomass production was observed (45 U/mL and 24 mg/mL respectively) and at pH 8, high protease activity was observed (110 U/mL and 58 mg/mL respectively). Generally, the pH of the textile effluent was found to be in alkaline condition which favours the potential use of the isolated alkaline protease enzyme for degradation. Similarly, protease production was found to be active between the temperature values 30°C and 50°C. At 35°C, maximum protease and biomass production (107 U/mL and 59 mg/mL respectively) was observed (Figure 2). Optimum activity of the enzyme was exhibited at 35°C and further increase or decrease in temperature (30°C and 50°C) caused hindrance in protease activity. Annapurna reported on production and molecular characterization of extracellular alkaline protease enzyme isolated from *Bacillus thuringiensis* and recorded the optimum conditions for highest activity at pH 8 and 37°C which were coincided with our study [29]. Olajuyigbe evaluated the protease production from *Bacillus* sp. with different ranges of pH and temperature [30]. Generally, wastewater processing plants in India operate at the temperature range of 30-37°C. Therefore, the isolated alkaline protease could be used for the effective degradation of the textile effluent.

© 2022 - Kalpana Corporation

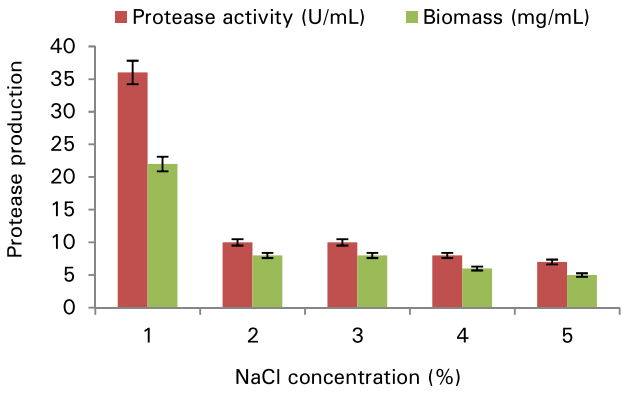


Figure 3. *Effect of different NaCl concentrations on the alkaline protease production in B. cereus RRJS2*

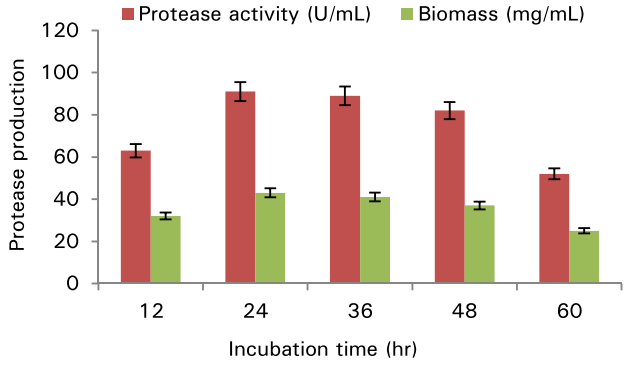


Figure 4. *Effect of different incubation times on the alkaline protease production in B. cereus RRJS2*

3.3 Effect of NaCl concentration

The influence of NaCl concentration on the alkaline protease activity and biomass production by *B. cereus* RRJS2 is shown in figure 3. The highest amount of alkaline protease and biomass production were seen in a medium that contained 1% of NaCl concentration. Increasing NaCl concentration from 2-5% caused a relative decrease in both alkaline protease and biomass production. Abdelnasser studied the effect of the optimum salinity concentration on the alkaline protease production from *Bacillus* sp. NPST-AK15 was comparable to our study [31]. Generally, chlorides ion concentration was found in the textile effluent and the effect of bacterial strain and its alkaline protease with salt tolerance showed the potential applicability in textile effluent degradation study.

3.4 Effect of incubation time

The effect of incubation time on protease activity and biomass production by *B. cereus* RRJS2 is shown in figure 4. The alkaline protease production was examined with different incubation periods, like 12, 24, 36, 48 and 60 hr. The optimum pH 8, temperature 35°C and 1% NaCl concentration induced the bacterial strains to secrete maximum protease at 24 hr of incubation (91 U/mL and 43 mg/mL respectively). Further increasing incubation time (36 – 60 hr) caused a drop in enzyme activity. Suganthi worked on screening and optimization of protease production from a halotolerant *Bacillus licheniformis* and studied optimum conditions for highest protease production at 24 hr of incubation resembled our study [32]. Generally, textile wastewater effluent treatment plants store the wastewater for 5 days treatment process. Hence the isolated protease could be used for effective degradation of textile effluent within the prescribed time.

Table 1. *Characteristics of raw and the alkaline protease treated textile effluent*

Parameters	Raw textile effluent	Alkaline protease treated effluent
pH	8.7	7.9
Temperature (°C)	32.0	30.5
BOD (mg/L)	532	304
COD (mg/L)	1047	571
TDS (mg/L)	2109	715
TH (mg/L)	578	329
Chloride (mg/L)	541	296

3.5 Biodegradation of textile effluent

The physico-chemical analysis of raw textile effluent showed higher values when compared to the values of alkaline protease treated effluent. Lokhande studied organic, inorganic pollutants and abnormal physico-chemical parameter values in raw wastewater effluent of Taloja industrial estate of Mumbai were comparable to our study in raw textile effluent [33]. The pH and temperature of the raw textile effluent were recorded as 8.7 and 32°C whereas alkaline protease treated effluent showed 7.9 and 30.5°C with degradation of 9.2% of pH and 4.69% of temperature, respectively (Tables 1 and 2). Maximum degradation was achieved at pH 7.9 and 30.5°C and further decrease or increase of pH and temperature of the effluent reduced the activity of enzyme due to denaturation. There was a highly significant interaction between pH and temperature of the effluent with a concentration of enzyme-substrate for degradation [34]. The previous reports on dye decolourization and degradation by using fungal laccase enzyme revealed that the optimum pH and the added volume of enzyme resulted in the highest degree of degradation [35]. The enzyme

© 2022 - Kalpana Corporation

Table 2. *Percentage of biodegradation in the alkaline protease treated textile effluent*

Parameters	Biodegradation (%)
pH	9.2
Temperature	4.69
BOD	42.85
COD	45.46
TDS	66.09
TH	43.07
Chloride	45.29

azoreductase is secreted by the bacterium *Enterobacter* sp. SXCR under the optimum condition degraded the textile azo dye Congo Red in a wide range of pH 5-9 and temperature 22-40°C was comparable to our study [36].

Biological oxygen demand (BOD), chemical oxygen demand (COD), total dissolved solids (TDS), total hardness (TH) and chlorides are considered important parameters to analyze water quality. The BOD of raw textile effluent was 532 mg/mL and alkaline protease treated effluent was 304 mg/mL with a reduction of 42.85% was observed in the enzyme-treated effluent. Similarly, COD of textile effluent was reduced from 1047 mg/mL to 571 mg/mL with degradation of 45.46%; TDS was reduced from 2109 mg/mL to 715 mg/mL with degradation of 66.09%; TH was reduced from 578 mg/mL to 329 mg/mL with degradation of 43.07% and chlorides was reduced from 541 mg/mL to 296 mg/mL with degradation of 45.29% (Tables 1 and 2). The previous works on textile dye effluent treatments were mainly based on using microbes for decolourization and recently few reports suggested that the extracellular microbial enzymes were used as a potential degrading agent for reduction of toxicity with the formation of various intermediate products during the process of biodegradation [37]. Soune worked on the identification of the extracellular amylase, protease and lipase from *B. licheniformis* and their effect on degradation of BOD and COD in the wastewater was agreed with our findings [38]. The parameters, such as BOD, COD and TDS in the textile dye effluent were reduced by the treatment of fungal strain *Irpex lacteus* and bacterial strain *Bacillus* sp. [39,40]. The extracellular microbial enzymes, like oxidoreductase, laccase mediator system and laccase combined ionic based surfactants also reduced the pollutants of textile dye effluent [41-43].

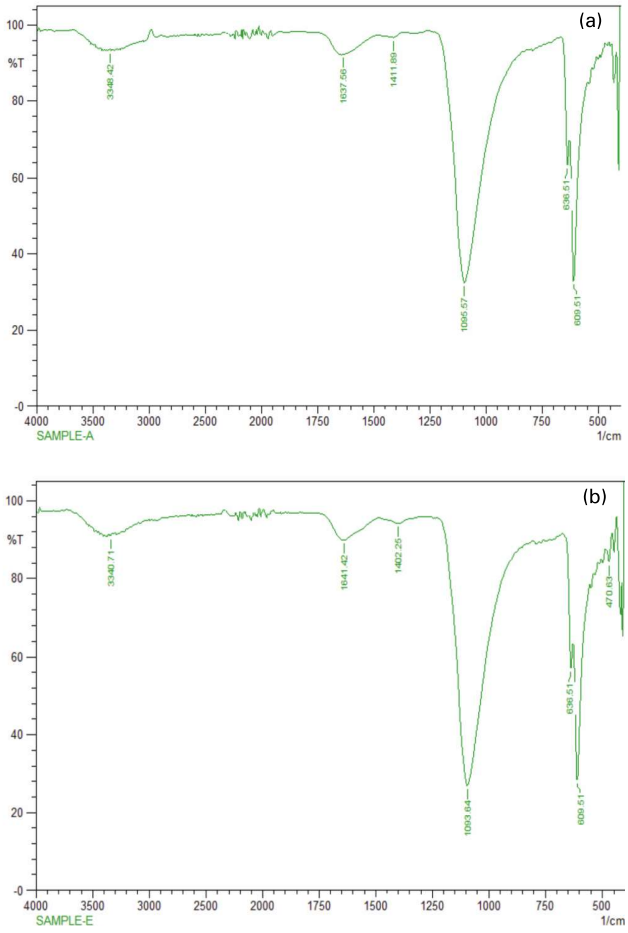


Figure 5. *FTIR spectrum of (a) raw textile effluent and (b) alkaline protease enzyme-treated textile effluent*

3.6 Fourier transform infrared analysis

The degradation products of raw textile effluent were determined by FTIR analysis. FTIR spectrum of raw textile effluent showed a peak at 3348.42/cm for N-H stretching in secondary amine compound, a peak at 1637.56/cm for C=C stretching in alkenes compound, a peak at 1411.89/cm for S=O stretching in sulphate compound, a peak at 1095.57/cm for C-O stretching in secondary alcohol compound and peaks at 636.51/cm and 609.51/cm, respectively for C-Br stretching in halogen compounds (Figure 5a). After degradation of textile effluent by an alkaline protease enzyme, the FTIR spectrum showed the appearance of a peak at 3340.71/cm for O-H stretching in alcohol compound, a peak at 1641.42/cm for C=N stretching in imine or oxime compound, a peak at 470.63/cm for aromatic ring stretching in the aromatic compound, a peak at 1402.25/cm for S=O stretching in sulphate compound and a peak at 1093.64/cm for C-O stretching in secondary alcohol. Similar peaks for C-Br stretching as

© 2022 - Kalpana Corporation

found in the FTIR spectrum of raw textile effluent are also displayed in the alkaline protease treated effluent (Figure 5b). The disappearance of peaks at 3348.42/cm and 1637.56/cm for N-H stretching and C=C stretching, respectively and the appearance of new peaks at 3340.71/cm, 1641.22/cm and 470.63/cm for OH stretching, C=N stretching and an aromatic ring stretching, respectively indicated that the alkaline protease enzyme secreted by the bacterial strain *B. cereus* RRJS2 has degraded various compounds in the raw textile effluent into subsequent compounds.

4. CONCLUSION

A salt-tolerant alkaline protease produced by *B. cereus* RRJS2 was isolated from a hyper saline environment from a solar saltpan. The effect of physico-chemical factors, such as pH, temperature, NaCl and incubation time was studied for a maximum alkaline protease production at pH 8, 35°C temperature, 1% NaCl concentration and 24 hr of incubation. The enzymatic degradation study was performed by the treatment of raw textile effluent with alkaline protease enzyme under optimized conditions and the results showed the reduction of contaminants. The FTIR analysis of raw textile effluent and alkaline protease enzyme treated effluent further revealed the degradation of contaminants in the textile effluent. The results obtained in the study illustrated the viability of using halotolerant alkaline protease for the degradation of textile effluent.

ACKNOWLEDGEMENT

The authors are grateful to UGC, New Delhi for the JRF fellowship under a major research project.

REFERENCES

1. Pearce, C.J., J.R. Llyod and J.T. Guthrie. 2003. Review : The removal of colour from textile wastewater using whole bacterial cells. *Dyes Pigments.* 58:179-196.
2. Schleiphake, K., *et al.* 2000. Transformation and degradation of the disazo dye Chicago Sky Blue by a purified laccase from *Pycnoporous cinnabarinus. Enzyme Microbial. Tech.,* 27:100-107.
3. Verma, P. and D. Madamwar. 2002. Decolourization of synthetic textile dyes by lignin peroxidase of *Phanecrochaete chrysosporium. Folia Microbiol.,* 47:283-286.
4. Keharia, H. and D. Madamwar. 2003. Bioremediation concepts for treatment of dye containing wastewater : A review. *Indian J. Expt. Biol.,* 41: 1068-1075.
5. Nigam, P., *et al.* 1996. Microbial process for the decolourization of textile effluent containing azo, diazo and reactive dyes. *Proc. Biochem.,* 31:435-442.
6. Banat, I.M., *et al.* 1996. Microbial decolourization of textile dye containing effluents : A review. *Bioresour. Tech.,* 58:217-227.
7. McMullan, G., *et al.* 2001. Microbial decolourization and degradation of textile dyes. *Appl. Microbiol. Biotech.,* 56:81-87.
8. Garcha, S., N. Verma and S.K. Brar. 2016. Isolation, characterization and identification of microorganisms from unorganized diary sector wastewater and sludge samples and evaluation of their biodegradability. *Water Resour. Ind.,* 16:19-28.
9. Asad, S., *et al.* 2007. Decolourization of textile azo dyes by newly isolated halophilic and halotolerant bacteria. *Bioresour. Tech.,* 98:2082-2088.
10. Asgher, M., *et al.* 2008. Recent development in biodegradation of industrial pollutants by white rot fungi and their enzyme system. *Biodegrad.,* 19:771-783.
11. Kalme, S., *et al.* 2009. Textile dye degradation laccase from *Pseudomonas desmolyticurn* NCIM 2112. *Enzyme Microbial Tech.,* 44:65-71.
12. Joo, H.S., *et al.* 2003. Oxidant and SDS-stable alkaline protease from *Bacillus clausii* 1-52 : Production and some properties. *J. Appl. Microbiol.,* 5:267-272.
13. Shamasehar, Abdulhameed. 2011. Extracellular alkaline protease by a newly isolated halophilic *Bacillus* sp. *Global J. Biotech. Biochem.,* 6:142-148.
14. Johnvely, B., B.R. Manjunath and G.R. Naik. 2002. Pigeon pea waste as a novel, inexpensive, substrate for production of a thermostable alkaline protease from thermo alkalophilic *Bacillus sp.* JP-99. *Bioresour. Tech.,* 82:61-64.
15. Naseimento, W.C.A. and M.L.L. Martins. 2004. Production and properties of an extracellular protease from thermophilic Bacillus sp. *Brazilian J. Microbiol.,* 3:591-596.
16. Campos, R., *et al.* 2001. Indigo degradation with purified laccases from *Trametes hirsuta* and *sclerotium rolfsii. J. Biotech.,* 89:131-139.
17. Agarry, S.E. and O. Ayobami. 2011. Evaluation of microbial systems for biotreatment of texile waste effluents in Nigeria : Biodecolourization and biodegradation of textile dye. *J. Appl. Sci. Env. Manage.,* 15:79-86.
18. Suzuki, T., *et al.* 2001. Correlation of aerobic biodegradability of sulphonated azo dyes with the chemical structure. *Chemosphere.* 45:1-9.
19. Robinson, T., *et al.* 2001. Remediation of dyes in textile effluent : A critical review on current treat-

© 2022 - Kalpana Corporation

ment technologies with a proposed alternative. *Bioresour. Tech.,* 77:247-255.
20. Maier, J., *et al.* 2004. A new alkali-thermostable azo reductase from *Bacillus sp.* strain SF. *Appl. Env. Microbiol.,* 70:837-844.
21. Bergey, D.H. and J.G. Holt. 2000. Bergey's mannual of determinative bacteriology (9th edn). Lippincott Williams and Wilkins, Philadelphia.
22. Salwan, R., A. Gulati and R.C. Kasana. 2010. Phylogenetic diversity of alkaline protease producing psychrotrophic bacteria from glacier and cold environments of Lahual and Spiti, India. *J. Basic Microbiol.,* 50:150-159.
23. Anson, M.L. 1938. The estimation of pepsin, trypsin, papain and cathepsin with hemoglobin. *J. Gen. Physiol.,* 22:79-89.
24. Bakri, Y., M. Magali and P. Thonart. 2009. Isolation and identification of a new fungal strain for amylase biosynthesis. *Polish J. Microbiol.,* 58:269-273.
25. APHA. 2005. Standard methods for the examination of water and wastewater (21st edn). American Public Health Association, Washington D.C.
26. Goodacre, R., *et al.* 2000. Detection of the dipicolinic acid biomarker in Bacillus spores using Curie-point pyrolysis mass spectrometry and Fourier transform infrared spectroscopy. *Anal. Chem.,* 72:119-127.
27. Jeyachandra, S., *et al.* 2012. Characterization of extracellular hydrolytic enzyme producing extremely halophilic bacterium *Virgibacillus sp.* JS5. *World J. Sci. Tech.,* 2:23-26.
28. Moreno, M.L. *et al.* 2013. Halophilic bacteria as a source of novel hydrolytic enzymes. *Life.* 3:38-51.
29. Annapura, S., *et al.* 2013. Isolation and molecular characterization of alkaline protease producing *Bacillus thuringiensis. Cell. Biochem. Biophy.,* 66:45-51.
30. Olajuyigbe, F.M. and J.O. Ajele. 2005. Production dynamics of extracellular protease from Bacillus species. *African J. Biotech.,* 4:776-779.
31. Abdelnasser, S.S., *et al.* 2015. Production of extracellular protease by new halotolerant alkaliphilic *Bacillus sp.* NPST-AK15 isolated from hypersaline soda lakes. *Electronic J. Biotech.,* 18:236-243.
32. Suganthi, C., *et al.* 2013. Screening and optimization of protease production from a halotolerant *Bacillus licheniformis* isolated from saltern sediments. *J. Gen. Eng. Biotech.,* 11:47-52.
33. Lokhande, R.S., P.U. Singare and D.S. Pimple. 2011. Toxicity study of heavy metal pollutants in wastewater effluent collected from Taloja, industrial estate of Mumbai, India. *Resour. Env.,* 1:13-19.
34. Benzina, O., *et al.* 2012. Enhanced decolourization of the azo dye Sirius Rose BB by laccase-HBT system. *3 Biotech.,* 2:149-157.
35. Devi, V.M., *et al.* 2012. Dye decolourization using fungal laccase. *Bull. Env. Pharmacol. Life Sci.,* 1:67-71.
36. Prasad, S.S. and K. Aikat. 2014. Study of biodegradation and biodecolourization of azo dye by *Enterobacter sp.* SXCR. *Env. Tech.,* 35:956-965.
37. Sarkar, S., *et al.* 2017. Degradation of synthetic azo dyes of textile industry : A sustainable approach using microbial enzymes. *Water Conser. Sci. Eng.,* 2:121-131.
38. Soune, N. and A. Garode. 2018. Isolation characterization and identification of extracellular enzyme producer *Bacillus licheniformis* from municipal wastewater and evaluation of their biodegradability. *Biotech. Res. Innovation.* 2:37-41.
39. Shin, K.S. 2004. The role of enzymes produced by white rot fungus *Irpex lacteus* in the decolourization of the textile industry effluent. *Microbiol.,* 42:37-41.
40. Pourbabaee, A., *et al.* 2006. Aerobic decolourization and detoxification of a disperse dye in textile effluent by a new isolate of Bacillus sp. *Biotech. Bioeng.,* 93:631-635.
41. Husain, Q. 2006. Potential applications of the oxidoreactive enzymes in the decolourization and detoxification of textile and other synthetic dyes from polluted water : A review. *Critical Reviews Biotech.,* 26:201-221.
42. Khilfi, R., *et al.* 2010. Decolourization and detoxification of textile industry wastewater by the laccase-mediator system. *J. Hazard. Mater.,* 175:802-808.
43. Bento, R.M.F., M.R. Almeida and P. Bharmoria. 2020. Improvements in the enzymetic degradation of textile dyes using ionic liquid-based surfactants. *Sep. Purif. Tech.,* 235:1-7.

© 2022 - Kalpana Corporation

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 322-332
Published Online November 2021 in IJEAST (http://www.ijeast.com)



IMPACT OF VARIATION IN SIZES OF BORON CARBIDE ON PROPERTIES OF NOVEL COMPOSITE OF ALUMINIUM ALLOY 6063-T6 AND BORON CARBIDE

Vibhu Singh
PhD Scholar, Department of Mechanical Engineering, Delhi Technological University, Delhi, India.

Soni Kesarwani
PhD Scholar, Department of Mechanical Engineering, Delhi Technological University, Delhi, India.

M.S. Niranjan
Associate Professor, Department of Mechanical Engineering, Delhi Technological University, Delhi, India

***Abstract*-Composites are continuously evolving as the most demanding materials among various industries because of their reinforced mechanical properties and lesser weight. The present work aimed to fabricate a novel aluminium matrix composite using Aluminium Alloy 6063 T6 as base metal and ceramic boron carbide particles with three different micro sizes viz.104µm, 74µm, and 53µm as reinforcement. For composite fabrication, stir casting technique has been utilized. Micro-structure and mechanical behavior of base alloy and prepared composites has been analyzed in this work. Observing through optical microscopy reveal a uniform and proper distribution of B_4C particle in Al matrix, along with cluster formation at few sites. The results also exhibit that hardness of reinforced composite has been found to be more than that of base alloy by 81.35 %, in which composite with particle size 53µm has shown maximum hardness of 75.77HV. The trend in tensile strength indicates that the ultimate tensile strength of composite with boron carbide of mesh size 53µm is more than the base alloy by 20.79% and is also highest among the rest of composites with a value of 219.196MPa.**

***Keywords*-Aluminium alloy 6063 T6, Boron Carbide (B_4C), Aluminium Metal Composite (AMC), Micro-structure, Micro-Hardness, Ultimate Tensile Strength**

I. INTRODUCTION

In current scenario, composites are emerging as materials replacing the monolithic materials and occupying the place of most demanding materials in various industrial sectors listing from automation & aerospace to electronics and healthcare [1]. Encompassing the properties like high specific strength, good stiffness, low thermal expansion, excellent corrosion and wear resistance along with lighter weight helps it to satisfy the global need at various level [2]. Among various composites, metal matrix composites are ones attracting the eyeballs of researchers most. Basically, metal matrix composites are made from two components- one is matrix which is continuous in nature and perform functions such as it helps to hold the reinforcement fibers together and protects it from environment and abrasion. The other component is reinforcement which is discontinuous in nature, stronger, harder and helps in bearing the load that is applied over the composite and transfers the required strength to the matrix [3]. Metal matrix composites can be basically categorized based on the type of reinforcement induced like particulate, laminar or fibre & on type of matrix like metal, ceramic etc. Centering on particulate metal matrix composites, here the reinforcement is dispersed over the entire volume of base matrix in the form of particles. Being isotropic in nature, the strength-to-weight ratios of the metal matrix composite is generally higher than most of the base alloys.

This investigation aims to prepare composites having aluminium alloy of grade 6063 T6 as base matrix since aluminium is one of the lightest metals with properties like good corrosion resistance, high reflectivity and strength that can some time smatch up with construction steel. Boron carbide (B_4C) is used as reinforcement which is extremely

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 322-332
Published Online November 2021 in IJEAST (http://www.ijeast.com)



hard ceramic and covalent in nature, find its utilization in tank defensive layer, impenetrable vests, motor damage powders and in various other mechanical applications[2]. It is accepted as one of the hardest materials in present times, behind cubic boron nitride and precious stone diamond. Though various techniques are available to manufacture the composites, stir casting is chosen as its highly economical method requiring only one-third of cost to prepare composites as compared to other methods available[4].In recent times, double stircasting or two- step mixing process is gaining more popularity. This process involves melting base metal above its liquidus temperature and then dropping the temperature to a value lying in between liquidus and solidus so that when cools, it forms a semi-solidstate. Assoonasmelt reaches at semi solid state, preheated reinforcement particles are added to it. After that, composite slurry is again heated to higher temperature to gain the liquid state. This process is able to break the layer of gases which form around the surface of the particles hindering the wettability of the reinforcement particle with molten metal matrix. Majorly, there are three factors which influence the output of the stir casting process-speed of mixing/stirring, time length of mixing/stirring and mixing temperature [3]. Stirring can enhance the uniformity and homogeneity in the microstructure of composite achieved.

The primary contributions of this research are as follows-(1) Manufacturing of composite has been done through convention alstircasting technique(2)Novelty of the research lies in the combination of base matrix and reinforcement chosen i.e. aluminium alloy of series 6063 T6 and reinforcement boron carbide ,as till now properties of a composite with this particular combination hasn't been explored or touched upon.3)For the first time, these three mesh sizes(MS) of boron carbide particle -150,200 and 270 are imparted with same weight of 6% (4) Micro structural observations, tensile strength and micro hardness measurements have been carried out on the surface of the composite samples and base alloy.

II. EXPERIMENTAL METHODS ANDMATERIALS

Materials

Aluminium alloy 6063 T6 is put to use as a base metal matrix which comprises of magnesium and silicon as main alloying elements. Here T6 stands for solution that is heat-treated and artificially aged. This alloy of aluminium portrays properties like good surface completion, high corrosion resistance, and can go easy with welding operation. Reinforcement of aluminium has been done by imparting boron carbide powder, ceramic by nature and one of the hardest elements known. In this experiment, three novel and different sizes of B_4C has been incorporated-viz.104μm, 74μm, 53μm with a weight of6%.

Table 1 Chemical distribution of Al alloy 6063

Elements	**Percentage**
%Mg	0.45-.090
%Si	0.20-0.60
%Fe	0.0-0.35
%Cu	0.0-0.10
%Mn	0.0-0.10
%Cr	0.0-0.10
%Zn	0.0-0.10
%Ti	0.0-0.10
%Others	0.0-0.15
%Al	Balance

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 322-332
Published Online November 2021 in IJEAST (http://www.ijeast.com)

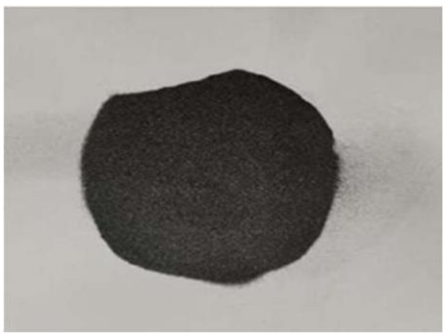


Figure 1 Boron Carbide Powder

Table 2 Particle size of B4C powder used in the experiment

Mesh size	**Microns(µm)**
150	104
200	74
270	53

Though many techniques are available in the market to prepare the composite like stir casting, infiltration, squeeze casting but stir casting is one that dominates among all as it is economical, facilitates the fabrication of large size product and can handle huge volumes of production [4].

Fabrication process

Aluminium6063–T6alloyisusedasthebasematrixand boron carbide (B_4C) of mesh size 150 (104µm), 200 (74µm) and 270 (53µm) at 6 % wt. are used as reinforcement in the manufacturing.

For the fabrication purpose, stir casting method is employed which involves various steps from melting of metal alloy to the preparation of final composites. Elaborating these steps as follows [5,6]-

Step1-A cylindrical graphite crucible of required dimension and having a maximum working temperature of 2700℃ is kept inside a 220V, three phase electrical resistance furnace (maximum operating temperature 1500℃) for base alloy melting and to prepare the slurry. It is equipped with thermocouple (K type) having a maximum range of 1200℃ and speed regulated motorized stirring system.

Step 2- Using muffle furnace, the boron carbide powder is heated to 350℃ for 45 min before being mixed with the molten alloy. In order to oxidize particles surface, to eliminate moisture and improving the wettability in matrix content, preheating of ceramic particle is done.

Step 3-In order to achieve proper mixing of ceramic particles, vortex is being developed on the surface of the molten alloy using a stirrer running at a speed of 400 rpm. After this, boron carbide particulates are added slowly into the vortex of molten alloy using funnel and rod. Speed of the stirrer is increased slowly and continuously for up to 10 minutes. The stirrer bladesare positioned from the surface atadepthofone third of the molten slurry, to enable propermixing.

Step4-Aftermixingthereinforcementforabout15min with the help of stirrer, stirrer is taken out and composite slurry is allowed to cool down and solidify. After this, casted composite is removed out of the graphite crucible.

Step 5-Proper turning and finishing is done over the composite's outer surface through conventional lathe machine using carbide tip cutting tool. After finishing, required specimen are being cut along the transverse section using wire electro discharge machine which is a thermal non-contact machining technique.

Step 6-Two specimens are being cut from each prepared AMC's, one for tensile test and other common for microstructure and micro hardness.

Step 7-Hardness test is performed on Vickers micro hardness machine with model name FISCHERSCOPE HM2000 S having test load range-0.1-2000 mN, F=300,000 mN/20sec and Creep-5.0 sec.

Step 8-To analyze the microstructure, specimens are cut and prepared as the per standard metallographic procedure performed in sequential order-Sectioning, Mounting, Sample surface. polishing, Grinding, Course polishing, Etching.

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 322-332
Published Online November 2021 in IJEAST (http://www.ijeast.com)

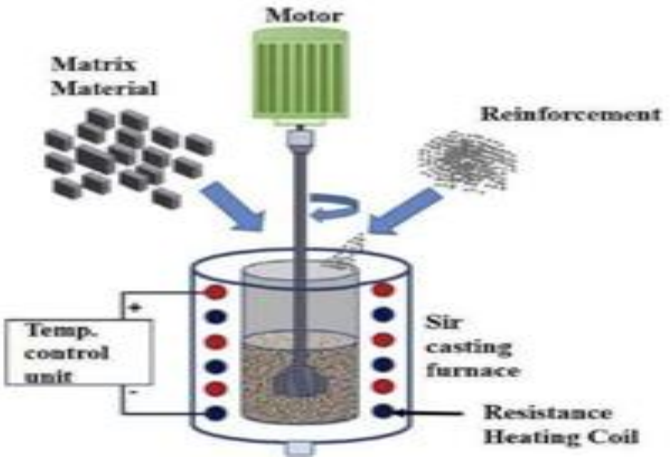


Figure 2 Stir Casting Techniques [2]

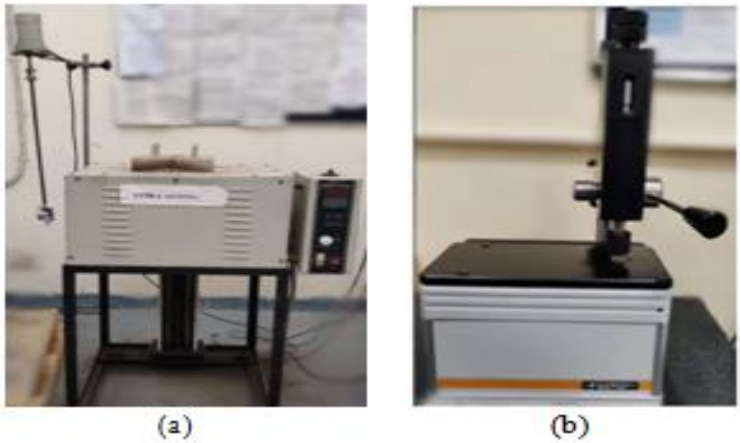


Figure 3 (a) Conventional Stir Casting setup (b) Vickers's micro hardness equipment

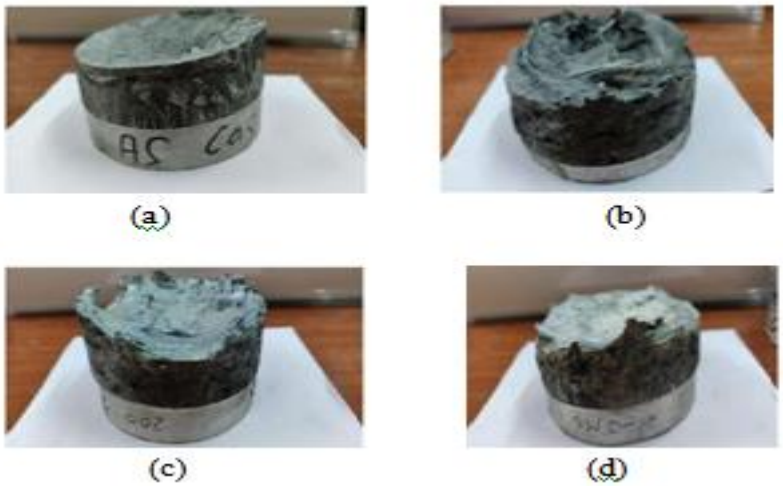


Figure 4 Casting of (a) Base alloy (b) AMC MS-150

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 322-332
Published Online November 2021 in IJEAST (http://www.ijeast.com)

(c) AMC MS-200 (d) AMC MS-270

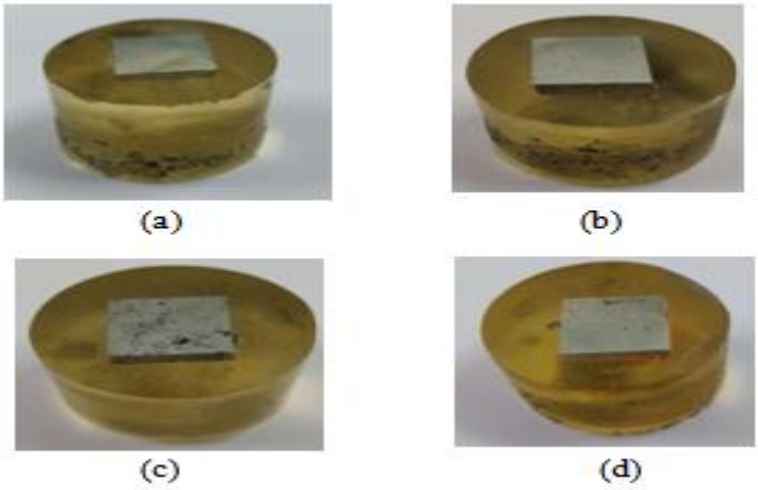


Figure 5 Specimens for microstructure and micro hardness analysis (a)Base alloy (b) AMC MS-150 (c) AMC MS-200 (d) AMC MS-270

III. RESULT AND DISCUSSIONS

Micro-Hardness

Hardness is the measure of the resistance of a material to indentation or damage on the surface [7]. In this experiment,hardnessofalloyandthreecompositeshas been calculated using Vickers's Micro Hardness test at four different locations on the prepared transverse section of composite and calculating its mean value in terms ofHV.

Table 3 Hardness value for various specimens

Specimens	**Hardness (HV)**
As casted Al-6063	41.78
AMC (MS-150)	49.94
AMC (MS-200)	60.02
AMC (MS-270)	75.77

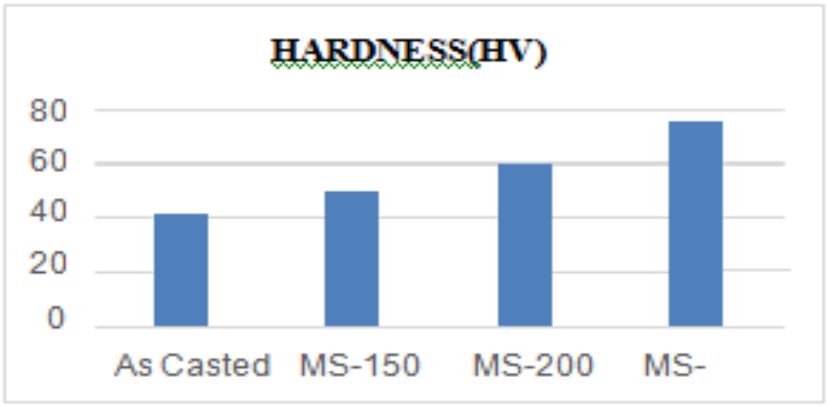


Figure 6 Bar graph representing Hardness value

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 322-332
Published Online November 2021 in IJEAST (http://www.ijeast.com)

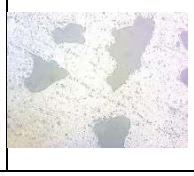
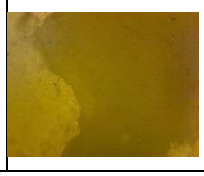
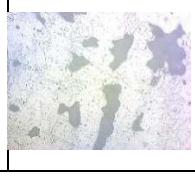
From above graph, it is revealed that hardness of reinforced composite specimen is more than that of basemetalAl-6063[8],[7,9].Thiscanbeexplainedon the basis of two facts that, firstly B_4C is entitled as one of the hardest known materials and secondly the presence of reinforcement particles come with a large amount of dislocation density during the solidification. Hence, presence of its particle on the surface of the prepared composite increases resistance towards indentation or plastic deformation, resulting in increase of composite hardness[10].

Among the AMC's, the hardness is found maximum for the composite with B_4Cmeshsize270whichis81.35% more than base alloy and gradually goes on decreasing for composite with B_4C mesh size 200 and 150 accordingly. As the size of there in forcement decreases, it results in more uniform mixing with themetal matrix and shorter inter-particle distances within the matrixas compared to larger size reinforcement. As a result, composite start behaving as more monolithic with less spatial dependent material variations. Hence results in increase in hardness.[11]

Microstructure

To investigate the microstructure, proper sectioning and mounting has been done one achaluminium matrix composite carrying different size of B_4C. Samples are observed through transverse section under the optical microscope at magnification level i.e., 10X,20X,50X& 100Xintwostages-firstbeforeetchingandsecondafter etching.

Table 4 Micrographs of various specimens

Specimen	**Magnifications**				
	Etching	**10X**	**20X**	**50X**	**100X**
As casted Al alloy 6063	**Before**				
	After				
AMC(MS-150)	**Before**				
	After				
AMC(MS-200)	**Before**				

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 322-332
Published Online November 2021 in IJEAST (http://www.ijeast.com)

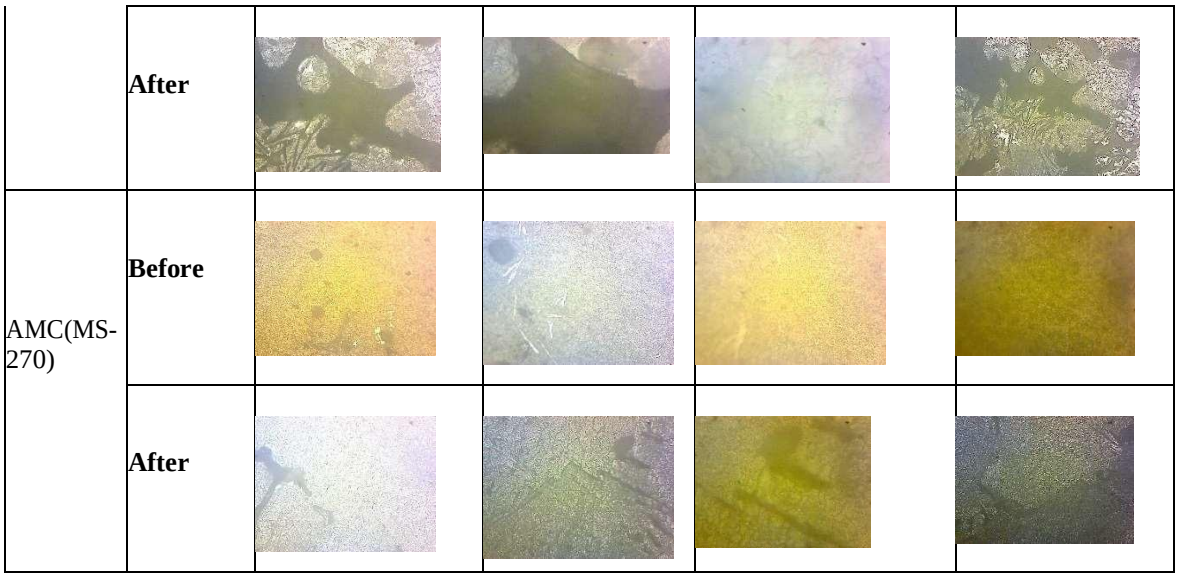




Figure 7 Optical micrograph of a composite

From micrographs, it is observed that the boroncarbide powder with different sizes has fairly uniform distribution in maximum portion within the core of molten matrix of Al-6063 and only some areas encounter a cluster formation [11]. This is because of proper stirring action and the usage of appropriate process parameters like preheating temperature, stirring speed, stirring temperature, stirring time, [12]. Clustering may have occurred due to inappropriate wetting behavior of boron carbide particles in molten aluminium alloy [12]. Some porosity also occurred in the microstructure of the composite due to entrapment of air bubbles while mixing of reinforcement in the molten alloy.[12].

Tensile Strength

On performing tensiletest on five samples of base alloy and composites each on TINIUS OLSEN H50K-S UTM, following mean values of tensile strength are obtained along with respective standard deviation.

Table 5 Tensile strength & Standard deviation values for various specimens from five tests.

Specimens	**Tensile strength Mean value (MPa)**	**Standard Deviation**
As casted Al-6063	181.486	0.439064

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 322-332
Published Online November 2021 in IJEAST (http://www.ijeast.com)

AMC (MS 150)	194.770	2.003
AMC (MS-200)	211.498	1.249696
AMC (MS-270)	219.196	1.152464

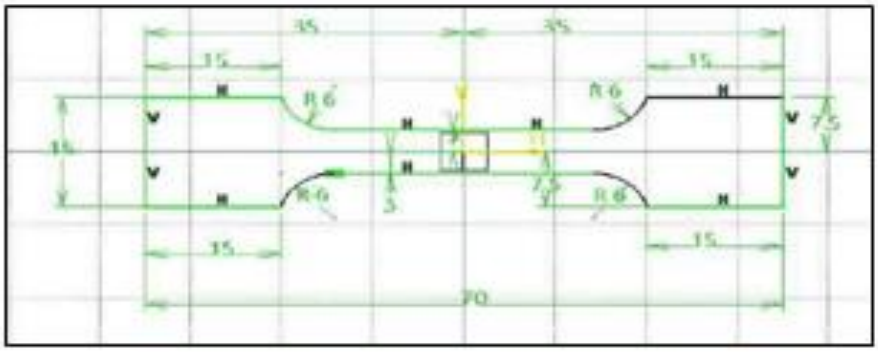


Figure 8 Schematic of standard tensile test specimen

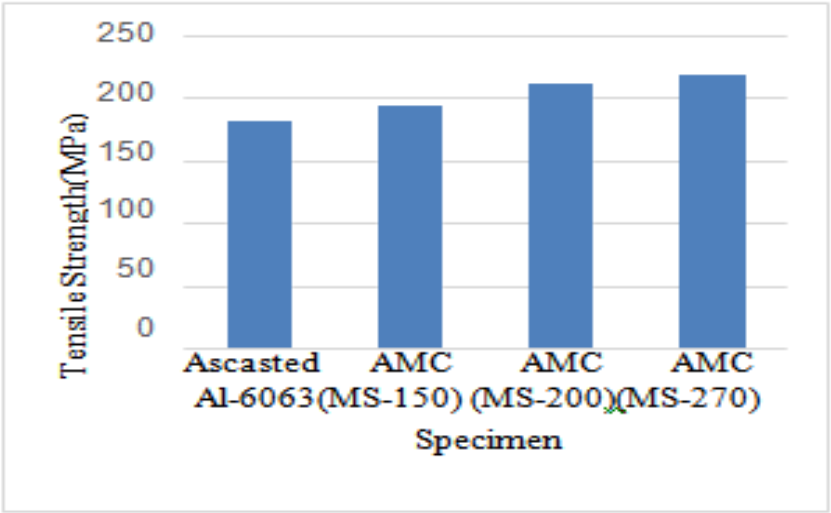
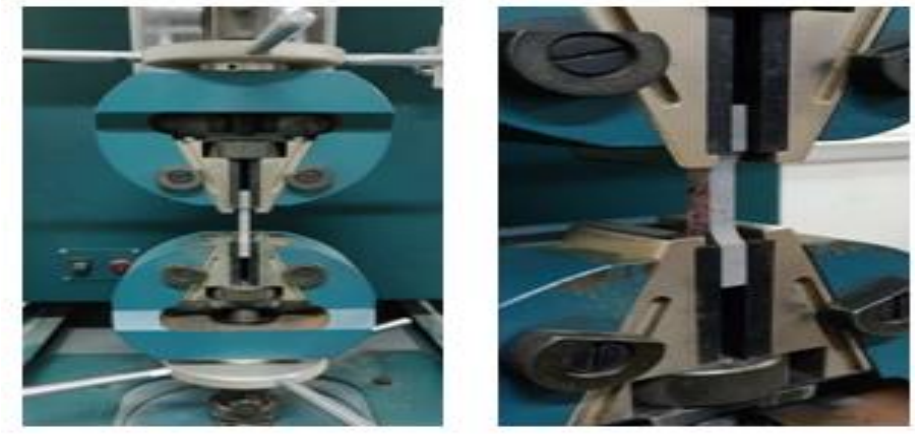


Figure 9 Bar graph between Tensile strength & various specimens

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 322-332
Published Online November 2021 in IJEAST (http://www.ijeast.com)



(a) (b)

Figure 10 Specimen subjected to tensile loading at UTM

From above, it is found that tensile strength of base alloy aluminium 6063 is less than the fabricated composites by 20.79 %. [9], [13] Among the composites, AMC's with B_4C mesh size 270 has highest tensile strength as compared to composite with B_4C of mesh size of 200 and 150.

This enhancement in the tensilestrength of the fabricated composite is a result of two strengthening effects which are independent but work simultaneously to produce a combine effect. One is direct effect in which the load is shifted from soft aluminium matrix to the hard boron carbide particles via matrix-particle interface. The other effect is indirect in which the presence of reinforcement increases the dislocation density in the metal matrix[1].



Figure 11 Base alloy specimens-Before and after tensile test



Figure 12 AMC MS-150 specimens-Before andafter tensiletest

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 322-332
Published Online November 2021 in IJEAST (http://www.ijeast.com)



Figure 13 AMC MS-200 specimens-Before andafter tensiletest



Figure 14 AMC MS-270 specimen-Before andafter tensile test

IV. CONCLUSIONS

The following significant conclusions are derived from the above investigation:

(1) Aluminum Metal Matrix composites has been preparedsuccessfullyusingaluminium6063T6asbase matrix and boron carbide powder with three different mesh sizes (viz 150,200,270) by stir castingtechnique.

(2) Optical microscopy study reveals a uniform distribution of B_4C powder in maximum region of aluminum core along with agglomerations at few sites. By developing composites through stir casting, it isobserved that clustering of reinforcement particles and their wettability with metal matrix are major concern that has be taken care for future studies.

(3) The hardness of the reinforced composites is found more than that of un-reinforced aluminium alloy. Maximum hardness is found for the composite with B_4C mesh size 270 (53μm) and shows a decreasing trend as size increases from 200 (74μm) to 150 (104μm).

(4) Ultimate tensile strength of base alloy aluminium 6063 has been found less than the prepared AMC's. AmongAMC's,thecompositewithB_4Cmeshsize270 has maximum tensile strength as compared to one with mesh size of 200 and 150respectively.

REFERENCES

[1]. Tianbing He, OnurErtuğrul,NevafCiftci,VolkerUhlen winkel,KorneliusNielsch & Sergio Scudino , "Effect of particle size ratio on microstructure and mechanical properties of aluminum matrix composites reinforced with $Zr_{48}Cu_{36}Ag_8Al_8$ metallic glass particles, Material Science and Engineering: A ,2019,742(10),517-525.

[2]. P.B.Pawar, R.M.Wabale, A.A.Utpat,"A ComprehensiveStudyofAluminumBasedMetalMatrix Composites: Challenges and Opportunities", Materials Today:Proceedings,5,(2018),23937–23944.

[3]. Pulkit Garg, Anbesh Jamwal, Devendra Kumar, Kishor Kumar,Sadasivuni Chaudhery Mustansar Hussain & Pallav Gupta, "Advance research progresses in aluminium matrix composites: manufacturing & applications", Journal of material research and tehnology,(2019),4924-4939.

[4]. Mohammed Imran, A.R.Anwar Khan, "Characterization of Al 7075 metal matrix composites",Journal of material research and technology,(2019),3347–3356.

[5]. Narender Panwar, Amit Chauhan, "Fabrication methods of particulate reinforced Aluminium metal matrix composite-A review", Materials Today: Proceedings, 5(2),(2018),5933-5939.

[6]. K.T.Akhil,Jerry Varghese,Arun Raphel,K.Vinoj Freosh & K.Francis, "To Study the Cooling Rate and



Influence of Boron Carbide on Mechanical Properties of Aluminium LM13 Matrix B4C Reinforced Composites", Materials Today: Proceedings, 4(8), (2017), 7202-7207.

[7]. Ulhas K.Annigeri,G.B.Veeresh Kumar," Effect of Reinforcement on Density, Hardness and Wear Behavior of Aluminum Metal Matrix Composites: A Review", Materials Today: Proceedings, 5(5), 2018, 11233-11237.

[8]. Dinesh Patidar, R.S.Rana, "Effect of B4C particle reinforcement on the various properties of aluminium matrix composites: a survey paper",Materials Today: Proceedings, 4(2),2017, 2981-2988.

[9]. Keshav singh, R.S.Rana& Anjaney Pandey,"Fabrication and Mechanical properties characterization of aluminium alloy LM24/B4C composites", Materials Today: Proceedings, 4(2), (2017), 701-708.

[10]. K.Shirvanimoghaddam,H.Khayyam,H.Abdizadeh, M.Karbalaei Akbari,A.H.Pakseresht & E.Ghasali,"Boron carbide reinforced aluminium matrix composite: Physical, mechanical characterization and mathematical modelling", Material Scienceand Engineering: A, (2016), 658,135-149.

[11]. Andy Nieto,Hanry Yang,Lin Jiang & Julie M.Schoenung," Reinforcement size effects on the abrasive wear of boron carbide reinforced aluminum composites", Wear,(2017), 390–391,228-235.

[12]. Bhaskar Chandra Kandpal,Jatinder Kumar & Hari Singh, "Manufacturing and technological challenges in Stir casting of metal matrix composites– A Review", Materials Today: Proceedings, 5(1),(2018), 5-10.

FOREX Publication
Open Access | Rapid and quality publishing

International Journal of Electrical and Electronics Research (IJEER)
Review Article | Volume 9, Issue 3 | Pages 66-75 | e-ISSN: 2347-470X

Implementation of GF-HOG Technique for Effective Commercial and Industrial Load Clustering and Classification for Better Demand Response

Aniruddha Bhattacharya[1] and Madhusudan Singh[2]

[1]*Assistant Professor, Department of Electrical Engineering, Delhi Technological University, Delhi, India*
[2]*Professor, Department of Electrical Engineering, Delhi Technological University, Delhi, India*

***Correspondence:** Aniruddha Bhattacharya; Email: a.b.bhattacharya@dtu.ac.in ; Tel.: +91-9910977405

ABSTRACT: With increased penetration of renewables since last decade has evolved measures from the regulator for robust distribution network mostly catering to residential load. With increasing future demand of commercial and industrial load (CIL) due to aspiring GDP growth and insistence of popular governments to encourage renewable use, large chunk of the CIL will be served by renewables. So increasing the robustness alone on the renewable supply side will be in vain unless effective Demand response with rationalized tariff system for CIL which is more profitable energy market than subsidized residential tariff in India .The present method employed for clustering and classifying Load profile of CIL loads with wild swings will impede effective demand response implementation due to tariff cartelization in favor of not so popular energy systems. The method of GF-HOG technique with SVM will adopt cluster and classification of pattern of Load profiles for better recognition generated out from large CIL data. This will help produce more vibrant tariff structure for effective CIL load management. The algorithms were tested on a local network which 70% -30 % of load in favor of CIL and found to generate better DR.

Keywords: Gradient function histogram of oriented gradients (GF-HOG), Commercial and Industrial Load (CIL), Targeted Load Profile (TLP), Support vector mechanics (SVM), Demand Response (DR), Probability density function (PDF)

ARTICLE INFORMATION

Author(s): Aniruddha Bhattacharya and Madhusudan Singh
Received: Sep 24, 2021; **Accepted**: Jan 25, 2022; **Published**: Jan 30, 2022;
e-ISSN: 2347-470X;
Paper Id: IJEER-090307;
Citation: doi.org/10.37391/IJEER.090307
Webpage-link:
https://ijeer.forexjournal.co.in/archive/volume-9/ijeer-090307.html

1. INTRODUCTION

THE development of the electrical power distribution policy makers offer distribution companies fresh area of evaluating rationalized prices, for different types of consumers. For the effective tariff mechanism, the electricity buyer's categorization is based upon the type of customer (*viz.* residential, commercial and industrial) are sparsely correlated to the actual electrical consumption providing distorted image of load profiles distribution [1]. The load profiles of the customers and clubbing consumers exhibiting similar load profile is the hallmark of residential load distribution studies [2], [3]. The customer groups, is divided into major categories i.e. residential, commercial and industrial, Inside each category, fresh clustering and classification can be identified by their actual loading patterns of the consumers. A residential load customer, the load behavior for a specific loading condition refer to as Targeted Load profile(TLP) *They* can be produced by averaging the load data monitored from the Load dispatch center for a period 15 minutes to 30 minutes of measurements of load. The 24-hour load profile data are normalized and few numbers of clustering methods have been used in the previous literature building load profile, traditional statistical clustering methods [1], [4][5][6][7], fuzzy logic and neural networks [6], [8][9][10][11], The clustering will create formation of the *load patterns* suitable for residential load. Each pattern is created with load patterns of optimal number of clusters with smallest variance during classification. However, since the Load pattern required for commercial and industrial load profiling offer large swing of load due to intermittent industrial and commercial load operations so proper classification and clustering with traditional methods will cause irrational tariff structure and weird DR, various types of clustering methods have been investigated for accurate classification of customer load profile of wild swinging CIL with shortlisted substation servicing load mix which is predominantly CIL type From the initial set of TLPs (Target Load Profiles) three clustering algorithms i.e., K-means, fuzzy hierarchical clustering, and Summon maps are adopted. This implies applications of classical statistical clustering techniques [1, 4, 5, 7, 6], fuzzy logic [8] [9] [10], and neural networks [10, 11]. Every profile is created on account of time averaged load (KW) profiles in the same class and each activity class of industrial and commercial load and may be Target Load Profile of the class. The TLPs used in clustering and classification methods are normalized, the load cluster centroids cannot provide two-dimensional clustering. The data point is always marked to the nearest center, the above said techniques were not able to find load clusters other than spherical clusters [7]. Which always remain typical for the clustering requirement of this type of commercial and industrial load distribution. Industrial load requires the load profile in 2D form with different class of customer and different classes of activities within it as Load profile. With wild swing in the above profiles for hard loads, the model load

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Review Article | Volume 9, Issue 3 | Pages 66-75 | e-ISSN: 2347-470X

profiles (MLP's) based on conventional clustering and classification approach will nonlinear quadratic with large statistical errors.

1.1 Contribution to clustering solution for effective demand response of commercial and industrial load

(a) Firstly For large distribution-based data vis load profile data points, clustering with strongly overlapping peaks like the case discussed above, The realization of TLP's load data points are the mix of kernel probability distribution functions (PDF) [8]; 2D distribution of Gradient density distribution subjected to Euclidian distance over the kernel .The precision of these techniques depends on the accuracy of probability of the template to represent data.. The cluster centers with a higher local density surrounded with lower local density in the neighborhood. The edge detection of these densities leads to Gradient distribution. Each data point, two quantities of density and gradients are computed: its density ρ_i and its distance δ_i from high density point and their gradients. They depend on the distances δ_{ij} between high density data points, which satisfies the equality.

(b) Density estimation techniques provide an accurate estimate of a observable PDF based on an illustrated collection of time averaged load data [1] , [2], [3] ,[4] . The gradient density functions obtained are known as (GF_HOG) Gradient Function of histogram of oriented gradients and are used to find, swing detection in Targeted Load profile (TLP) [10] [11]. The method developed in [12] [13] with effective application of GF-HOG [15] supported by Kernel PCA(K-Principal Component Analysis) and Support Vector Mechanism(SVM) for detecting sketch based high swinging high density distribution

(c) The GF-HOG are primarily employed for pattern recognition techniques [6, 24]. The approaches provided in [19, 1] demonstrate an HOG application in association with support vector machines [20] for identifying probability density cluster image of 2D. In [10], Gradient Field HOG (GF-HOG) is applied to retrieval of sketch-based image the density of is taken as a probability function over a 2D domain, and the density distribution in 1D-PCA and gradients as 1D –PCA or edge detection PCA with SVM is used as the feature identifier to describe the wild swing in CIL load.

2. PAPER STRUCTURE

Section (3) will review of advance clustering and classification techniques to find suitable method for CIL load clustering in 2D.

Section (4) In K-medoids method, K value it is calculated as distance within data points. In the mean shift technique, is used to detect only spherical load clusters which seems unlikely in commercial and Industrial Load (CIL) load clustering in 2D and to find automatically correct number of clusters.

Section (5) will explore the popular concept used in the density estimation, these gradient density functions which are referred as GF-HOG which are primarily used for pattern recognition likely to be used in this research work for detecting distributed density gradient images of CIL Load [12] [13]. The approaches adopted in [14] [15] demonstrate an application of HOG for detecting wild swings in non-circular CIL load density images in 2D. In the lately published article [16], These HOG's are called the GF-HOG which are used primarily for sketch-based image retrieval seems ideally suited for detection of CIL load clusters. The 2D GF-HOG will reconstructed to 1D Gradient PCA vector and 1D Density distribution PCA vector with non-vanishing Eigen values.

Section (6) we provide experimental result to validate the GF-HOG use in feature extraction of Gradient density distribution images obtained from CIL loads. Clusters. The present work was inspired by recent work on application of pattern recognition to quantum supremacy [17] [18] [19]. This method basically used for density determination and gradient feature extraction from the images created out from substation data.

3. ADVANCE CLUSTERING AND CLASSIFICATION TECHNIQUES

3.1 Substation load profile 2D data distribution has three major steps.

Step 1: The substation metered load profiles data are mined for two features (1) data density and (2) feature and shape extraction which uses normalization for shape data only. Substations are grouped as per the TLP's shape, producing load cluster images using various iterative algorithms. An average normalized template, load shape of TLP within load group, to represent the overall load profile shape from the substations metered data with each cluster for metered over days.

Step 2: The relationship between fixed retrieved data and these developed clusters of substations is characterized. A classification method uses the multinomial Logical Regression model. It can mark on the HOG plot undetected substations metered data to each cluster with a probability completely based on the fixed data. This is imitable by time periods and unknown areas, for which real-time metered load (KW) data.

Step 3: The normalized vectors will be scaled up by suitable Eigen value to load magnitude density distribution vectors and shape vectors. For unmonitored substations, the actual scaling process is sometimes used to estimate peak loads distribution as bright spots in the images meant for feature extraction.

4. CLOSE COMPARATIVE STUDIES

4.1 K-Means clustering technique

In Low Voltage networks servicing CIL loads cannot be classified into micro categories, the clustering handles large volume of intermittently swinging and sometimes wildly varying meter data through the study period [25]. It cannot do assessment of the optimal clusters number. Number of Cluster were investigated in terms of the dissimilarity within each group but cannot converge for largely fluctuating CIL. The difference in the cluster is identified by the sum of Euclidian

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Review Article | Volume 9, Issue 3 | Pages 66-75 | e-ISSN: 2347-470X

error between substation metered load profiles within CIL groups. With a smaller number of clusters, the targeted set will be at the centroid. The Euclidian distance of all clusters is calculated as the error and the primary disadvantage of this approach is its inability to scale and iteratively convergent solution for deciding precise number of clusters with large fix of CIL loads, So more effective and traditional techniques like histograms used for pattern recognition ,which can compute the densities and gradient of different shapes called as Histogram of Gradients (GF-HOG), by smoothing the image, and computing the gradient and binning the resulting gradients can be useful to cluster and classify CIL Loads.. For the discrete N samples of Set S, it is important to know the faster rate of convergence of Eigen values of the proposed density estimation vector for GF-HOG method. For the kernel density estimators and the histograms used for shape vector estimation [14, 15], the convergence rates are the expected value (N samples size) of the square of the error between the computed and the true PDF's and are equal to be $O(N^{-1/3})$ and $O(N^{-3/5})$ [17],[21]

4.2 Multinomial logistic regression (MLR) for classification within the cluster groups

Logistic regression is used to predict categorical dependent variable y, outcome is used here to produce classification tools [26]. The independent variables (template) and the dependent variable x (Load clusters), the regression coefficients will produce probability of outcomes by logistic function expressed in (5)

$$P(y=1) = F(z) = \frac{1}{1+e^{-(b_0+b_1\times z)}}. \quad (5)$$

Where z = $\{z_{i1}, z_{i2}, \ldots\ldots z_{im}\}$ is set of fixed independent variables $b_k = \{b_{1k} \ldots\ldots b_{mk}\}$ are constant of regressions P(y=1) : Probability of the substation to remain in the first cluster.

The result of equation (5) is between 0 and 1, For N number of clusters (MLR), multinomial logistic regression is used to define the relationship between z values and P(y) membership values. For N clusters, the model will give N-1 logistic regression equations and probability of metered data assigned to a cluster that can be determined by (6) and (7).

$$P(Y_i = n) = \frac{e^{b_n \times z_i}}{1+\sum_{k=1}^{K-1} e^{b_k \times z_i}}, \quad n = 1, 2, \ldots K-1 \quad (6)$$

$$P(Y_i = K) = \frac{1}{1+\sum_{k=1}^{K-1} e^{b_k \times z_i}} \quad (7)$$

A substation meter data assigned to the cluster with highest probability. We can able to develop an estimate of simple mean-shift. The approach will provide observations about a point and related to the density gradient. A simple modification provides k-nearest neighbor estimate of mean-shift class. This is same as Loftsgaarden and Quesenberry's [11] estimate of k-nearest neighbor density, where within number of observations do not change, but the other estimates the volume of region remain same

4.3 Density Gradient Estimation

In issues related to pattern recognition, probability density function (PDF) is known in greater details. Limited information about density function, nonparametric techniques is used to obtain density gradient estimates. The nonparametric density estimates as discussed in Cacoullos' [5] Kernels of multivariate problems of Parzen's [4] univariate kernel estimates.is is possible n-dimensional space.

Let X (n) are N independent n-dimensional random vectors identically. Kernel density estimators of multivariate type as

$P_N(X) = (Nh^n)^{-1} \sum k(h^{-1}(X - Xj))$ where K(x) is the scalar functions satisfying

Sup[k(Y)]<inf

Y€ R^n

$$\int_{R^n} [k(Y)]dy < infinity \quad (8)$$

$$\lim_{[[Y]]+=} [[Y]]^n k(Y) = 0 \quad (9)$$

$$\int_{R^n} k(Y)\, dY = 1 \quad (10)$$

5. CLUSTER REPRESENTED USING NON-PARAMETRIC 2D DENSITY DISTRIBUTION IMAGES

Clusters having irregular shapes detected based on densities of data points. In density-based clustering of noise [11], density threshold is chosen, which discards noise as the points in regions of densities where threshold is greater, it assigns the data point to different clusters as outliers This drawback is not present in the mean-shift method of clustering [10, 11] is choosing an appropriate threshold. For a chosen group of M customers, each targeted Load Profile (TLP) of CIL load is characterized by a vector $x^m = \{x_h^m, h = 1,2,3,4 \ldots. H\}$, where H represents time-domain data time intervals of 15 min, 30 min, representing the Target Load Profile [1]. The whole set of data is X = { x^m with m =1,2,3,4 ….M} . The clustering process associates the initial M Target Load Profile (TLP) to customer classes, with each cluster represented as X^k C X contains n^k TLPs, for k= 1,2,3, 4.K. Let us further call $c^{k,}$ as the component vector having the centroid of the cluster There set of points representing density distribution function. That converge to a maximum this method is useful for finding non-spherical clusters. [10], it is useful to find non spherical clusters and to calculate number of clusters. The cluster centers are defined, as maxima of the density of data points. The algorithm converges to cluster centers which are surrounded by lower local density neighbors and have large

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Review Article | Volume 9, Issue 3 | Pages 66-75 | e-ISSN: 2347-470X

distance from few points of higher local density. For every data point, we find two quantities: *δi as its distance from points of higher density and ρi as its local density*. It depends on the distances δ_{ij} (the distance between data points), satisfies the Euclidian criterion. The density ρi of data point can be found as

$$\rho_i = \sum_j \chi(d_{ij} - d_c) \tag{11}$$

δc is a cutoff distance and ρi is equal to density of point distribution. The algorithmic only take as input, relative magnitude of ρi in different points. The choice of critical distance is key.is key for large data sets for the results to be robust with respect to δi which is the distance between any point and other point of higher density the highest density points, we represent δi = max (d_{ij}). The δi is larger than nearest neighbor distance only for points which are global and are density local maxima. The cluster centers are identified as data points with large δij. These observations, which, is illustrated by the simple example in *Figure 1A*. Solved using Euclidean distance framework described below illustrated diagrammatically by *Figure 1B* shows 29 points embedded

The Euclidean distance is used to measure different types of distances represented as follows

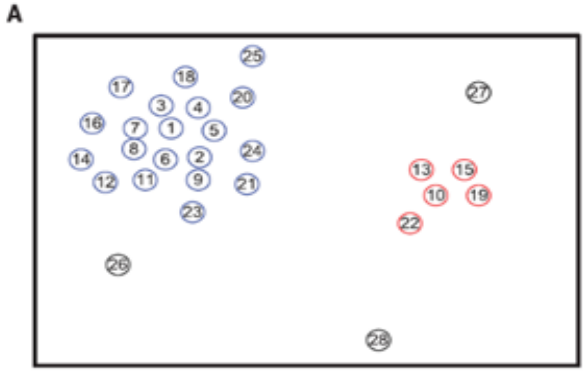
(1) Distance between the vectors with H components

$$d(\mathbf{y}, \mathbf{x}) = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(y_h - x_h)^2} \tag{12}$$

*(2) The distance calculated between the vector **y** and each of **M** members of set **X** i.e. Vector-to-set distance,*

$$d(\mathbf{y}, \boldsymbol{X}) = \sqrt{\frac{1}{M}\sum_{\mathbf{x}\in\boldsymbol{X}} d^2(\mathbf{y}, \mathbf{x})} \tag{13}$$

the mean distance of members x_q the set (for Q members) and y_i of set Y (for J members) The algorithm can be run iteratively for unlimited number of clusters with all TLP data accommodated in single image of density distribution clusters



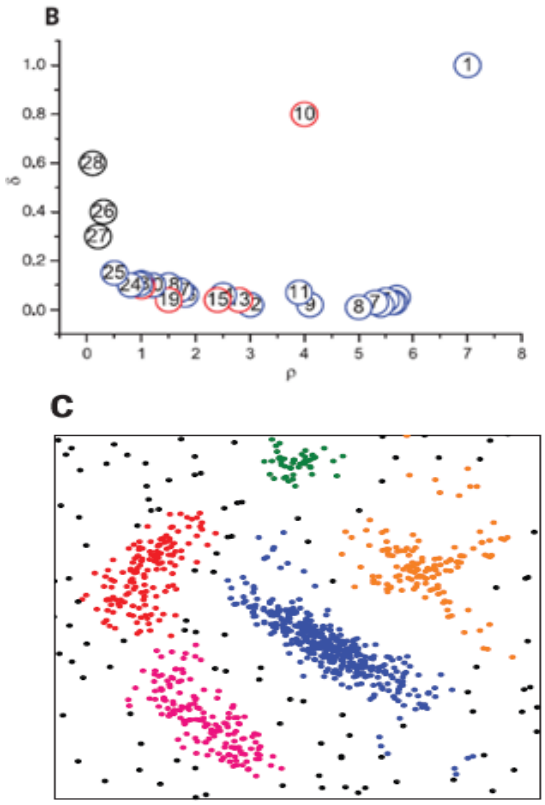
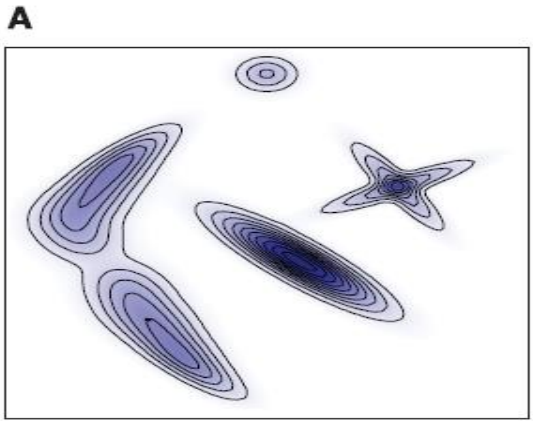
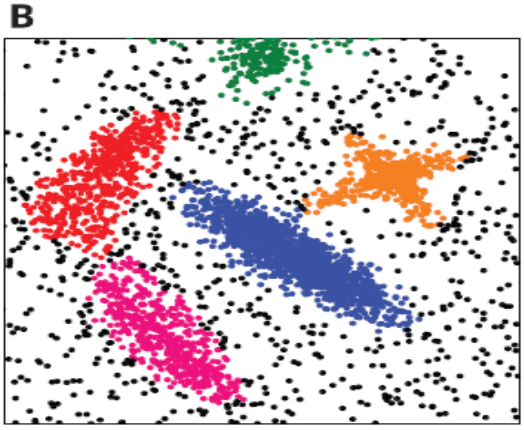


Figure 1: 2D representation. (A)Distribution of points... (B) Data Decision graph (C) different clusters represented with Different colors



(1) **X axis** represents **ρ** ***(2)*** **Y axis** represents **δ**



FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Review Article | Volume 9, Issue 3 | Pages 66-75 | e-ISSN: 2347-470X

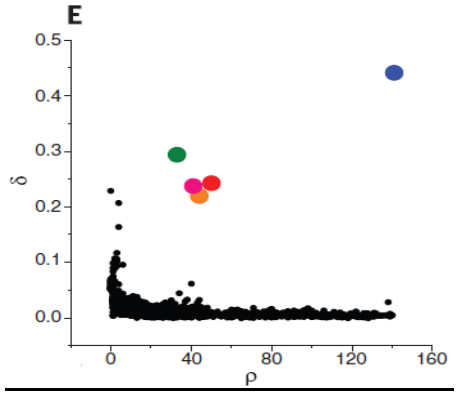


Figure 2: Images of point distributions. (A) The probability distribution. The regions with lowest density (B) Probability distributions for 5000 and 3000 sample points, respectively. Points are marked with different color according to the cluster. (E) The decision graphs, with colored centers of cluster.

5.2 Theoretical interpretation of the images obtained

The density distribution images represented in two-dimensional space distribution as shown in *Figure 2*. Then the maximum density is found at various points, which are centers of clusters. Figure 1B, the plot of ρi as a function of δi, mentioned as decision graph. The δ values of points 8 and 9, of same values of ρ, is different: Point 10 remains in cluster group of point 1, but higher ρ points are closest to cluster, but the neighborhood of high density point 9 belongs to different cluster. The only high δ and high ρ points are the centers of clusters. Points 25, 26, 27 are, outliers. When the cluster centers are found, other points are representing the same cluster as nearby points of highest cluster density. The cluster allocation is achieved in a single step, compared to other clustering and classification algorithms where objectivity is optimized iteratively [3, 7]. In clustering, it is also useful to asses reliability of an assignment. The approaches are obtained by function optimization [2, 8], and convergence is a natural defined measure.

5.3 Special kernel functions for identification non-regular density distribution

The 2D density distribution clusters are investigated using two components (1) Density (2) Density gradients by retaining Hessians functions at each point of the clusters and extracting one dimensional feature vector for easy identification used depending upon its prudence. The density gradient estimates do normally have kernel functions for gradient estimation. This result offers intuitive estimate for the density gradient mean shift. It can be expanded to a K nearest- neighbor approach in gradient estimation as discussed Loftsgaarden and Quesenberry's [11] extension of kernel density estimates.

5.4 Mean-Shift gradient estimates

The kernel Gaussian function is often the differentiable kernel function of multivariate type which satisfies the conditions for consistency of estimates of density gradient. The kernel stating probability Gaussian density function with zero mean with covariance matrix stated as

$$k(X) \equiv (2\pi)^{-n/2} \exp\left(-\tfrac{1}{2}X^T X\right). \tag{12}$$

$$\hat{\nabla}_x p_N(X) = N^{-1} \sum_{i=1}^{N} (X_i - X)(2\pi)^{-n/2} h^{-(n+2)} \cdot \exp\left[-(X - X_i)^T \left(\frac{X - X_i}{2h^2}\right)\right]. \tag{13}$$

Intuitive interpretation of (13) is a mean shift of the measure of weighted the observations about the point X. The mean samples of the weighted shifts are taken as estimate of gradient. This will give the kernel function is of the form

$$k(X) = g(X^T X). \tag{14}$$

$$k(X) = \begin{cases} c(1 - X^T X), & X^T X \le 1 \\ 0, & X^T X > 1 \end{cases} \tag{15}$$

Where

$$c = \pi^{-n/2} \left(\frac{n+2}{2}\right) \Gamma\left(\frac{n+2}{2}\right) \tag{16}$$

C is the constant for normalizing required to integrate the kernel function to 1. This kernel function can satisfy the conditions for consistency of the estimate of gradient and similar to the kernel stated in Epanechnikov [lo]. By putting (15) in (17), we compute the gradient in (18)

$$\hat{\nabla}_x p_N(X) \equiv (Nh^n)^{-1} \sum_{j=1}^{N} \nabla_x k(h^{-1}(X - X_j)) \tag{17}$$

$$\hat{\nabla}_x p_N(X) = (Nh^{n+2})^{-1} 2c \sum_{X_i \in S_h(X)} (X_i - X) \tag{18}$$

$$= \left(\frac{k}{N v_h(X)}\right) \frac{n+2}{h^2} \left(\frac{1}{k} \sum_{X_i \in S_h(X)} (X_i - X)\right)$$

$$v_h(X) \equiv \int_{S_h(X)} dY = \frac{h^n \pi^{n/2}}{\Gamma(n + 2/2)} \tag{19}$$

The volume of the region

$$S_h(X) \equiv \{Y : (Y - X)^T (Y - X) \le h^2\}$$

and k is the number of TLP data points within $S_h(X)$ and, Equation (18) provides with estimation of the gradient estimation process. The last term in (19) is a mean shift od data points of the observations in S(X) of X.

$$M_h(X) \equiv \frac{1}{k} \sum_{X_i \in S_h(X)} (X_i - X) \tag{20}$$

If the gradient values are zero, then it represents uniform density over cluster S(X), the average mean shift is zero with a nonzero density gradient heading to quick increase in PDF, the time average data points will align along the direction than anywhere else in S(X). The mean shift average will head along the direction and magnitude of the gradient is proportional to length [13]

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Review Article | Volume 9, Issue 3 | Pages 66-75 | e-ISSN: 2347-470X

5.5 Gradient Clustering Algorithm

The method of clustering a data points is to assign each data points to the nearest mode along the direction of the gradient at that point. One could move each observation with small step in the direction of gradient and iteratively repeat the process till denser clusters produces near the modes. The other approach is to shift each observation by similar amount proportional to the gradient of the observation point [14]

$$X_j^0 \equiv X_j, \quad j = 1,2,\cdots,N \tag{21}$$

$$X_j^{i+1} = X_j^i + a\nabla_x \ln p(X_j^i). \tag{22}$$

Value of (a) is optimally chosen to obtain convergence. N-dimensional analog can be found putting into the roots of the equation

$$\nabla_x p(X) = 0 \tag{23}$$

Above equation represents the combined density p(X). The minima will be the roots of (23), the algorithm resulted from (22) will move data points away as the gradients moves away from them. From every iteration, points will move closer to center. The following two observations are

(1) The step size for the gradient which is same near the centroid will be lesser than that of farther point. This will cause data points move far from the centroid and near to the local minimum which move to the mode quicker.
(2) Gradient normalization is considered if a mean M Gaussian density and covariance matrix represented as

$$p(X) = (2\pi)^{-n/2} \exp\left[-\tfrac{1}{2}(X - M)^T(X - M)\right]. \tag{24}$$

$$X_j^1 = X_j + \nabla_x \ln p(X_j) = X_j - (X_j - M) = M. \tag{25}$$

Gaussian density, cluster merged to single point after one iteration. This will help mix of density problem of which each component densities which look Gaussian

(3) The normalized gradient, convergence can be achieved as mean-square sense. [23][29]

$$Y = X + a\nabla_x \ln p(X) \tag{26}$$

Minimize covariance of our samples with smaller (a) as shown in (27) & (28),

$$E\{Y - M_Y)^T(Y - M_Y)\} \le E\{(X - M_X)^T(X - M_X)\} \tag{27}$$

$$(E\{Z\})_i \equiv \left(\int_{R^n} \frac{\nabla_x p(Y)}{p(Y)}\, p(Y)\, dY\right)_i \tag{28}$$

5.6 Complete Gradient Cluster Algorithm

1. Set of *n*-dimensional vectors as random sample for n-dimensional random variable *X*, which has density distribution of *E as* described in *Equations (24)-(28)*.
2. Using the methodology described, he kernels estimator Eˆ* can be computed.
3. Taking the clusters particularly to local maxima of Eˆ function, by transposing those elements in the gradient ∇Eˆ direction described in *Equation (28)*, with the appropriate step.

The Gradient Clustering Algorithm, are of classic Newton method, defined as

$$x_j^{k+1} = x_j^k + b\frac{\Delta f(x_j^k)}{f(x_j^k)} \text{ For j=1…m and k=0, 1……..k*} \tag{29}$$

4. Formulate the kernel estimator Eˆ;
5. Identify condition to stop the program (find also number of steps represented as *k**);
6. Initiation of process to create cluster with particular elements are assigned to set after the last, *k** step;
7. Analyze the effect of the values of parameters on obtained results.

6. HOG FOR CLUSTER TYPE DETECTION

This section explains feature extraction chain of the 2D density model functions, which is summarized in *Figure 3*. Evaluation of local histograms normalized over image gradient in a dense grid as described in Section 2 and 3. Similar features is explained in pattern recognition methods [16, 18, 19, and 20].

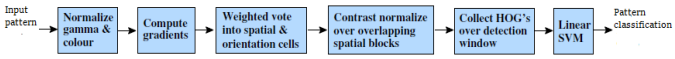


Figure 3: Object detection chain and feature extraction block chain

The local density distribution gradients or sometimes called as edge directions can be characterized by local object appearance and shape, without having corresponding edge positions or gradients. Then by splitting the image into proportional spatial cell and constructing a local 2-D histogram of density and direction of gradient. This is a measure of local histogram over larger distributed regions and to normalize all cells. This is known as 2 D Histogram of Oriented Gradient (HOG) descriptors. HOG descriptors are combined with feature vector for conventional SVM classifier provides detection chain the orientation histograms offer many precludes [22, 23, 24], but when combined with distributed histogram and normalized in Lowe's Scale Invariant Feature Transformation (SIFT) to precise image matching [30], either image patch descriptor for matching scale invariant points. SIFT perform well in this application [30, 31]. "The Shape Context studied in [32] provides alternative block shapes, although initially using only edge features without orientation histogram that makes the representation effective". The HOG representation has many advantages. It detects density and gradient structure that is very unique of local shape, and easily has controllable invariance. The 2-D histogram is represented, two parameters of image i.e. one is the image density, which is denoted and the other is the gradient of the image [6] [15].

6.1 Improved 2D Gradient Histography Image

The 2-D histogram is composed of density image, which is denoted by f(x, y), and gradient image. For the given function

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Review Article | Volume 9, Issue 3 | Pages 66-75 | e-ISSN: 2347-470X

f(x, y), the gradient at the location (x, y) is defined as the 2-D column vector [6]

$$\nabla f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right]^T. \quad (48)$$

For a discrete image representing HOG-GF, This will generate the gradient which signifies running difference of density over the horizontal and vertical axis of the image, which is as follows:

$$\frac{\partial f}{\partial x} = f(x+1, y) - f(x-1, y) \quad (49)$$

$$\frac{\partial f}{\partial y} = f(x, y+1) - f(x, y-1). \quad (50)$$

Horizontal and vertical gradients effectively can be used to find combined gradient g(*x*, *y*) to find gradient of images as follows. This will remove noise and outliers from the space for better image analysis

$$g(x, y) = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}. \quad (51)$$

Each distributed cluster image has its GF pair which consist of united gradient (G) and density (F) representing density. Traversing clusters of images, and computing the frequency of G-F pair is easily possible. If p_{ij} is the frequency of G-F pair (i, j), The cluster density will be represented by

$$p_{ij} = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \delta_{ij} \quad (52)$$

6.2 Mathematical analysis for retrieving two 1D Principal Component Analysis (PCA) components using Unsupervised Learning Algorithm

The consolidate noise free 2 D images are sorted by unsupervised algorithm to retrieve two 1D density images of density and gradient. Following is a sample numerical exercise solved and can be replicated for large data set of running data from substations

Step 1: Data Set

Substation Nos	Mean data of 15 minutes (D1)	Mean data of the next 15 minutes (D2)	Mean data of the next 15 minutes (D3)
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

Step 2: Spread it in matrix and compute mean
The mean matrix A = [66 60 60]
Step 3: Compute covariance

	D1	D2	D3
D1	504	360	180
D2	360	360	0
D3	180	0	720

Step 4: Calculating one dimensional PCA using Eigen values and vector

Coefficients are values are λ= 44.81 λ= 629.11 λ=910.01
Eigen values = -3.75; -0.505; -1.05

Step 5: Sort the eigenvectors with decreasing eigenvalues and choose eigenvectors of largest eigenvalues to form a d × k dimensional matrix W.

Eigen values in decreasing order
910.06; 629.11; 44.81

Eigen vector corresponding to two highest eigenvalues

W= 1.05594 -0.594
0.69108 -0.675
0 0

6.3 HOG-GF image segmentation through 2D Principal Component Analysis (PCA)

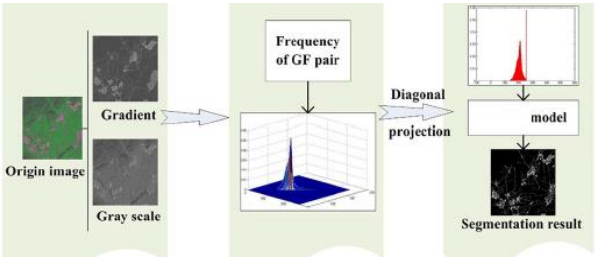


Figure 4(i)

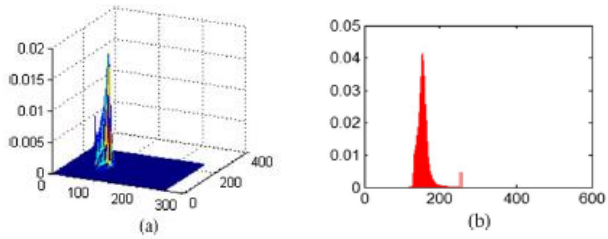


Figure 4(ii)

Figure 4: ***(i)*** Improved gradient of histogram of 1D PCA histogram density image; ***(ii)*** (**a**) Improved 1D PCA gradient histogram. (**b**) Two dimensional PCA of gradient histogram

6.4 MAD (Mean Average Distance) Model

For 1D histogram, k is considered as threshold segment value for binary bits. The binary segments, density distribution and its gradient images are splitted into two classes i.e. of density and its gradient corresponding to r $\in \{0, 1, 2 \ldots k\}$ in the histogram of 1D size [23]

$$\begin{cases} \omega_0(k) = \sum_{r=0}^{k} p_r \\ \omega_1(k) = \sum_{r=k+1}^{2(L-1)} p_r \end{cases} \quad (54)$$

$$\begin{cases} \mu_0(k) = \sum_{r=0}^{k} r p_r \\ \mu_1(k) = \sum_{r=k+1}^{2(L-1)} r p_r. \end{cases} \quad (55)$$

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Review Article | Volume 9, Issue 3 | Pages 66-75 | e-ISSN: 2347-470X

Next, we can produce mean absolute deviation on the 1D histogram, where sample mean is μ, and sample variance is σ where MAD is Mean Average Distance models in 1D space

$$\begin{cases} \mathrm{MAD}_0(k) = \sum_{r=0}^{k} |r - \mu_0(k)|\, p_r/\omega_0(k) \\ \mathrm{MAD}_1(k) = \sum_{r=k+1}^{2(L-1)} |r - \mu_1(k)|\, p_r/\omega_1(k). \end{cases} \tag{56}$$

The residential loads, the gradients and densities are same within one type of image. Unlike in CIL load, 2D histogram images have density and its gradient, it appears as a Gaussian curve, and threshold will be the lowest descendent in between modes. The modes of two images are represented by $f(x|\mu 0, \sigma 0)$ and $f(x|\mu 1, \sigma 1)$,respectively. Each probability of $f(x|\mu 0, \sigma 0)$ and $f(x|\mu 1, \sigma 1)$ is $\omega 0(x)$ and $\omega 1(x)$, and the function of the 1-D histogram is shown as follows [23]:

$$h(x) = \omega_0(x) f(x|\mu_0, \sigma_0) + \omega_1(x) f(x|\mu_1, \sigma_1)$$

$$\text{where } 0 \le x \le 2(L-1). \tag{57}$$

For the optimal threshold is represented as $k*$, where $k*$ satisfies

$$k^* = \frac{\frac{\sigma_0 \sigma_1}{\sqrt{2}} \ln\left(\frac{\omega_0 \sigma_1}{\omega_1 \sigma_0}\right) + \mu_1 \sigma_0 + \mu_0 \sigma_1}{\sigma_0 + \sigma_1}. \tag{58}$$

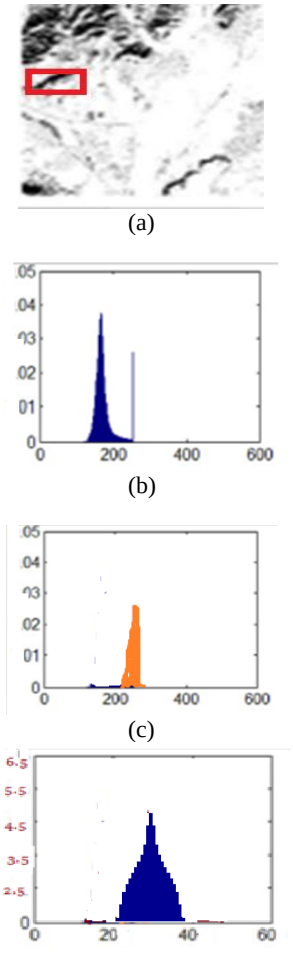
The mean deviation in absolute of σ_1 *and* σ_0 is shown as follows:

$$\sum_{x=0}^{k} f(x|\mu_1, \sigma_1)\, |x - \mu_0(k)| + \sum_{x=k+1}^{2(L-1)} f(x|\mu_1, \sigma_1)\, |x - \mu_1(k)|$$
$$= f(x|\mu_1, \sigma_1)\, |\mu_1 - \mu_1(k)| = |\mu_1 - \mu_1(k)|\,. \tag{59}$$

7. EXPERIMENT AND RESULT ANALYSIS

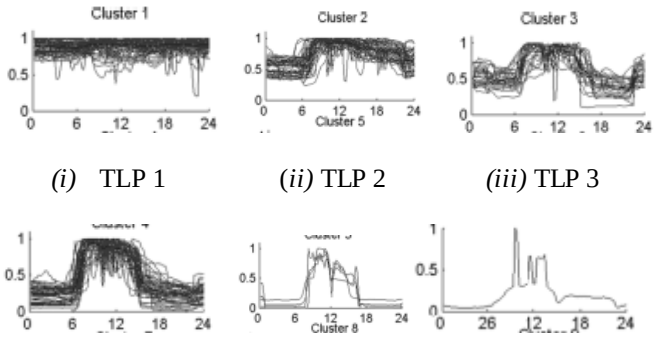
A set of CIL load customers connected to the Medium Voltage distribution system with reasonable penetration of renewables having 162 Nos 33KV/411KV 3 Phase AUSGRID NSW Load dispatch center at Australian State of New South Wales has been considered [12]. The system was considered due to mix of residential and nonresidential loads having 75% in favor of CIL loads with the Target Load Profile (TLP) of a customer is found by calculating the mean of 15-min interval in a day data od Load. The original vector space with specific set of data will then be with magnitude and shape of the TLP which are closely monitored. The clustering algorithm assigns to each TLP of every customer to a particular cluster, providing predominantly without overlapping for all customers. The Gradient clustering algorithms used require an optimal number of clusters to be formed. The clusters numbers decrease with increase in the distance threshold [9], and the shape of the clusters considering CIL loads remain non-overlapping. Some indications optimum number of clusters from the view point of classification may be obtained from cluster density and gradient cluster function optimization. The best number of clusters has been found to be the knee of the curve principle [27]. The cluster patterns are obtained from algorithms described in *Section 4*.

Sample images – 1 of TLP cluster image constructed and its ID conversion



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

Figure 5: (a). Sample marked in Red (b). PCA (Eigen Value) of (c). PCA of Gradient Distribution Represents of typical CIL Load densities distribution of the sample marked in Red (d). Covariance distribution



(i) TLP 1 *(ii)* TLP 2 *(iii)* TLP 3

(iv) TLP 4 *(v)* TLP5 *(vi)* TLP 6

Figure 6: Images of the clusters formed during 24-hour period of the day from the Sample images- 1 Clustering results for the HOG-GF METHOD with K = 6 clusters

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Review Article | Volume 9, Issue 3 | Pages 66-75 | e-ISSN: 2347-470X

Sample images – 2 extracted from the substation and its ID conversion

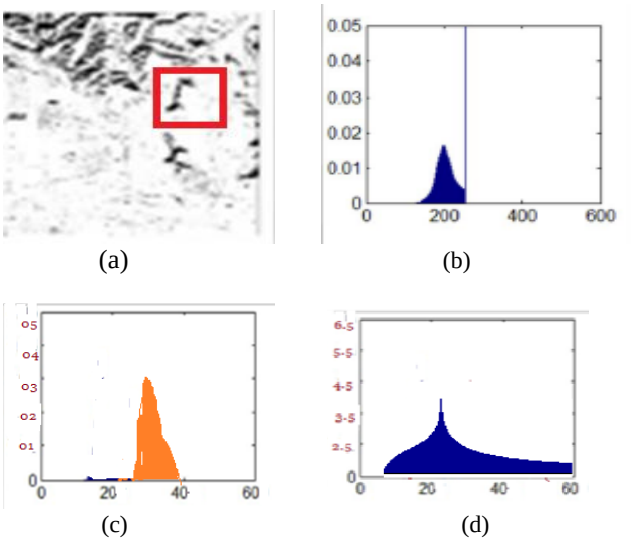


Figure 7: (a). Sample marked in Red (b) PCA of densities distribution (c). PCA of Gradient Represents of typical CIL Load of the sample marked in Red Distribution Cluster distribution (d) Covariance distribution

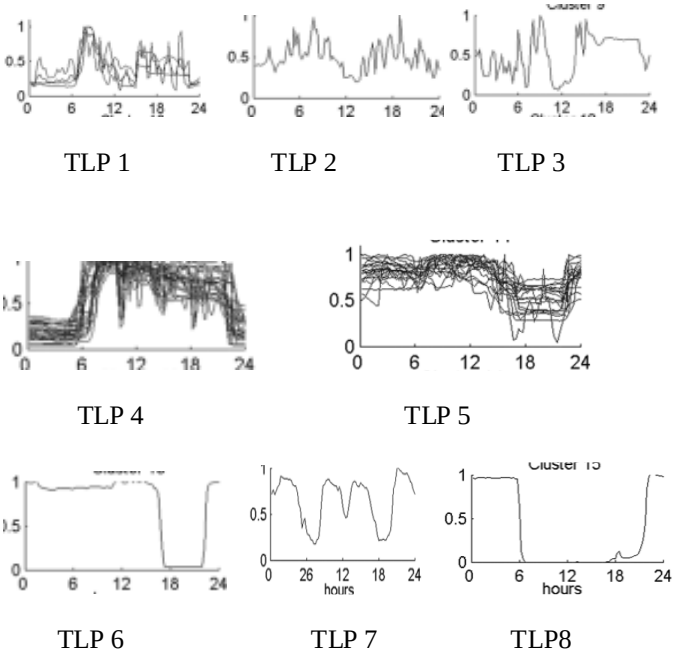


Figure 8: Images of the clusters formed during 24-hour period fo the day from the Sample image 2 Clustering results for the HOG-GF with K = 8 clusters

7.1 Result analysis

The results clearly indicate the importance of the HOG-GF method compare to other techniques to represent the density and its shape. This superiority also depends on techniques to build separate classes of TLP's and to discriminate load patterns for CIL Loads, as in Fig. 3 unlike hierarchical clustering. The various clustering techniques will be useful to find the number of TLP's, equal to number of clusters i.e., is a cluster center. However, with increasing fluctuations in the load pattern with decrease in cluster numbers, every method adopts a different strategy for load pattern grouping, the customer partitioning resulting, compute the centroids from the set of data points. [27][28]

8. CONCLUSION

The distribution service providers use Load clustering techniques for density distribution of load pattern for electricity customer classification. The study carried here emphasize most important aspects for the distribution companies for the possible use of clustering and Pattern recognition techniques for gradient density distribution. The indicative number of CIL load customer classes using the GF-HOG method take into consideration the intermittently turbulent CIL load, even though criteria to find out the best number of clusters could result in higher values The results of the clustering quality assessment done in this paper show that one algorithms—HOG-GF, will give highly detailed isolation of the clusters, discriminating load patterns which seems perfectly suited to large swinging CIL Loads. These avenues provide algorithms will perform classification, by making customer groups into a number of customer classes particularly fit for a customer for tariff formation purposes for CIL loads. The various algorithms are used to distribute the TLP's within some clusters and, as such, are less effective for CIL Load.

REFERENCES

[1] G. Chicco, R. Napoli, P. Postolache, M. Scutariu, and C. Toader, "Customer characterization options for improving the tariff offer," IEEE Trans. Power Syst., vol. 18, no. 1, pp. 381–387, Feb. 2003

[2] J. MacQueen, in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, L. M. Le Cam,J. Neyman, Eds. (Univ. California Press, Berkeley, CA, 1967), vol. 1, pp. 281–297.

[3] E. Parzen, "On estimation of a probability density function and mode," Ann. Math. Statist. yol. 33, pp. 1065-1061, 1962

[4] T. Cacoullos. "Estimation of a multivariate densltv." Ann. Instkatist. Math:, vol. 18, pp. 179-189, 1966

[5] N. N. Cencov, Estimation of an unknown distribution density from observations, Soviet Math., 3 (1962), pp. 1559–1562.

[6] A. Pentland, "Looking at People: Sensing for Ubiquitous and Wearable Computing," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 107-119, Jan. 2000

[7] C. S. Chen, J. C. Hwang, Y. M. Tzeng, C. W. Huang, and M. Y. Cho,"Determination of customer load characteristics by load survey system at Taipower," IEEE Trans. Power Del., vol. 11, pp. 1430–1436, 1996.

[8] G. J. McLachlan, T. Krishnan, The EM Algorithm and Extensions (Wiley Series in Probability and Statistics vol. 382, Wiley-Interscience, New York, 2007

[9] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, in Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, E. Simoudis, J. Han, U. Fayyad, Eds. (AAAI Press, Menlo Park, CA, 1996), pp. 226–231.

[10] G. Chicco, R. Napoli, F. Piglione, M. Scutariu, P. Postolache, and C. Toader, "Load pattern-based classification of electricity customers," IEEE Trans. Power Syst., vol. 19, no. 2, pp. 1232–1239, May 2004.

[11] D. 0. Loftsgaarden and C. P. Quesenberry, "A nonparametric estimate of a multivariate density function," Ann. Math. Statist.,vol. 36, pp. 1049-1051, 1965

[12] "Sample report of Northern India state load dispatch " provided by SLDC Delhi

[13] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review,"ACM Comput. Surv., vol. 31, no. 3, pp. 264–323, 1999.

[14] C. S. Chen, M. S. Kang, J. C. Hwang, and C. W. Huang, "Synthesis of power system load profiles by class load study," Elect. Power Energy Syst., vol. 22, pp. 325–330, 2000.

[15] C.M. Bishop, Pattern recognition and machine learning (Information science and statistics),Springer, New York, NY, 2006.

[16] W. T. Freeman and M. Roth. Orientation histograms for hand gesture recognition. Intl. Workshop on Automatic Faceand Faceand Gesture- Recognition, IEEE Computer Society, Zurich,Switzerland, pages 296–301, June 1995.

[17] D. Scott, On optimal and data-based histograms, Biometrika, 66 (1979), pp. 605–610.

[18] W. T. Freeman, K. Tanaka, J. Ohta, and K. Kyuma. Computer vision for computer games. 2nd International Conference on Automatic Face and Gesture Recognition, Killington,VT, USA, pages 100–105, October 1996

[19] D. G. Lowe. Distinctive image features from scale-invariant keypoints. IJCV, 60(2):91–110, 2004

[20] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. IJCV, 60(1):63–86, 2004

[21] G. Wahba, Optimal convergence properties of variable knot, kernel, and orthogonal series methods for density estimation, Ann. Stat., 3 (1975), pp. 15–29.

[22] R. K. McConnell. Method of and apparatus for pattern recognition,January 1986. U.S. Patent No. 4,567,610.

[23] N. N. Cencov, Estimation of an unknown distribution density from observations, Soviet Math., 3 (1962), pp. 1559–1562.

[24] W. T. Freeman, K. Tanaka, J. Ohta, and K. Kyuma. Computer vision for computer games. 2nd International Conference on Automatic Face and Gesture Recognition, Killington, VT, USA, pages 100–105, October 1996

[25] D. W. Hosmer and S. Lemeshow, Applied Logistic Regression. New York, NY, USA: Wiley-Interscience, 2000, vol. 354.

[26] D. Gerbec, S. Gasperic, I. Smon, and F. Gubina, "Allocation of the load profiles to Trans.Power Syst., vol. 20, no. 2, pp. 548–555, May 2005.

[27] G. Schwarz, "Estimating the dimension of a model," Ann. Stat., vol. 6, pp. 461–464, 1978.

[28] D. M. Gavrila. The visual analysis of human movement: A survey. CVIU, 73(1):82–98, 1999.

[29] D. G. Lowe. Distinctive image features from scale-invariant keypoints. IJCV, 60(2):91–110, 2004.

[30] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. PAMI, 2004. Accepted

[31] S. Belongie, J. Malik, and J. Puzicha. Matching shapes. The 8th ICCV, Vancouver, Canada, pages 454–461, 2001



© 2021 by the Aniruddha Bhattacharya and Madhusudan Singh. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Investigations on Plasma Pretreatment of Catalyst Film and Catalyzed Growth of Carbon Nanotubes

Umang Sharma and Suresh C. Sharma, *Senior Member, IEEE*

***Abstract*— The catalyzed-growth process of carbon nanotubes (CNTs) in a C_2H_2/H_2 plasma is established theoretically in the current study to explore the effect of plasma pretreatment of catalyst nanofilm and its effect on CNT growth. We have also deliberated different catalyst nanoparticles and their consequence on the nanotube growth considering a theoretical model which includes the dynamics of catalyst nanoparticle, plasma sheath, and plasma species (ions, neutrals, and electrons) and the generation of atomic C and H species on the catalyst surface, taking into contemplation several processes essential for the plasma-assisted CNT growth. The theoretical study aims to improve the understanding of using different catalysts for various scientific fields by unraveling their effect on CNT growth. The pretreatment effects on the catalyst nanofilm for nanoparticle formation are deliberated. The effects of increasing catalyst nanofilm thickness on catalyst size and density were attained. For the same plasma conditions, different metal catalysts (Ni, Fe, and Co) were studied for their influence on CNT growth. The results of this theoretical study were compared with the scanning electron microscopy (SEM), transmission electron microscopy (TEM), and other experimental observations and can provide a better insight into plasma aided catalytic synthesis of CNT growth for various applications.**

***Index Terms*— Carbon nanotubes (CNTs), catalytic growth, plasma-enhanced chemical vapor deposition (PECVD), plasma parameters.**

I. Introduction

FOR a few decades, carbon nanotubes (CNTs) [1] have emerged out as a significant material for numerous applications in the growing nanotechnology area, which are produced via various methods, such as arc discharge [2], laser ablation [3], chemical vapor deposition (CVD) [4], and plasma-enhanced CVD (PECVD) [5], [6]. The synthesis of CNTs via PECVD has developed a considerable interest of researchers due to their vertical alignment in the potential applications in field-emission devices [7]–[9]. The plasma produces an electrostatic force on the growing nanotube because of the plasma sheath field, aligning it vertically to its growth axis [10]. The synthesis of CNTs can be of catalytic or noncatalytic type. The catalytically grown CNTs have improved characteristics, such as yield and quality of the CNTs [11], [12]. The transition metal nanoparticles are of utmost effectiveness owing to their high diffusion rate and carbon solubility. Thorough studies on the simple *in situ* production of heteroatom-doped carbon nanomaterials for efficient catalytic applications have recently been demonstrated [13], [14]. A catalyst nanofilm deposited over the substrate for the catalytic method transforms into catalyst nanoparticles (via plasma processing or etching) [15] which acts as the seed layer, essential for the nucleation of CNTs [16], [17]. Numerous experiments determine that the catalyst nature and size govern the nanotube's growth, which can be altered on varying the plasma conditions [18], [19]. The transition metals, namely, iron (Fe), cobalt (Co), and nickel (Ni), can be used as a pure metal catalyst for CNT synthesis [20]–[24]. A metal nanofilm predeposited over a substrate is etched in the presence of plasma to obtain metal catalyst nanoparticles. This pretreatment can alter the surface with preferred characteristics by adjusting the plasma chemistry and plasma operating conditions. Yen *et al.* [25] suggested that hydrogen pretreatment is essential for catalytic activity preceding CNT growth for the CVD process. They obtained worm-like CNFs for thermally annealed nanofilm and spaghetti-like CNTs for hydrogen plasma etching of the film. Chang *et al.* [19], [26] used pure hydrogen for pretreatment of nickel catalyst film to break it into nanoparticles and showed that plasma pressure and power influence the catalyst size greatly and hence the diameter of CNT. Yang *et al.* [27] showed plasma treatment using hydrogen, nitrogen, and oxygen for a nickel film and deliberated improvement in CNT number density and quality. Wang *et al.* [28] exhibited consistent growth of multiwalled CNTs on pretreatment of catalyst nanofilm with NH_3. They also described the influence of hydrogen gas flow on the dimension of nickel catalyst [29]. The catalyst and CNT parameters were plasma power and pretreatment time-dependent. Hoyos-Palacio *et al.* [23] produced CVD grown CNTs using nickel, cobalt, iron, and molybdenum as catalysts and observed the finest growth with nickel. Huang *et al.* [24] studied the effect of Ni, Fe, and Co catalyst on nanotube growth via PECVD. Valentini *et al.* [30] observed the effect of catalyst layer thickness on CNT growth parameters. Labbaye *et al.* [20] also investigated the catalytic effect on CNT production. Wang *et al.* [22] studied the formation of catalyst nanoparticles and their underlayer for the CNT growth. Chang *et al.* [26] and Jian [31] obtained that CNTs synthesized on catalsyt were well-aligned and had a similar diameter and density as catalyst on treatment with H_2 or H_2/N_2 plasmas.

Manuscript received October 12, 2021; revised January 26, 2022; accepted February 22, 2022. The work of Umang Sharma was supported by the Department of Science and Technology (DST), Government of India, India, through the DST-Innovation in Science Pursuit for Inspired Research (INSPIRE) Fellowship Program. The review of this article was arranged by Senior Editor S. J. Gitomer. *(Corresponding author: Suresh C. Sharma.)*

The authors are with the Department of Applied Physics, Delhi Technological University (DTU), Delhi 110042, India (e-mail: suresh321sharma@gmail.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPS.2022.3155172.

Digital Object Identifier 10.1109/TPS.2022.3155172

0093-3813 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 23,2022 at 07:05:40 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
PLASMA SPECIES DELIBERATED IN THE CURRENT STUDY

	Neutral species	Positive species
A	$C_2H, C_2H_2, C_2H_3, C_2H_4$	$C_2H_2^+, C_2H^+, C_2H_3^+, C_2H_4^+$
	$C_4H_3, C_4H_2, C_6H_2, C_6H_4$	$C_4H_3^+, C_4H_2^+, C_6H_4^+, C_6H_2^+$
B	H, H_2	H^+, H_2^+

In the current study, we examine plasma pretreatment effects on catalyst nanoparticles and CNT growth. The dynamics of catalyst film thickness, catalyst size, and their effects on the growth parameters of the nanotube is also studied. The model will also consider the consequences of using different catalysts for CNT nucleation and growth in plasma. The model developed in Section II shows the catalyzed-growth mechanism of CNT and consists of the kinetics of the plasma sheath, various time-differential equations of plasma species (namely ions, neutrals, and electrons), the kinetics of the different catalyst particles, involving various plasma processes, such as diffusion and dissociation, of the carbon species on the nanoparticle, discharge losses of the plasma chamber, and the CNT growth calculations. The outcomes of the present model are deliberated in Section III, which would help in revealing the role of different catalysts in the CNT growth. The inference from the study is mentioned in Section IV.

II. CATALYZED CNT GROWTH MODEL

The catalyst-assisted growth of CNTs aims to increase nanotube production, better growth, enhance properties, or modify their structures for various applications. The most commonly used catalyst for CNT production via PECVD is transition metals, such as iron (Fe), cobalt (Co), and nickel (Ni). The nanoparticles are achieved from the nanofilms via the plasma processing technique. Along with being fast, this method helps achieve catalyst nanoparticles of high quality and activity compared with conventional methods. Plasma processing leads to well-dispersed and smaller catalyst particles having a strong interaction with the substrate leading to better activity and support. Many chemical reactions and species are considered to demonstrate the phenomena of nucleation and growth mechanism of CNTs [10]. An analytical study has been performed investigating the effect of plasma pretreatment on the catalyst nanoparticles and their result on the CNT growth. The role of different catalyst nanoparticles on the growth of CNT is also deliberated. The study incorporates particle balance equations, energy-momentum balance equations, and Poissons equation that are solved numerically for a reactive plasma constituted by a primary hydrocarbon gas-acetylene (C_2H_2) represented as A and etchant gas-hydrogen (H_2) represented as B throughout the work, respectively. The most prevailing hydrocarbon and carrier species considered in the study are listed in Table I.

The study considers eight acetylene ions, two hydrogen ions, eight acetylene neutrals, and two hydrogen neutrals, respectively. The remaining species were omitted because they primarily sputter and etch out amorphous carbon from the catalyst surface and have a diminishing role in the CNT growth. The graphic illustration of the catalytic CNT growth

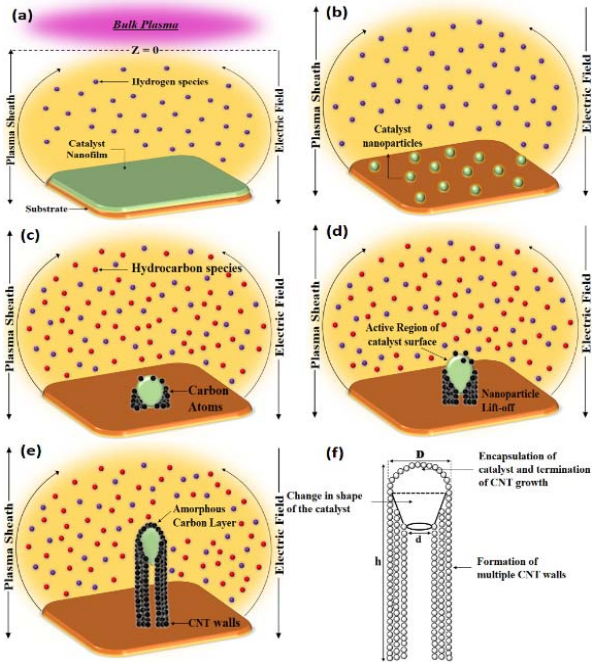


Fig. 1. Catalytic CNT growth representation (a) nanofilm deposited on substrate, (b) hydrogen plasma segregates nanofilm into nanoparticles, (c) hydrocarbon species dissociate on catalyst surface into C atoms which diffuse over catalyst surface and bulk and initiate growth, (d) with diffusion, C atoms precipitate at the rear end of catalyst forming graphitic walls, (e) when catalyst is covered with amorphous carbon layer, growth terminates, and (f) catalyst gets fully encapsulated inside CNT, multiple walls are formed, and catalyst shape also changes due to wall stresses.

model in a reactive plasma is represented in Fig. 1. The initialization of nanotube growth starts on the facet of metallic catalyst nanoparticles required to disintegrate hydrocarbon species. The carbide formation allows the decomposition segregation of carbon atoms to penetrate through the nanoparticle and regenerate into graphitic cylinders. An interface between the bulk plasma and plasma sheath at $z = 0$ is considered to study the plasma sheath model, electronic, ionic, neutral fluxes, and plasma species energies. The following steady-state fluid equations [32], [33] constitute this 1-D model within the plasma, which is composed of a neutral ($z < 0$) and a nonneutral region ($z > 0$):

$$\frac{\partial}{\partial z}\left(n_{ij}u_{ij}\right) = n_e \nu_{\text{if}} \tag{1}$$

$$\frac{\partial}{\partial z}(n_e u_e) = \sum_{ij} n_e \nu_{\text{if}} \tag{2}$$

where n_{ij} and n_e are the number densities, u_{ij} and u_e are the velocities, and m_{ij} and m_e are masses of the ionic and electronic species undertaken in the investigative model, respectively. The above equations represent the particle balance equations for charged plasma species

$$m_{ij}u_{ij}\left(\frac{\partial u_{ij}}{\partial z} + \nu_{\text{cf}}\right) = e\frac{\partial \varphi}{\partial z} \tag{3}$$

$$m_e u_e \nu_{\text{cf}} = e\frac{\partial \varphi}{\partial z}. \tag{4}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 23,2022 at 07:05:40 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Equations (3) and (4) describe the momentum balance equations for charged plasma species, respectively. ν_{cf} is defined as the neutral-ion collision frequency, and ν_{if} is the neutral atom ionization frequency due to electron collision. The sheath potential of the plasma is defined as φ

$$\frac{1}{4\pi e}\frac{\partial^2\varphi}{\partial z^2} = n_e - \sum_{ij}\zeta_{ij}n_{ij}. \tag{5}$$

Equation (5) defines Poisson's equation that defines the sheath potential by considering the ion and electron density, where ζ_{ij} is the ratio of number density for jth an ion to the electron, $\sum_{ij}\zeta_{ij} = 0$, and $0 < \zeta_{ij} < 1$.

The nanoparticles on the substrate are charged due to immersion in the reactive plasma as it contains various charged species (i.e., electrons, ions, etc.). The nanoparticles usually attain a negative charge, and the magnitude of this charge developed depends significantly upon the plasma conditions and particle sizes. Following Fig. 1.

1) Catalyst thin film is predeposited over the substrate and is treated with hydrogen plasma to obtain the fundamental growth point, i.e., metal catalyst nanoparticles over the substrate [34].
2) The hydrocarbon gases decay on the particle's surface due to substrate heating and high power plasma. Some of these adsorbed ions are desorbed from the catalyst's surface due to thermal dehydrogenation and dissociation. The active area of the catalyst nanoparticle generates numerous carbon atoms that diffuse over the surface and the volume of the catalyst [35]–[37].
3) These carbon atoms dissolve into the catalyst until it is supersaturated, after which they start precipitating out of the rear end of the metal nanoparticle and initiate the formation of graphitic layers around the nanoparticle.
4) The growth of CNT and the etching due to enchants species present in the plasma happen simultaneously. The nanoparticle's shape starts to deform due to the stresses exerted by these graphitic layers on the catalyst [38], [39].
5) The CNT growth concludes due to the encapsulation of the catalyst under amorphous carbon layers that covers the nanoparticle's active area [17].
6) The nanotube is aligned vertically simultaneous to the growth due to the electrostatic forces provided by the plasma sheath [10].

A. Pretreatment of Nanofilm and Creation of Nanoparticles

Prior to CNT growth, a catalyst nanofilm is predeposited on the substrate surface. The nanofilm is processed in hydrogen plasma (plasma processing) to convert it into a seed layer of catalyst nanoparticles. The energy balances in plasma processing, i.e., the total power input on the surface of the substrate is the sum of contributions from different energy fluxes per unit area [40], i.e., power interchange of plasma species (neutrals, ions, and electrons), various reactions involve recombination or collection and sputtering of the nanofilm

$$\text{Power Input}=\begin{pmatrix}\mathrm{P}_{\text{neutrals}} + \mathrm{P}_{\text{ions \& electrons}}\\ +\mathrm{P}_{\text{processes}} + \mathrm{P}_{\text{reactions}} + \mathrm{P}_{\text{external}}\end{pmatrix}.$$

Considering this power balance which takes into account the input plasma power of the PECVD reactor and its effect on the power gain and loss over substrate surface and nanofilm which fragments it into nanoparticles, we get

$$\begin{aligned}&\partial_\tau(m_{\mathrm{nf}}C_{\mathrm{nf}}T_S)\\ &= T_S C_{\mathrm{nf}} A_{\mathrm{nf}}\rho_{\mathrm{nf}}\partial_\tau(V_{\mathrm{ct}}n_{\mathrm{ct}})\\ &= T_S C_{\mathrm{nf}} A_{\mathrm{nf}} t_{\mathrm{nf}}\rho_{\mathrm{nf}}\frac{1}{n_{\mathrm{ct}}}\partial_\tau(n_{\mathrm{ct}}) = \left[\xi_p^{\mathrm{nf}}\varepsilon_p^{\mathrm{nf}} + \xi_{\mathrm{ip}}^{\mathrm{nf}}\varepsilon_{\mathrm{ip}}^{\mathrm{nf}} + \xi_e^{\mathrm{nf}}\varepsilon_e^{\mathrm{nf}}\right]\\ &\quad+\left[\frac{3}{2}k_{\mathrm{B}}T_S(1-\gamma_{\mathrm{ip}})\xi_{\mathrm{ip}}^{\mathrm{nf}} + \delta(1-\gamma_e)\left[2k_{\mathrm{B}}T_e - \frac{3}{2}k_{\mathrm{B}}T_S\right]\xi_e^{\mathrm{nf}}\right]\\ &\quad+A_{\mathrm{nf}}\begin{bmatrix}\upsilon_0\sigma_{\mathrm{ads}}U_{\mathrm{B}}n_p(k_{\mathrm{B}}T_p/m_{\mathrm{B}})^{1/2}\\ +y_d E_i(1-\theta_t)n_{\mathrm{ip}}(k_{\mathrm{B}}T_{\mathrm{ip}}/m_{\mathrm{ip}})^{1/2}\end{bmatrix}.\end{aligned} \tag{6}$$

Equation (6) refers to the disintegration of nanofilm into nanoparticles deliberating the dynamics involved at the nanofilm surface. m_{nf} is the nanofilm mass, C_{nf} denotes the nanofilm specific heat, and ρ_{nf} is the nanofilm density. $V_{\mathrm{ct}}(= 2\pi R_{\mathrm{ct}}^3/3)$ is the catalyst volume, where R_{ct} is catalyst radius. A_{nf} and t_{nf} are nanofilm area and thickness, and n_{ct} is density of nanoparticles. y_d is the sputtering yield. U_{B} is the surface binding energy for the catalyst. $\varepsilon_e^{\mathrm{nf}}$, $\varepsilon_{\mathrm{ip}}^{\mathrm{nf}}$, and $\varepsilon_p^{\mathrm{nf}}$ are the mean energies of the electrons, ionic, and neutral species of hydrogen at the metal nanofilm surface given by $\varepsilon_k^{\mathrm{nf}} = [(2 - Z\gamma_k/1 - Z\gamma_k) - Z\gamma_k]k_{\mathrm{B}}T_k$ and $\gamma_k = (E_b/k_{\mathrm{B}}T_k)$. n_{ip}, n_p, and n_e are the number densities of plasma species. Also, ξ_{ik}^{nf}, ξ_k^{nf}, and ξ_e^{nf} are the collection current of the ions, neutrals, and electrons at the nanofilm surface defined as

$$\begin{aligned}\xi_{\mathrm{ip}}^{\mathrm{nf}} &= n_{\mathrm{ip}}(z)A_{\mathrm{nf}}\left(\frac{8k_{\mathrm{B}}T_{\mathrm{ip}}}{\pi m_{\mathrm{ip}}}\right)^{1/2}\exp\left[-\frac{E_b}{k_{\mathrm{B}}T_S}\right]\exp\left[-\frac{eU_S}{k_{\mathrm{B}}T_S}\right]\\ \xi_p^{\mathrm{nf}} &= n_p(z)A_{\mathrm{nf}}\left(\frac{8k_{\mathrm{B}}T_p}{\pi m_p}\right)^{1/2}\\ \xi_e^{\mathrm{nf}} &= n_e(z)A_{\mathrm{nf}}\left(\frac{8k_{\mathrm{B}}T_e}{\pi m_e}\right)^{1/2}\exp\left[\frac{eV_S}{k_{\mathrm{B}}T_e} + \frac{eU_S}{k_{\mathrm{B}}T_S}\right].\end{aligned}$$

The left-hand side of (6) refers to the increased power due to the catalyst nanofilm etching. On the right-hand side of (6), the first term takes into account the change in energy due to the collection of neutral, ionic, and electronic species at nanofilm surface, the second term considers neutral atom formation and loss of electrons due to sticking and collisions, and the last term denotes plasma etching and sputtering of the nanofilm [41].

B. Charge Balance of Carbon Nanotubes

$$\partial_\tau(\mathrm{Z}) = \sum_{A=1}^{8}\begin{pmatrix}\overset{\text{tip}}{\xi}\\ {\scriptstyle iA}\end{pmatrix} + \sum_{B=1}^{2}\begin{pmatrix}\overset{\text{tip}}{\xi} + \overset{\text{cyl}}{\xi}\\ {\scriptstyle iB}\end{pmatrix} - \gamma_e\begin{pmatrix}\overset{\text{tip}}{\xi} + \overset{\text{cyl}}{\xi}\\ {\scriptstyle e}\end{pmatrix}. \tag{7}$$

The time progression of charge over the nanotube surface is represented by (7). Due to positive ions and electronic species deposition, a charge is developed over the nanotube's tip and cylindrical surface. The above equation considers Z as the charge generated on the CNT surface and γ_e is the sticking coefficient of electrons. The first and second term

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 23,2022 at 07:05:40 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE II
INTERPRETATION OF FUNCTIONS/PARAMETERS TERMED IN (8)–(11)

Terms	Description
V_S	CNT tubular surface potential
U_S, T_S	Substrate bias and temperature
E_b	Energy necessary for bulk diffusion
$n_{ij}(z)=n_{ij0}\left(1-\frac{2e\varphi(z)}{m_{ij}u_{ij0}^2}\right)^{-1/2}$	Plasma sheath ion density
$n_e(z)=n_{e0}\exp\left(\frac{\lvert e\rvert\varphi(z)}{k_BT_e}\right)$	Plasma sheath electron density
$\varphi(z)=\varphi_0\exp\left(-\frac{z}{\lambda_D}\right)$	Electrostatic sheath potential

TABLE III
INTERPRETATION OF FUNCTIONS/PARAMETERS TERMED IN (12)

Terms	Description
β_j	Ionization coefficient of neutrals
$\alpha_j(T_e)=\alpha_{j0}(300/T_e)^k$	Recombination coefficient for ions and electrons where $k=-1.2$ [41]
$\underset{wall}{\overset{e}{\mathrm{K}}}=\left(\frac{\gamma_e u_e A_{chm}}{4V}\right)$	Loss rate of electrons over the volume on the discharge wall[42]
A_{chm}, V	Cylindrical plasma chamber's surface area and volume (L=23cm and D=32cm)[42]

of (7) can be expanded in the form of (8) and (9), respectively, (where $j = A,\ B$)

$$\overset{\text{tip}}{\xi}_{ij} = n_{ij}(z)\pi r^2\left(\frac{8k_{\mathrm{B}}T_{ij}}{\pi m_{ij}}\right)^{1/2}\left[1-\mathrm{Z}\alpha_{ij}\right]\exp\left[-\frac{E_b}{k_{\mathrm{B}}T_S}\right]\times\exp\left[-\frac{eU_S}{k_{\mathrm{B}}T_S}\right] \tag{8}$$

$$\overset{\text{cyl}}{\xi}_{ij} = n_{ij}(z)rl\left(\frac{2\pi k_{\mathrm{B}}T_{ij}}{m_{ij}}\right)^{1/2}\exp\left[-\frac{E_b}{k_{\mathrm{B}}T_S}\right]\exp\left[-\frac{eU_S}{k_{\mathrm{B}}T_S}\right]\times\left\{\frac{2}{\sqrt{\pi}}\left(\frac{eV_S}{k_{\mathrm{B}}T_{ij}}\right)^{1/2}+\exp\left[\frac{eV_S}{k_{\mathrm{B}}T_{ij}}\right]\mathrm{erfc}\left[\left(\frac{eV_S}{k_{\mathrm{B}}T_{ij}}\right)^{1/2}\right]\right\}. \tag{9}$$

These equations define the amount of charge collected due to ionic species at the nanotube's tip and cylindrical regions, respectively. The third term in (7) can be expanded in the form of the following equations, respectively:

$$\overset{\text{tip}}{\xi}_{e} = n_e(z)\pi r^2\left(\frac{8k_{\mathrm{B}}T_e}{\pi m_e}\right)^{1/2}\exp\left[\mathrm{Z}\alpha_e+\frac{eU_S}{k_{\mathrm{B}}T_S}\right] \tag{10}$$

$$\overset{\text{cyl}}{\xi}_{e} = n_e(z)rl\left(\frac{2\pi k_{\mathrm{B}}T_S}{m_e}\right)^{1/2}\exp\left[\frac{eV_S}{k_{\mathrm{B}}T_e}+\frac{eU_S}{k_{\mathrm{B}}T_S}\right]. \tag{11}$$

These equations define the charge collected due to electrons at the nanotube's tip and cylindrical regions, respectively. $\alpha_e(=e^2/rk_{\mathrm{B}}T_e)$ and $\alpha_{ij}(=e^2/rk_{\mathrm{B}}T_{ij})$. The other terms in the above equations are termed in Table II.

C. Time-Evolution of Plasma Electron Density

$$\partial_\tau(n_e) = \left[\sum_{A=1}^{8}\beta_{\mathrm{A}}n_{\mathrm{A}}+\sum_{B=1}^{2}\beta_{\mathrm{B}}n_{\mathrm{B}}\right]-\left[\sum_{A=1}^{8}\alpha_{\mathrm{A}}n_en_{iA}+\sum_{B=1}^{2}\alpha_{\mathrm{B}}n_en_{iB}\right]\times\frac{n_{\mathrm{CNT}}}{\lambda_s}\left(\xi_e^{\text{tip}}+\xi_e^{\text{cyl}}\right)\gamma_e-n_e\mathrm{K}_{\text{wall}}^{e}. \tag{12}$$

The time progression of bulk plasma electron density is defined by (12). The first term in the above equation defines the gain in density owing to the dissociative-ionization of neutrals $j = (A,\ B)$; the second term calculates loss due to the ion-electron recombination. The decrease in electron density due to the accumulation of electrons at the nanotube surface is interpreted in the third term, and at last, the fourth term is the electron discharge loss to the reactive plasma chamber walls. Other terms in (12) are mentioned in Table III.

D. Time-Evolution of Plasma Ion Density

$$\partial_\tau(n_{iA}) = \overbrace{\left\{\beta_{\mathrm{A}}n_{\mathrm{A}}+\left(\sum_{\ell AB}k_\ell n_{\mathrm{A}}n_{iB}-\sum_{qBA}k_q n_{\mathrm{B}}n_{iA}\right)+\left(\frac{P}{E_{\mathrm{CH}}^{\text{diss}}V}\right)\right\}}^{\text{Gain}} - \overbrace{\left\{\alpha_{\mathrm{A}}n_en_{iA}+\frac{n_{\mathrm{CNT}}}{\lambda_s}\left(\xi_{iA}^{\text{tip}}+\xi_{iA}^{\text{cyl}}\right)+n_{iA}\mathrm{K}_{\text{wall}}^{iA}\right\}}^{\text{Loss}} \tag{13}$$

$$\partial_\tau(n_{iB}) = \overbrace{\left\{\beta_{\mathrm{B}}n_{\mathrm{B}}+\left(\sum_{qBA}k_q n_{\mathrm{B}}n_{iA}-\sum_{\ell AB}k_\ell n_{\mathrm{A}}n_{iB}\right)+\left(\frac{P}{E_{H}^{\text{diss}}V}\right)\right\}}^{\text{Gain}} - \overbrace{\left\{\alpha_{\mathrm{B}}n_en_{iB}+\frac{n_{\mathrm{CNT}}}{\lambda_s}\left(\xi_{iB}^{\text{tip}}+\xi_{iB}^{\text{cyl}}\right)+n_{iB}\mathrm{K}_{\text{wall}}^{iB}\right\}}^{\text{Loss}}. \tag{14}$$

Equations (13) and (14) denote the time progression of bulk plasma ion number density of A- and B-type ions. In the above equations, the first part refers to the rise in the ion density owing to dissociative ionization of neutrals $j = (A,\ B)$, changes due to ion-neutral atom plasma reactions, and ion dissociation by plasma power, respectively. The second part of the equations refers to the fall in ion number density as the ions and electrons recombine, accumulation of ions on the nanotube surface, and discharge loss of ions to the reactive plasma chamber walls. $\mathrm{K}_{\text{wall}}^{ij}$ is the ion discharge loss to the chamber [42], k_q, k_l are the neutral-ion reaction coefficients [43], $\delta E_{\mathrm{CH}}^{\text{diss}}(\approx 2.1$ eV), $\delta E_{H}^{\text{diss}}(\approx 2.1$ eV) are the hydrocarbon and hydrogen dissociation energies [33], and P is applied plasma power.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 23,2022 at 07:05:40 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

E. Time-Evolution of Plasma Neutral Atom Density

Equations (15) and (16) denote the time progression of bulk plasma neutrals of A and B types

$$\partial_\tau(n_{\mathrm{A}}) = \overbrace{\left\{\alpha_{\mathrm{A}} n_e n_{iA} + \frac{n_{\mathrm{CNT}}}{\lambda_s}\left(\xi_{iA}^{\mathrm{tip}} + \xi_{iA}^{\mathrm{cyl}}\right)(1-\gamma_{iA}) + J_{\mathrm{A}}^{\mathrm{desp}} + \left(\mathrm{Q}_{\mathrm{A}}^{\mathrm{IF}} - \mathrm{Q}_{\mathrm{A}}^{\mathrm{OF}}\right) + \overbrace{\left(\sum_{qAB} k_q n_{\mathrm{B}} n_{iA} - \sum_{lBA} k_l n_{\mathrm{A}} n_{iB}\right)}^{\text{Gain}}\right\}}^{\text{Gain}} - \overbrace{\left\{\beta_{\mathrm{A}} n_{\mathrm{A}} + \frac{n_{\mathrm{CNT}}}{\lambda_s}\left(\xi_{\mathrm{A}}^{\mathrm{tip}} + \xi_{\mathrm{A}}^{\mathrm{cyl}}\right)\gamma_{\mathrm{A}}\right.}^{\text{Loss}} \overbrace{\left. + J_{\mathrm{A}}^{\mathrm{adsp}} + n_{\mathrm{A}} \mathrm{K}_{\mathrm{wall}}^{\mathrm{A}}\right\}}^{\text{Loss}} \quad (15)$$

$$\partial_\tau(n_{\mathrm{B}}) = \overbrace{\left\{\alpha_{\mathrm{B}} n_e n_{iB} + \frac{n_{\mathrm{CNT}}}{\lambda_s}\left(\xi_{iB}^{\mathrm{tip}} + \xi_{iB}^{\mathrm{cyl}}\right)(1-\gamma_{iB}) + J_{\mathrm{B}}^{\mathrm{desp}} + \left(\mathrm{Q}_{\mathrm{B}}^{\mathrm{IF}} - \mathrm{Q}_{\mathrm{B}}^{\mathrm{OF}}\right) + \overbrace{\left(\sum_{\ell AB} k_\ell n_{\mathrm{A}} n_{iB} - \sum_{qBA} k_q n_{\mathrm{B}} n_{iA}\right)}^{\text{Gain}}\right\}}^{\text{Gain}} - \overbrace{\left\{\beta_{\mathrm{B}} n_{\mathrm{B}} + \frac{n_{\mathrm{CNT}}}{\lambda_s}\left(\xi_{iB}^{\mathrm{tip}} + \xi_{iB}^{\mathrm{cyl}}\right)\gamma_{\mathrm{B}}\right.}^{\text{Loss}} \overbrace{\left. + J_{\mathrm{B}}^{\mathrm{adsp}} + \mathrm{K}_{\mathrm{wall}}^{\mathrm{B}} n_{\mathrm{B}}\right\}}^{\text{Loss}} + J_{\mathrm{thd}}. \quad (16)$$

In the above equations, the first part denotes the gain in the number density of neutrals due to ion-electron recombination, ion neutralization after their accumulation at the nanotube surface, ion desorption from the catalyst surface, changes due to neutral species inflow and outflow in the chamber, and changes owing to neutral atom-ion reactions deliberated in the study [43], respectively. The second part of the equations refers to the loss in density due to the neutral atom dissociative ionization, loss owing to the collection of neutrals on nanotube surface, neutral atom adsorption, and neutral atom discharge loss to the reactive plasma chamber walls, respectively. In (16), the last term determines the rise in hydrogen neutral density due to the thermal dehydrogenation process [54]. Other terms described in (15) and (16) are designated in Table IV.

TABLE IV
INTERPRETATION OF FUNCTIONS/PARAMETERS TERMED IN (15) AND (16)

Terms	Description
$\mathrm{Q}_j^{IF} = \left(\frac{4.4\times10^{17}}{V} J_j\right)$	Inflow of neutrals from the reactive plasma chamber[42]
$\mathrm{Q}_j^{OF} = \left(\frac{\upsilon_{pump}}{V} n_j\right)$	Outflow of neutrals from the reactive plasma chamber[42]
J_j^{adsp}, J_j^{desp}	adsorption and desorption fluxes from the catalyst-substrate surface
J_{thd}	thermal dehydrogenation flux[44]
$\underset{wall}{\overset{j}{\mathrm{K}}} = \sigma_j \frac{D_j}{\Lambda^2}$	neutral atom discharge loss with Λ as cylindrical chamber effective diffusion length[32], [45]
σ_j, D_j	wall reaction and diffusion coefficient respective species[32], [45]

F. Evolution of C and H Atoms Over the Catalyst

The C species density n_c over the catalyst is signified by (17), as shown at the bottom of the page. The first part of the equation construes the rise in carbon density as the acetylene neutrals get adsorbed over the catalyst, thermal dissociation, ion-induced hydrocarbon dissociation, positive ion decomposition, direct carbon influx, and hydrocarbon and hydrogen ion-ion interaction, respectively. The second part of the equation construes the fall in the density of carbon owing to the hydrocarbon desorption from catalyst surface, hydrocarbon-hydrogen interaction, C species evaporation, and interaction between catalyst surface generated C and H species. Finally, the continuous hydrocarbon species dissociation over the active area of catalyst nanoparticles creates an excessive amount of carbon species, leading to an amorphous carbon layer formation that decreases the active area over the catalyst and concludes nanotube growth. The last term of (17) indicates the catalyst-poisoning rate that depends on the density of C atoms generated throughout the growth over the catalyst surface

The hydrogen species number density n_H over the catalyst is signified by (18), as shown at the bottom of the next page. The first part of the equation construes the rise in density owing to the hydrogen neutrals adsorption over catalyst surface, thermal dissociation, ion-induced dissociation

$$\partial_\tau(n_{\mathrm{C}}) = \left[\overbrace{\left\{\sum_{A=1}^{8}(1-\theta_t) J_{\mathrm{A}} + \sum_{A=1}^{8} n_{\mathrm{A}}^{\mathrm{surf}} \upsilon \exp\left(-\frac{\delta E_{\mathrm{CH}}^{\mathrm{diss}}}{k_{\mathrm{B}} T_S}\right) + \sum_{iA=1}^{8}\left(\sum_{A=1}^{8} \frac{n_{\mathrm{A}}^{\mathrm{surf}} y_d}{\upsilon_0}\right) J_{iA}\right.}^{\text{Gain}} \overbrace{\left. + \sum_{iA=1}^{8} J_{iA} + J_{\mathrm{C}} + \sum_{iB=1}^{2}\left(\sum_{iA=1}^{8} \frac{J_{iA}\sigma_{\mathrm{ads}}}{\upsilon}\right) J_{iB}\right\}}^{\text{Gain}} - \overbrace{\left\{\sum_{A=1}^{8} n_{\mathrm{A}} \upsilon \exp\left(-\frac{\delta E_{dCH}}{k_{\mathrm{B}} T_S}\right)\right.}^{\text{Loss}} \overbrace{\left. + \sum_{B=1}^{2}\left(\sum_{A=1}^{8} n_{\mathrm{A}}^{\mathrm{surf}} \sigma_{\mathrm{ads}}\right) J_{\mathrm{B}} + n_{\mathrm{C}} \upsilon \exp\left(-\frac{\delta E_{\mathrm{ev}}}{k_{\mathrm{B}} T_S}\right) + n_{\mathrm{C}} \sigma_{\mathrm{ads}} \upsilon n_H\right\}}^{\text{Loss}}\right] \times d_\tau(\Omega_{\mathrm{C}}) \quad (17)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 23,2022 at 07:05:40 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE V
INTERPRETATION OF FUNCTIONS/PARAMETERS TERMED IN (17) AND (18)

Terms	Description
$\upsilon_0(\approx 10^{15}\ cm^{-2})$	Adsorption sites per unit area[44]
$\upsilon(\approx 10^{13}\ Hz)$	Thermal vibration frequency[46]
σ_{ads} $\left(=6.8\times10^{-16}cm^2\right)$	Interaction cross-section amongst various species[44]
$n_j^{surf}\left(=\theta_j\upsilon_0\right)$	Surface number density of neutrals[12], [44]
$\delta E_{CH}^{diss}\left(=2.1\ eV\right)$, $\delta E_{ev}(=1.8\ eV)$, $\delta E_{dCH}\left(=1.8\ eV\right)$, $\delta E_{dH}\left(=1.8\ eV\right)$	Hydrocarbon dissociation, carbon evaporation, hydrocarbon and hydrogen desorption energies[33]
$d_\tau(\Omega_C)$ $=\left(1-4n_C\pi r^2\right)$	Poisoning rate of catalyst governed by n_C that is generated throughout the growth process[10]

TABLE VI
INTERPRETATION OF FUNCTIONS/PARAMETERS TERMED IN (19)

Terms	Description
$D_{Surf}=a_0^2\upsilon\exp\left(-\frac{E_{Surf}}{k_BT_S}\right)$	Surface diffusion coefficient of carbon atoms[44]
$D_{Bulk}=\frac{\upsilon R_{ct}}{\pi}\exp\left(-\frac{E_{Bulk}}{k_BT_S}\right)$	Bulk diffusion coefficient of carbon atoms[44]
$D_m=D_{m0}\exp\left(-\frac{\delta E_{sd}}{k_BT_S}\right)$	Self-diffusion coefficient for metal atoms[44]
$\kappa=a_0\upsilon\exp\left(-\frac{\delta E_{inc}}{k_BT_S}\right)$	Carbon atom incorporation rate into the CNT wall[12], [44]
$E_{Surf}\left(=0.3\ eV\right)$, $E_{Bulk}\left(=1.6\ eV\right)$	Carbon atom energy barrier for surface and bulk diffusion on the catalyst surface[44]
δE_{inc} $\left(=\delta E_{sd}=0.8\ eV\right)$	Carbon atom energy barrier to diffuse along nanotube-catalyst boundary[12]

of hydrocarbons, and inflow of hydrogen around the catalyst. The second part construes the fall in density owing to its desorption from the catalyst surface, the interaction between the hydrogen adsorbed over the catalyst and neutral atom, interaction between the adsorbed hydrogen over the catalyst and hydrogen ions, and C and H atom interaction. Other terms stated in (17) and (18) are termed in Table V.

G. Catalytic Growth of CNT in Plasma

$$\partial_\tau\left[\pi\left(R_{\rm ct}^2-r^2\right)l\right] = \left[\left\{D_{\rm Surf}+D_{\rm Bulk}+2\kappa\,\upsilon\,R_{\rm ct}^2\right\}\times\frac{P(8\pi\,R_{\rm ct})n_{\rm C}}{\upsilon\rho_{\rm CNT}\sum_{A=1}^{8}\xi_{iA}} +\left\{\frac{D_m}{4\pi\,R_{\rm ct}^2}+n_H\left(4\pi\,R_{\rm ct}^2\right)\sum_{B=1}^{2}\xi_{iB}\right\}\right]\times\frac{m_{\rm ct}}{\rho_{\rm ct}}. \tag{19}$$

The time progression of CNT growth is estimated by (19). The nanotube growth is assessed assuming a change in CNT volume (dependent on its radius and height). Several growth processes considered in the above equation are explained in Table VI. The first part of the equation considers the surface diffusion over the catalyst surface, bulk diffusion of carbon through the catalyst volume, and the carbon atom precipitation from the rear end of the catalyst, respectively, forming a tubular graphitic structure of the nanotube. The supremacy of one diffusion process over another mainly governs the synthesis of single walled CNTs (SWCNTs) or multi walled CNTs (MWCNTs), i.e., if bulk diffusion dominates over surface diffusion, MWCNTs are formed or else SWCNTs [47]. The second part of the equation contemplates the self-diffusion of catalyst as the CNT growth proceeds; and the amorphous carbon atom etching attached at the extremity that hinders nanotube growth. Finally, the growth and alignment of CNT proceed simultaneously due to the electrostatic force applied to the growing nanotube structure [10]. The change in the CNT diameter also changes the number of the graphitic layers or the CNT walls. For simplification, CNT walls can be estimated by the difference between the inner and outer diameter of the CNT, assuming these CNTs are perfectly aligned concentric graphitic tubes, i.e., $(D_o-D_i)\ \sim$ number of CNT walls, where $D_o(\sim\ D_{\rm ct})$ is the nanotube outer diameter which assumed to be comparable with catalyst diameter, and D_i is the nanotube inner diameter.

III. RESULTS AND DISCUSSION

The results for the current study helps in understanding the catalyzed growth of CNT in acetylene-hydrogen plasma. To study the plasma pretreatment, we have considered a

$$\partial_\tau(n_H)=\overbrace{\left\{\sum_{B=1}^{2}J_{\rm B}(1-\theta_t)+\sum_{B=1}^{2}n_{\rm B}^{\rm surf}\upsilon\exp\left(-\frac{\delta E_{\rm CH}^{\rm diss}}{k_{\rm B}T_S}\right)+\sum_{iA=1}^{8}\left(\sum_{A=1}^{8}\frac{n_{\rm A}^{\rm surf}y_d}{\upsilon_0}\right)J_{iA}\right.}^{\text{Gain}}\overbrace{\left.+\sum_{iB=1}^{2}J_{iB}\right\}}^{\text{Gain}}$$
$$-\overbrace{\left\{\sum_{B=1}^{2}n_{\rm B}^{\rm surf}\upsilon\exp\left(-\frac{\delta E_{dH}}{k_{\rm B}T_S}\right)+\sum_{B=1}^{2}n_{\rm B}^{\rm surf}\sigma_{\rm ads}J_{\rm B}\right.}^{\text{Loss}}\overbrace{\left.+\sum_{B=1}^{2}n_{\rm B}^{\rm surf}\sigma_{\rm ads}J_{iB}+n_H\sigma_{\rm ads}\upsilon n_{\rm C}\right\}}^{\text{Loss}} \tag{18}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 23,2022 at 07:05:40 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE VII
INITIAL CONFIGURATION OF VARIOUS PARAMETERS IN THE PRESENT COMPUTATIONAL STUDY

Terms	Initial value
T_e (electron temperature)	2.0 eV
$T_{ij} = T_n$ (ion and neutral temperature)	0.20 eV
$T_s = T_{ct}$ (substrate and catalyst temperature)	350 $^{\circ}C$
Plasma power	300 W
Working pressure	50 $mTorr$
$\gamma_e, \gamma_{ij}, \gamma_j$ (sticking coefficient)	1
U_S (substrate bias)	-500 V
θ_t (surface coverage)	0.01
$J_{C_2H_2}$ (C_2H_2 gas flow)	50 $sccm$
J_{H_2} (H_2 gas flow rate)	200 $sccm$
A_{nf} (catalyst nanofilm area)	1 cm^2

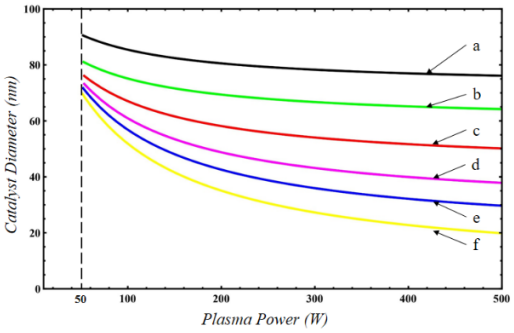


Fig. 2. Variation of catalyst diameter with plasma power at different process pressures where a–f correspond to 10, 20, 30, 40, 50, and 60 mTorr, respectively.

Ni catalyst nanofilm of thickness 15 nm, and the effect of plasma power, pressure, and gas flow on the nanoparticles is investigated. Some of the initial plasma parameters used in the study are mentioned below. For studying the different catalyst effects, Ni, Fe, and Co catalyst films are considered. The first-order time-differential equations mentioned in the previous section are solved for CNT growth considering the experimentally obtained primary conditions and glow-discharge plasma parameters assumed from [48]–[51], and some are mentioned in Table VII.

The variation of catalyst diameter with the plasma power is depicted in Fig. 2. For constant plasma pressure, increase in plasma power during plasma pretreatment of nanofilm decreases the nanoparticle diameter. For high plasma power, the gaseous species' energy increases and bombarding effect on the nanofilm increases, leading to smaller nanoparticles. The effect of plasma processing pressure on the catalyst diameter is observed in Fig. 3. For fixed plasma power, increase in the plasma pressure during pretreatment decreases the nanoparticle diameter. This is attributed to the gas density increase for high pressure that decreases the mean free path of particles. This increases the collisional rate of gaseous species,

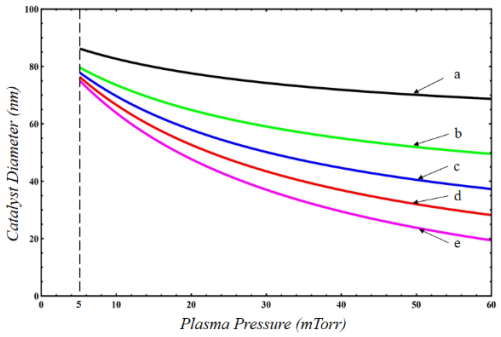


Fig. 3. Variation of catalyst diameter with plasma pressure at different process powers where a–e correspond to 100, 200, 300, 400, and 500 W, respectively.

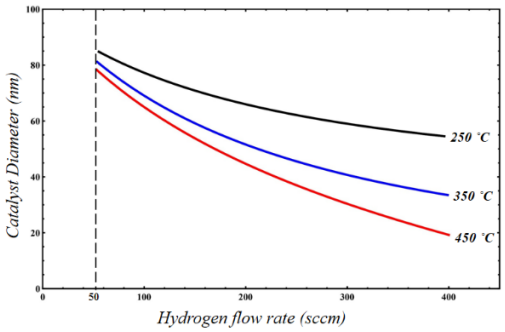


Fig. 4. Variation of catalyst diameter with hydrogen gas flow rate at different substrate temperatures of 250 °C, 350 °C, and 450 °C, respectively.

increasing the gas flux toward the nanofilm, and decreasing the nanoparticle size. The result agrees with the experimental observations of Chang *et al.* [19], [26] whose scanning electron microscopy (SEM) images of Ni nanofilm pretreated with hydrogen show similar trend of decreasing nanoparticle size with plasma power and pressure. The effect of changing hydrogen flow rate during plasma pretreatment is observed in Fig. 4. On increasing the hydrogen flow rate during pretreatment, the flux of plasma species toward the nanofilm increases breaking it into smaller nanoparticles. In addition, it is also evident that smaller nanoparticles are obtained for higher pretreatment temperature as the catalyst can move conveniently than at lower pretreatment temperature. The SEM and transmission electron microscopy (TEM) results from the experimental findings of Wang *et al.* [29] and Jian *et al.* [31], [52] also suggest that hydrogen pretreatment of nanofilm produces uniform and smaller particles for higher gas flow rate. The growth time variation of the catalyst particle density over the substrate for Ni catalyst for varying nanofilm thickness is represented in Fig. 5. It can be observed that the catalyst number density decreases on increasing the nanofilm thickness. This is attributed to the agglomeration of the larger catalyst particles on the substrate for a thicker nanofilm, decreasing their density over the substrate. Moreover, this number density increases with time as the hydrogen plasma species are rigorously and incessantly bombarded on the nanofilm until it is wholly converted into nanoparticles and attains saturation. It can be concluded that on increasing the film thickness, the catalyst obtained has increased diameter. Chhowalla *et al.* [6] and

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 23,2022 at 07:05:40 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

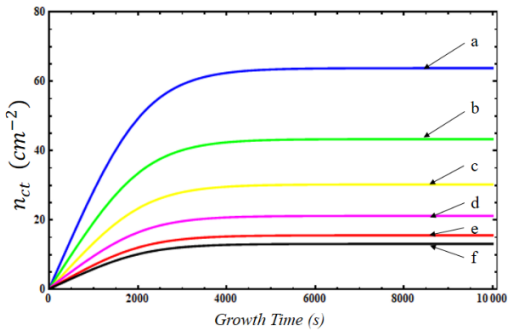


Fig. 5. Time variation of catalyst number density for different nanofilm thickness for Ni catalyst where a–f correspond to 2, 4, 6, 8, 10, and 15 nm, respectively.

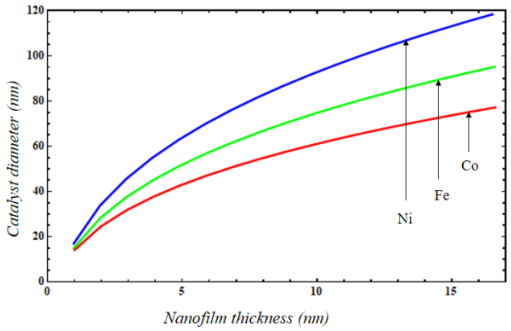


Fig. 6. Variation of catalyst diameter with nanofilm thickness.

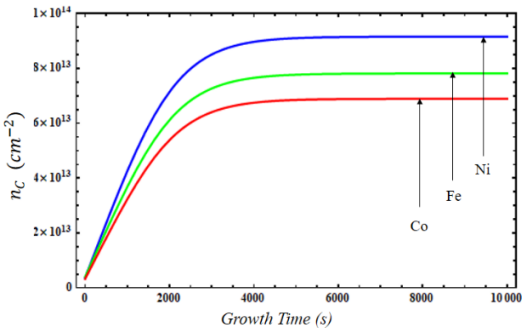


Fig. 7. Rate of creation of *C* species on the catalyst.

Bower *et al.* [53] showed similar dependency of nanofilm thickness on catalyst size experimentally through their SEM and TEM results for Ni and Co, respectively. The variation of catalyst diameter with nanofilm thickness for Ni, Fe, and Co is represented in Fig. 6. It can be concluded that on increasing the film thickness, the catalyst nanoparticle obtained has increased diameter. It is also observed that Ni catalysts are bigger in diameter than Fe and Co catalysts. Huang *et al.* [24] in their experimental study showed the SEM and TEM images of plasma-assisted CNT growth for Ni, Fe, and Co catalysts and observed that for the same nanofilm thickness, the grain size of Ni is more significant than that of Fe and Co and, hence, observed that Ni-CNT had a larger diameter. The rate of carbon species generation over the nanoparticle for the different catalysts is denoted in Fig. 7. For a large catalyst, a larger active region over the catalyst surface allows more hydrocarbon species to dissociate

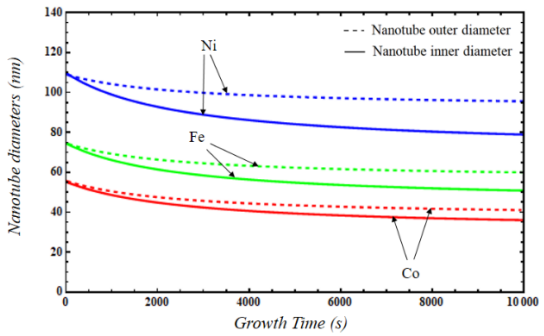


Fig. 8. Temporal variation of CNT inner and outer diameters.

into carbon species hence, creating more carbon species over the catalyst surface. From the observations of Fig. 6, it is comprehended that the carbon species generated over Ni is higher than Fe and Co for the same nanofilm thickness. The temporal variation of CNT inner and outer diameter is depicted in Fig. 8. During the growth process, the carbon atoms generated over the nanoparticle diffuse out at its rear end, forming graphitic tubular layers and causing the nanoparticle to lift off. These layers exert stresses on the nanoparticle leading to its shape change. These walls produce compressive stress at the rear end of the nanoparticle surface. Hence, the inner diameter of the growing CNT is reduced due to these stresses and changes the shape of the catalyst during CNT growth that gives rise to diameter reduction which is observed experimentally through SEM, high-resolution TEM (HRTEM), and environmental TEM (ETEM) results of various groups [6], [37], [54]. The outer CNT diameter is assumed comparable with the catalyst diameter at the top of the catalyst surface; hence not much reduction is observed. Jung and Lee [55] studied the PECVD growth of CNT over Ni catalyst film and suggested a lift-off mechanism. They also observed a more significant reduction in CNT diameter for an increasing catalyst film thickness. The nanoparticle surface produces carbon species and this region gets encapsulated once the catalyst gets saturated as these abundant carbon atoms form an amorphous carbon layer over the catalyst surface. The result shows that for a larger catalyst diameter, the reduction in the inner diameter is more. In addition, CNT with Ni catalyst has larger CNT diameters than Fe and Co. Experimentally, the field Emission SEM (FESEM) and HRTEM results of Hoyos-Palacio *et al.* [23] and Huang *et al.* [24] comply with our observations. The time variation in CNT length for different hydrocarbon densities is observed in Fig. 9(i)–(iii) for Ni, Fe, and Co, respectively. It can be seen that greater acetylene density leads to greater CNT length. When the acetylene density increases, more plasma species are present for dissociation at the catalyst surface that generate more carbon species leading to higher diffusion and precipitation on the catalyst surface. This rise in the CNT growth processes leads to a rise in the CNT length. The time variation of CNT length for the different catalysts is represented in Fig. 10. For a catalyst of smaller size, the carbon species that precipitate out at the rear end of the catalyst cannot circumvent the etching of carbon species by the hydrogen species. Hence, the CNT

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 23,2022 at 07:05:40 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

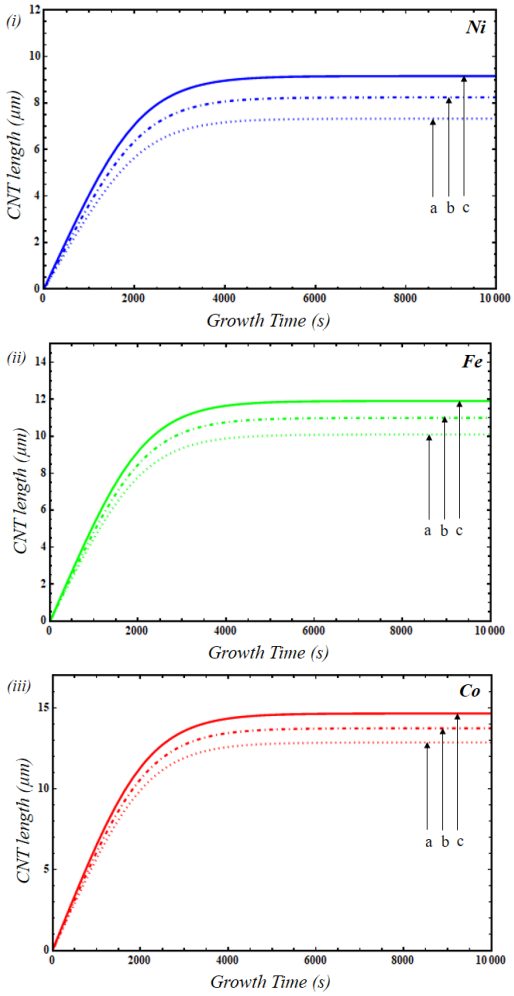


Fig. 9. Temporal variation of CNT length for different hydrocarbon densities (where a, b, and c correspond to 10^{12}, 10^{13}, and 10^{14} cm^{-3}, respectively) for (i) Ni, (ii) Fe, and (iii) Co.

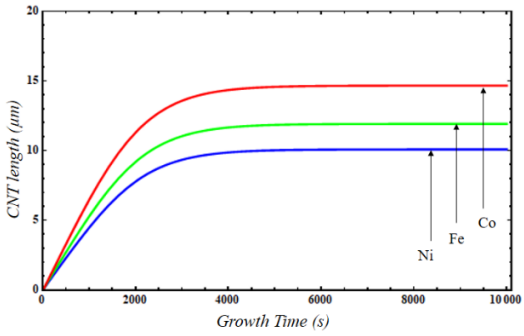


Fig. 10. Growth time variation of CNT length.

starts growing in such a way that it promotes vertical growth by precipitating more at the catalyst rear end but a larger catalyst particle would cover CNT walls growing underneath from hydrogen etching; hence, the number of layers increases. Chhowalla *et al.* [6] and Hofmann *et al.* [56] through their SEM results and data analysis showed this inverse proportionality of CNT length with diameter. Hence, CNT length for Co catalyst is greater than Ni and Fe. The dependency of number

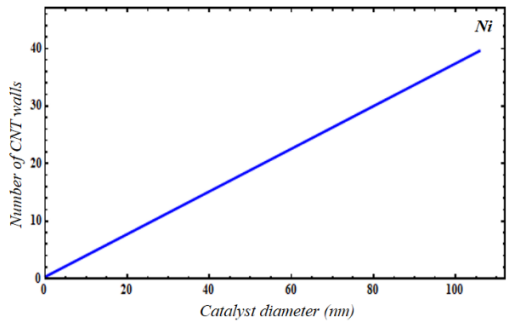


Fig. 11. Dependence of the number of CNT walls on the catalyst diameter for Ni.

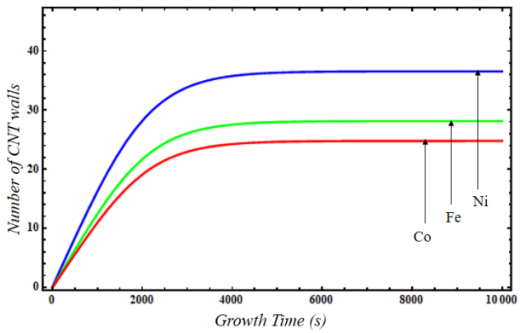


Fig. 12. Time evolution of the number of CNT walls.

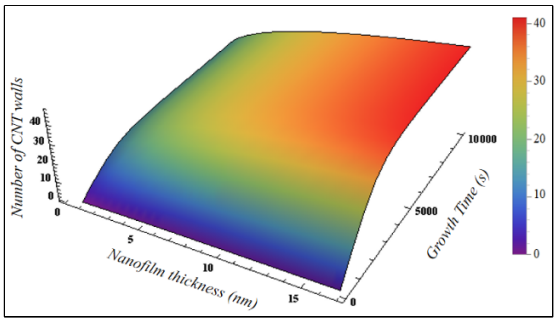


Fig. 13. Variation of CNT walls with the catalyst thickness and growth time for Ni.

of walls on the Ni catalyst diameter is shown in Fig. 11. As catalyst diameter increases, amount of carbon species generated on the catalyst increases that diffuse and precipitate to form more graphitic tubular walls at the bottom of the catalyst. Hence, CNT with larger diameter would have more number of walls and is experimentally confirmed by Hoyos-Palacio *et al.* [23] and Huang *et al.* [24]. Similar dependency was observed for Fe and Co catalysts. Fig. 12 shows the time variation of the number of CNT walls for different catalysts. As CNT grows, the number of CNT walls also increases, and when the growth terminates, these walls attain a saturation. Ni having the bigger nanoparticle size has higher number of CNT walls compared with Fe and Co. The relationship of the number of CNT walls with the catalyst diameter and nanofilm thickness and its variation with growth time

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 23,2022 at 07:05:40 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

has been verified by the data analysis of Xie *et al.* [57], Ma *et al*. [58], and Prawoto *et al.* [59]. The temporal variation of CNT walls along with catalyst thickness for Ni catalyst is depicted in Fig. 13. The result shows that CNT walls increase with time for varying nanofilm thickness. This is attributed to the increase in catalyst size with increasing nanofilm thickness (cf. Fig. 6). This increase in catalyst size increases the number of CNT walls (cf. Fig. 11). A similar outcome is observed using Fe and Co.

IV. Conclusion

A computational model was developed and studied for the plasma pretreatment of nanofilms, and the effect of process parameters, i.e., power, pressure, gas flow, and plasma temperature, was deliberated. The model also accomplishes the catalyzed growth of CNTs in a reactive plasma considering different catalysts, namely, iron (Fe), nickel (Ni), and cobalt (Co). The effect of the nature and thickness of the catalyst and its nanofilm on CNT growth parameters, such as nanotube diameter, nanotube length, number of CNT walls, and so on, was deliberated in the current study. The relation between the catalyst diameter and nanotube walls was established. The time evolution of various parameters (CNT diameter, length, and number of CNT walls) for different catalysts was also studied. The study would also be helpful in producing various other nanostructures and altering their properties to serve the necessary objective for their use in various applications.

References

[1] S. Iijima, "Helical microtubules of graphitic carbon," *Nature*, vol. 354, pp. 56–58, Nov. 1991.

[2] Z. Shi *et al.*, "Mass-production of single-wall carbon nanotubes by arc discharge method," *Carbon*, vol. 37, no. 9, pp. 1449–1453, 1999, doi: 10.1016/S0008-6223(99)00007-X.

[3] R. L. Vander Wal, G. M. Berger, and T. M. Ticich, "Carbon nanotube synthesis in a flame using laser ablation for *in situ* catalyst generation," *Appl. Phys. A: Mater. Sci. Process.*, vol. 77, no. 7, pp. 885–889, Dec. 2003, doi: 10.1007/s00339-003-2196-3.

[4] M. Jung, "Growth of carbon nanotubes by chemical vapor deposition," *Diamond Related Mater.*, vol. 10, nos. 3–7, pp. 1235–1240, 2001, doi: 10.1016/S0925-9635(00)00446-5.

[5] M. Tanemura *et al.*, "Growth of aligned carbon nanotubes by plasma-enhanced chemical vapor deposition: Optimization of growth parameters," *J. Appl. Phys.*, vol. 90, no. 3, pp. 1529–1533, Aug. 2001, doi: 10.1063/1.1382848.

[6] M. Chhowalla *et al.*, "Growth process conditions of vertically aligned carbon nanotubes using plasma enhanced chemical vapor deposition," *J. Appl. Phys.*, vol. 90, no. 10, pp. 5308–5317, Nov. 2001, doi: 10.1063/1.1410322.

[7] H. Murakami, M. Hirakawa, C. Tanaka, and H. Yamakawa, "Field emission from well-aligned, patterned, carbon nanotube emitters," *Appl. Phys. Lett.*, vol. 76, no. 13, pp. 1776–1778, Mar. 2000, doi: 10.1063/1.126164.

[8] J. M. Bonard *et al.*, "Tuning the field emission properties of patterned carbon nanotube films," *Adv. Mater.*, vol. 13, no. 3, pp. 184–188, Feb. 2001, doi: 10.1002/1521-4095(200102)13:3%3C184::AID-ADMA184%3E3.0.CO;2-I.

[9] G. S. Bocharov and A. V. Eletskii, "Theory of carbon nanotube (CNT)-based electron field emitters," *Nanomaterials*, vol. 3, no. 3, pp. 393–442, 2013, doi: 10.3390/nano3030393.

[10] U. Sharma and S. C. Sharma, "A parametric study to unravel the alignment mechanism of carbon nanotubes during its plasma-assisted growth," *Phys. Plasmas*, vol. 25, no. 10, Oct. 2018, Art. no. 103509, doi: 10.1063/1.5049866.

[11] C. Ducati, I. Alexandrou, M. Chhowalla, J. Robertson, and G. A. J. Amaratunga, "The role of the catalytic particle in the growth of carbon nanotubes by plasma enhanced chemical vapor deposition," *J. Appl. Phys.*, vol. 95, no. 11, pp. 6387–6391, Jun. 2004, doi: 10.1063/1.1728293.

[12] I. Denysenko and N. A. Azarenkov, "Formation of vertically aligned carbon nanostructures in plasmas: Numerical modelling of growth and energy exchange," *J. Phys. D: Appl. Phys.*, vol. 44, no. 17, May 2011, Art. no. 174031, doi: 10.1088/0022-3727/44/17/174031.

[13] X. Wang, A. Dong, Y. Hu, J. Qian, and S. Huang, "A review of recent work on using metal–organic frameworks to grow carbon nanotubes," *Chem. Commun.*, vol. 56, no. 74, pp. 10809–10823, Sep. 2020, doi: 10.1039/d0cc04015k.

[14] J. Ding *et al.*, "Chitosan hydrogel derived carbon foam with typical transition-metal catalysts for efficient water splitting," *Carbon*, vol. 177, pp. 160–170, Jun. 2021, doi: 10.1016/j.carbon.2021.01.160.

[15] R. Mohan, *Plasma Processing of Nanomaterials*. Boca Raton, FL, USA: CRC Press, 2017.

[16] M. Cantoro *et al.*, "Effects of pre-treatment and plasma enhancement on chemical vapor deposition of carbon nanotubes from ultra-thin catalyst films," *Diamond Rel. Mater.*, vol. 15, nos. 4–8, pp. 1029–1035, Apr. 2006, doi: 10.1016/j.diamond.2006.01.007.

[17] S. Hofmann *et al.*, "Effects of catalyst film thickness on plasma-enhanced carbon nanotube growth," *J. Appl. Phys.*, vol. 98, no. 3, Aug. 2005, Art. no. 034308, doi: 10.1063/1.1989432.

[18] S. K. Srivastava, A. K. Shukla, V. D. Vankar, and V. Kumar, "Growth, structure and field emission characteristics of petal like carbon nanostructured thin films," *Thin Solid Films*, vol. 492, nos. 1–2, pp. 124–130, Dec. 2005, doi: 10.1016/j.tsf.2005.07.283.

[19] S.-C. Chang, T.-C. Lin, and C.-Y. Pai, "Low-temperature process in growing carbon nanotube," *Microelectron. J.*, vol. 38, nos. 6–7, pp. 657–662, Jun. 2007, doi: 10.1016/j.mejo.2007.06.002.

[20] T. Labbaye *et al.*, "Enhancement of catalytic effect for CNT growth at low temperature by PECVD," *Appl. Surf. Sci.*, vol. 453, pp. 436–441, Sep. 2018, doi: 10.1016/j.apsusc.2018.05.059.

[21] M. W. Lee, M. A. S. M. Haniff, A. S. Teh, D. C. S. Bien, and S. K. Chen, "Effect of Co and Ni nanoparticles formation on carbon nanotubes growth via PECVD," *J. Experim. Nanosci.*, vol. 10, no. 16, pp. 1232–1241, Nov. 2015, doi: 10.1080/17458080.2014.994679.

[22] Y. Wang *et al.*, "Comparison study of catalyst nanoparticle formation and carbon nanotube growth: Support effect," *J. Appl. Phys.*, vol. 101, no. 12, pp. 1–8, 2007, doi: 10.1063/1.2749412.

[23] L. M. Hoyos-Palacio, A. G. García, J. F. Pérez-Robles, J. González, and H. V. Martínez-Tejada, "Catalytic effect of Fe, Ni, Co and Mo on the CNTs production," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 59, Jun. 2014, Art. no. 012005, doi: 10.1088/1757-899X/59/1/012005.

[24] Z. P. Huang, D. Z. Wang, J. G. Wen, M. Sennett, H. Gibson, and Z. F. Ren, "Effect of nickel, iron and cobalt on growth of aligned carbon nanotubes," *Appl. Phys. A: Mater. Sci. Process.*, vol. 74, no. 3, pp. 387–391, Mar. 2002, doi: 10.1007/s003390101186.

[25] J. H. Yen, I. C. Leu, C. C. Lin, and M. H. Hon, "Effect of catalyst pretreatment on the growth of carbon nanotubes," *Diamond Rel. Mater.*, vol. 13, nos. 4–8, pp. 1237–1241, Apr. 2004, doi: 10.1016/j.diamond.2003.11.045.

[26] S.-C. Chang, T.-C. Lin, T.-S. Li, and S.-H. Huang, "Carbon nanotubes grown from nickel catalyst pretreated with H_2/N_2 plasma," *Microelectron. J.*, vol. 39, no. 12, pp. 1572–1575, Dec. 2008, doi: 10.1016/j.mejo.2008.02.025.

[27] D. J. Yang *et al.*, "Effects of oxygen and nitrogen on carbon nanotube growth using a microwave plasma chemical vapor deposition technique," *Surf. Coatings Technol.*, vol. 167, nos. 2–3, pp. 288–291, 2003, doi: 10.1016/S0257-8972(02)00921-0.

[28] S. Wang, P. Wang, and O. Zhou, "Effects of NH3 plasma pretreatment on the growth of carbon nanotubes," *Diamond Rel. Mater.*, vol. 15, nos. 2–3, pp. 361–364, Feb. 2006, doi: 10.1016/j.diamond.2005.06.036.

[29] W.-P. Wang *et al.*, "The effects of hydrogen plasma pretreatment on the formation of vertically aligned carbon nanotubes," *Appl. Surf. Sci.*, vol. 253, no. 23, pp. 9248–9253, Sep. 2007, doi: 10.1016/j.apsusc.2007.05.060.

[30] L. Valentini, J. M. Kenny, L. Lozzi, and S. Santucci, "Formation of carbon nanotubes by plasma enhanced chemical vapor deposition: Role of nitrogen and catalyst layer thickness," *J. Appl. Phys.*, vol. 92, no. 10, pp. 6188–6194, Nov. 2002, doi: 10.1063/1.1515126.

[31] S.-R. Jian, "Effects of H_2 plasma pretreated Ni catalysts on the growth of carbon nanotubes," *Mater. Chem. Phys.*, vol. 115, nos. 2–3, pp. 740–743, Jun. 2009, doi: 10.1016/j.matchemphys.2009.02.018.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 23,2022 at 07:05:40 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[32] M. A. Lieberman and A. J. Lichtenberg, *Principles of Plasma Discharges and Materials Processing*, 2nd ed. Hoboken, NJ, USA: Wiley, 2005.

[33] Z. Marvi, S. Xu, G. Foroutan, and K. Ostrikov, "Contribution of radicals and ions in catalyzed growth of single-walled carbon nanotubes from low-temperature plasmas," *Phys. Plasmas*, vol. 22, no. 1, Jan. 2015, Art. no. 013504, doi: 10.1063/1.4905522.

[34] R. T. K. Baker, "Catalytic growth of carbon filaments," *Carbon*, vol. 27, no. 3, pp. 315–323, 1989, doi: 10.1016/0008-6223(89)90062-6.

[35] M. Pérez-Cabero, E. Romeo, C. Royo, A. Monzón, A. Guerrero-Ruíz, and I. Rodríguez-Ramos, "Growing mechanism of CNTs: A kinetic approach," *J. Catal.*, vol. 224, no. 1, pp. 197–205, May 2004, doi: 10.1016/j.jcat.2004.03.003.

[36] N. M. Rodriguez, A. Chambers, and R. T. K. Baker, "Catalytic engineering of carbon nanostructures," *Langmuir*, vol. 11, no. 10, pp. 3862–3866, Oct. 1995, doi: 10.1021/la00010a042.

[37] C. S. Cojocaru, A. Senger, and F. Le Normand, "A nucleation and growth model of vertically-oriented carbon nanofibers or nanotubes by plasma-enhanced catalytic chemical vapor deposition," *J. Nanosci. Nanotechnol.*, vol. 6, no. 5, pp. 1331–1338, May 2006, doi: 10.1166/jnn.2006.144.

[38] A. Malesevic, H. Chen, T. Hauffman, A. Vanhulsel, H. Terryn, and C. V. Haesendonck, "Study of the catalyst evolution during annealing preceding the growth of carbon nanotubes by microwave plasma-enhanced chemical vapour deposition," *Nanotechnology*, vol. 18, no. 45, Nov. 2007, Art. no. 455602, doi: 10.1088/0957-4484/18/45/455602.

[39] G.-B. Zheng, K. Kouda, H. Sano, Y. Uchiyama, Y.-F. Shi, and H.-J. Quan, "A model for the structure and growth of carbon nanofibers synthesized by the CVD method using nickel as a catalyst," *Carbon*, vol. 42, no. 3, pp. 635–640, 2004, doi: 10.1016/j.carbon.2003.12.077.

[40] H. Kersten, H. Deutsch, H. Steffen, G. M. W. Kroesen, and R. Hippler, "The energy balance at substrate surfaces during plasma processing," *Vacuum*, vol. 63, no. 3, pp. 385–431, 2001, doi: 10.1016/S0042-207X(01)00350-5.

[41] M. S. Sodha, S. Misra, S. K. Mishra, and S. Srivastava, "Growth of embryonic dust particles in a complex plasma," *J. Appl. Phys.*, vol. 107, no. 10, May 2010, Art. no. 103307, doi: 10.1063/1.3410676.

[42] I. B. Denysenko, S. Xu, J. D. Long, P. P. Rutkevych, N. A. Azarenkov, and K. Ostrikov, "Inductively coupled Ar/CH4/H2 plasmas for low-temperature deposition of ordered carbon nanostructures," *J. Appl. Phys.*, vol. 95, no. 5, pp. 2713–2724, Mar. 2004, doi: 10.1063/1.1642762.

[43] (2012). *UMIST RATE2012*. The UMIST Database for Astrochemistry. [Online]. Available: http://www.udfa.net/

[44] I. Denysenko and K. Ostrikov, "Plasma heating effects in catalyzed growth of carbon nanofibres," *J. Phys. D: Appl. Phys.*, vol. 42, no. 1, Jan. 2009, Art. no. 015208, doi: 10.1088/0022-3727/42/1/015208.

[45] Antonio Miguel da Cruz Baptista Dias. *Modeling of Low Pressure Plasmas in CH4-H2 Mixtures*. Lisbon, Portugal: Instituto Superior Técnico, Universidade de Lisboa, 2012.

[46] H. Mehdipour, K. Ostrikov, and A. E. Rider, "Low- and high-temperature controls in carbon nanofiber growth in reactive plasmas," *Nanotechnology*, vol. 21, no. 45, Nov. 2010, Art. no. 455605, doi: 10.1088/0957-4484/21/45/455605.

[47] S. Hofmann, G. Csányi, A. C. Ferrari, M. C. Payne, and J. Robertson, "Surface diffusion: The low activation energy path for nanotube growth," *Phys. Rev. Lett.*, vol. 95, no. 3, Jul. 2005, Art. no. 036101, doi: 10.1103/PhysRevLett.95.036101.

[48] C. Bower, W. Zhu, S. Jin, and O. Zhou, "Plasma-induced alignment of carbon nanotubes," *Appl. Phys. Lett.*, vol. 77, no. 6, pp. 830–832, Aug. 2000, doi: 10.1063/1.1306658.

[49] M. S. Bell, K. B. K. Teo, R. G. Lacerda, W. I. Milne, D. B. Hash, and M. Meyyappan, "Carbon nanotubes by plasma-enhanced chemical vapor deposition," *Pure Appl. Chem.*, vol. 78, no. 6, pp. 1117–1125, Jan. 2006, doi: 10.1351/pac200678061117.

[50] M. Mao and A. Bogaerts, "Investigating the plasma chemistry for the synthesis of carbon nanotubes/nanofibres in an inductively coupled plasma-enhanced CVD system: The effect of processing parameters," *J. Phys. D: Appl. Phys.*, vol. 43, no. 31, Aug. 2010, Art. no. 315203, doi: 10.1088/0022-3727/43/31/315203.

[51] J. Benedikt, "Plasma-chemical reactions: Low pressure acetylene plasmas," *J. Phys. D: Appl. Phys.*, vol. 43, no. 4, Feb. 2010, Art. no. 043001, doi: 10.1088/0022-3727/43/4/043001.

[52] S.-R. Jian, Y.-T. Chen, C.-F. Wang, H.-C. Wen, W.-M. Chiu, and C.-S. Yang, "The influences of H_2 plasma pretreatment on the growth of vertically aligned carbon nanotubes by microwave plasma chemical vapor deposition," *Nanosc. Res. Lett.*, vol. 3, no. 6, pp. 230–235, Jun. 2008, doi: 10.1007/s11671-008-9141-5.

[53] C. Bower, O. Zhou, W. Zhu, D. J. Werder, and S. Jin, "Nucleation and growth of carbon nanotubes by microwave plasma chemical vapor deposition," *Appl. Phys. Lett.*, vol. 77, no. 17, pp. 2767–2769, Oct. 2000, doi: 10.1063/1.1319529.

[54] S. Hofmann *et al.*, "*In situ* observations of catalyst dynamics during surface-bound carbon nanotube nucleation," *Nano Lett.*, vol. 7, no. 3, pp. 602–608, Mar. 2007, doi: 10.1021/nl0624824.

[55] H. K. Jung and H. W. Lee, "Effect of catalytic layer thickness on diameter of vertically aligned individual carbon nanotubes," *J. Nanomater.*, vol. 2014, pp. 1–6, Jan. 2014.

[56] S. Hofmann *et al.*, "Effects of catalyst film thickness on plasma-enhanced carbon nanotube growth," *J. Appl. Phys.*, vol. 98, no. 3, Aug. 2005, Art. no. 034308, doi: 10.1063/1.1989432.

[57] R. Xie, G. Zhong, C. Zhang, B. Chen, C. S. Esconjauregui, and J. Robertson, "Diameter and wall number control of carbon nanotubes by chemical vapor deposition," *J. Appl. Phys.*, vol. 114, no. 24, Dec. 2013, Art. no. 244302, doi: 10.1063/1.4851656.

[58] Z. Ma, S. Zhou, C. Zhou, Y. Xiao, S. Li, and M. Chan, "Synthesis of vertical carbon nanotube interconnect structures using CMOS-compatible catalysts," *Nanomaterials*, vol. 10, no. 10, pp. 1–13, 2020, doi: 10.3390/nano10101918.

[59] C. Prawoto, S. Li, and M. Chan, "Impact of CNT diameter distribution on CNT filled via scaling," in *Proc. IEEE Int. Conf. Electron Devices Solid State Circuits (EDSSC)*, Jun. 2018, pp. 1–2, doi: 10.1109/EDSSC.2018.8487148.



Umang Sharma was born in Delhi, India. She received the bachelor's degree from Hansraj College, University of Delhi, Delhi, in 2012, and the master's degree from the Visvesvaraya National Institute of Technology, Nagpur, India, in 2015. She is currently pursuing the Ph.D. degree with the Department of Applied Physics, Delhi Technological University (DTU), Delhi.

She is currently working in the fields of nucleation and growth of carbon nanotubes (CNTs) in a reactive plasma environment.

Ms. Sharma was a recipient of DST-INSPIRE Fellowship provided by the Department of Science and Technology (DST), Government of India.



Suresh C. Sharma (Senior Member, IEEE) was a Monbusho Post-Doctoral Fellow under Japanese Government Fellowship, Department of Physics, Faculty of Science, Ehime University, Matsuyama, Japan, from October 1997 to March 1999. In addition, he has been the Japan Society for the Promotion of Science (JSPS) (Invitation) Post-Doctoral Fellow and a Visiting Researcher with the Centre for Atomic and Molecular Technologies (CAMT), Osaka University, Osaka, Japan, from May 2004 to October 2005. He has been a Professor with the Department of Applied Physics, Delhi Technological University (DTU), Delhi, India, and also has administrative responsibility of Dean (Acad-PG). He established Plasma and Nano-Simulation Research Laboratory. He has guided 14 Ph.D. students and several M.Tech. and B.Tech. students. He has published 186 research articles in journals of international and national repute and proceedings of international and national conferences.

Dr. Sharma is a member of the American Physical Society (APS), USA, and many more. He received the Commendable Research Award for Excellence in Research by DTU for four consecutive years, i.e., March 2018, March 2019, March 2020, and February, 2021. He was awarded the Young Scientist Project as a Principal Investigator by the Department of Science and Technology (DST), Government of India, from 1997 to 1999. Also, he was awarded as a Senior Research Associate under the Scientist's Pool Scheme by the Council of Scientific and Industrial Research (CSIR), Government of India, from 1999 to 2002.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 23,2022 at 07:05:40 UTC from IEEE Xplore. Restrictions apply.

arXiv:2203.12192v1 [cs.CV] 23 Mar 2022

Learning to Censor by Noisy Sampling

Ayush Chopra[1], Abhinav Java[2], Abhishek Singh[1], Vivek Sharma[1], and Ramesh Raskar[1]

[1] Massachusetts Institute of Technology, Cambridge, MA
[2] Delhi Technological University, Delhi, India

Abstract. Point clouds are an increasingly ubiquitous input modality and the raw signal can be efficiently processed with recent progress in deep learning. This signal may, often inadvertently, capture sensitive information that can leak semantic and geometric properties of the scene which the data owner does not want to share. The goal of this work is to protect sensitive information when learning from point clouds; by censoring the sensitive information before the point cloud is released for downstream tasks. Specifically, we focus on preserving utility for perception tasks while mitigating attribute leakage attacks. The key motivating insight is to leverage the localized saliency of perception tasks on point clouds to provide good privacy-utility trade-offs. We realise this through a mechanism called *Censoring by Noisy Sampling* (*CBNS*), which is composed of two modules: i) Invariant Sampler: a differentiable point-cloud sampler which learns to remove points invariant to utility and ii) Noisy Distorter: which learns to distort sampled points to decouple the sensitive information from utility, and mitigate privacy leakage. We validate the effectiveness of CBNS through extensive comparisons with state-of-the-art baselines and sensitivity analyses of key design choices. Results show that CBNS achieves superior privacy-utility trade-offs on multiple datasets.

1 Introduction

Proliferation of 3D acquisition systems such as LiDARs, ToF cameras, structured-light scanners has made it possible to sense and capture the real-world with high fidelity. Point clouds are emerging as the preferred mode to store the outputs of these 3D sensors given that they are lightweight in memory and simple in form. Recent advances in deep learning have allowed to directly process the raw sensor output; which has enabled use of point clouds for diverse perception tasks across classification [1,2,3,4,5,6], semantic segmentation [7,8,9,10], object detection [11,12,13,14], and registration [15,16]. This is facilitating algorithms for critical applications across autonomous navigation, precision surgery and secure authentication.

The deployment of downstream algorithms in these critical domains implies that the sensor often captures sensitive information, which the user would like to keep private. This is then inadvertently encoded in representations learned from the signal [17], leaking several semantic and geometric properties of the scene.

Consider for instance, the robotic vacuum cleaners which use LiDAR sensors to efficiently navigate inside the house. The captured signal is also sufficient to localize and map the entire house (via SLAM) as well as track and surveil individuals (via object detection). Similarly, this is also valid for the popular *face-id* experience in recent smartphones which use structured light to capture point clouds of the owner(s) face and use it for authentication, locally on-device. It is well understood that a lot of semantic information (age, gender, expression etc.) can be perceived from the point cloud - which the user may not be willing to share. With the emergence of strict regulations on data capture and sharing such as HIPAA [18], CCPA [3], capturing such sensitive information can create legal liabilities. The goal of this paper is to alleviate such privacy concerns, while preserving utility, by transforming the point cloud to censor sensitive information *before* it is released for downstream utility tasks.

In practice, the design of such transformation functions depends upon the definition of the utility task and privacy attack. Most prior work has focused on preserving the utility of geometric tasks (image-based localization, SLAM, SfM, etc.) while protecting against input reconstruction attacks [19]. For these setups, the dominant idea is to transform the point cloud into 3D line cloud [20] which obfuscates the semantic structure of the scene while preserving utility for camera localization [21], SLAM [22], SfM [23] etc. In contrast, we focus on providing utility for perception tasks (classification, detection, segmentation etc.) while mitigating sensitive attribute leakage [24]. We posit that projecting to line clouds is an infeasible transformation for perception tasks because: i) line clouds disintegrates the semantic structure of the scene required for perception which worsens the utility. This is visualized in [20] and validated by our analysis in section 6; and ii) line clouds are now also vulnerable to inversion attacks, as recently shown in [25], which worsens the privacy. We propose *Censoring by Noisy Sampling (CBNS)* as an alternate transformation for censoring point clouds, which provides improved privacy-utility trade-offs.

The motivating insight for *CBNS* is that performance on perception tasks (utility) only depends upon only a small subset of points (critical points) such that removing (or *sampling*) other non-critical points does not change prediction. Leveraging this for censoring point clouds presents two challenges: *First*, conventional point cloud sampling methods are designed to improve compute efficiency while retaining maximal information about a specific task. Hence, we need to design methods that can jointly sample critical points for the utility task and remove information *invariant* to utility. *Second*, this invariant sampling is necessary but not sufficient, as critical points for task and sensitive attributes can overlap; as we observe through quantitative analysis in section 3.1. We develop *CBNS* to overcome these challenges - i) by introducing an invariant sampler that balances privacy-utility trade-off in its sampling via an adversarial contrastive objective (ℓ_{aco}); ii) by designing a noisy distortion network that adds sample-specific noise to minimize the overlap between task and sensitive information in

[3] `https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375`

an utility conducive manner. We demonstrate the effectiveness of our solution in section 4.

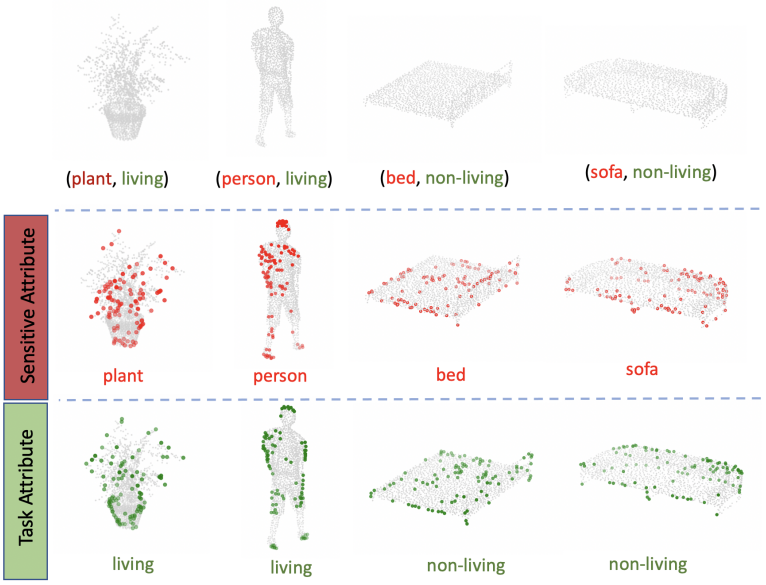
Contributions: Our CBNS is an end-to-end learning framework for protecting sensitive information in perception tasks by dynamically censoring point clouds. CBNS is composed of: i) an invariant sampler that learns to sample non-sensitive points in a task-oriented manner by balancing privacy-utility, ii) a noisy distorter that learns to randomize sampled points for utility conducive removal of sensitive information. We demonstrate the effectiveness of our framework through extensive comparisons against strong baselines and analyses of key design choices. Results show that CBNS significantly improves privacy-utility trade-offs on multiple datasets.

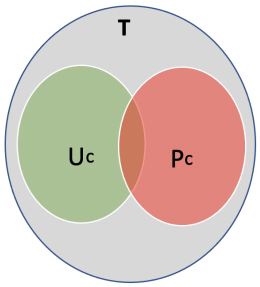
2 Problem Formulation

This section formalises the notation for our task, the threat and attack models, and our privacy definition.

Notation: Consider a data owner O with a point cloud dataset $D_O = (P, Y)$ of N datapoints and (p, y) denotes a paired sample, for $p \in P$ and $y \in Y$. Specifically, $p \in R^{m \times d}$ is a *point cloud* defined as an unordered set of m elements with d features; and y is a *label set* of k attributes describing p. For instance, p can be a 3D point cloud representing a human face ($p \in R^{m \times 3}$) with the *set* y containing categorical attributes that indicate the {*age, gender, expression*} ($k = 3$) of the individual. For every pair $(p, y) \in D_O$, certain attributes in the label set y represent sensitive information which the data owner (O) wants to keep private (y_s) but is comfortable sharing the non-sensitive (or task) information (y_t), such that ($y = y_s \cup y_t$). The risk of leaking this sensitive information prevents the data owner from sharing D_O with untrusted parties; especially with recent progress in deep learning where attackers can efficiently learn functions (F) that can directly map the raw point cloud p to any attribute $a \in y$, where $a = F(p)$ [14,6]. Trivially omitting y_s from y to share the dataset of paired samples $\{(p, y_t)\}$ is not enough since the sensitive information is encoded in p and can be inferred by attackers (*e.g.* using pre-trained models or auxilliary datasets). Hence, to facilitate data sharing with untrusted parties, it is essential to *censor* the sensitive information in p (that leaks y_s) *before* the dataset can be released. Our goal is to learn such a transformation function ($T(\theta_T; \cdot)$) that censors each sample in D_O by generating $\hat{p} = T(p)$. This allows to release $(\hat{p}, y_t)$ instead of (p, y_t). Henceforth, we denote this *censored dataset* as $(\hat{P}, \hat{Y})$. The key challenge for T is to preserve utility of the task information (y_t) while protecting privacy of sensitive information (y_s). In practice, the design of T depends upon definition of the utility task and privacy attack. We focus on providing utility for perception tasks while mitigating attribute leakage attacks [24].

Threat Model: We assume that the attacker gains access to a subset of censored point clouds $(\hat{P}, Y)$ intending to infer sensitive attributes (y_s). This is practical since data owners typically share data with external entities for storage and also for monetary incentives. Further, the threat is also valid if the attacker gains





Critical Points	mIoU (task, sensitive)	
	ModelNet	FaceScape
50	0.171	0.117
64	0.215	0.128
100	0.315	0.223
500	0.663	0.572

Fig. 1: **Premise Validation**. (*Left*) - perception on point clouds depends upon few critical points. (*Right - bottom*) table shows overlap of critical points for sensitive (P_c) and task (U_c) attributes; (*Right - top*) goal of a censoring mechanism is to remove $T - U_c$ and reduce $P_c \cap U_c$. We bridge these ideas in section 3.1 and introduce CBNS in section 3.2 to accomplish both goals.

access to the learned censoring function T which can be used to simulate a dataset that mimics the censored point cloud distribution. This is practical if the attacker is one of the data owners that has access to T. We note that unlike differential privacy [26] that protects identifiability, our threat model protects sensitive attribute leakage [24].

Attack Model: We model an attacker that uses the released dataset to train state-of-the-art DNN models that can directly predict the sensitive attribute from the point cloud. This attacker may use arbitrary models which are not accessible during training, and hence we mimic a proxy attacker for learning the censoring transformation. We represent the proxy attacker by a state-of-the-art DNN parameterized as $f_A(\theta_A; \cdot)$ and trained on censored point clouds $\hat{P}$.

Privacy: Following the setup described by Hamm *et al.* [27], we define privacy as the expected loss over the estimation of sensitive information by the attacker. This privacy loss L_{priv}, given ℓ_p norm, for an attacker can be stated as:

$$L_{priv}(\theta_T, \theta_A) \triangleq E[\ell_p(f_A(T(p; \theta_T); \theta_A), y_s)]$$

Under this definition, releasing sensitive information while preserving privacy manifests as a *min-max* optimization between the data owner and the attacker. However, for training the model parameters, we use a proxy adversary from which gradients can be propagated. We refer to the attack performed by this proxy attacker as an *online attack* and note that this allows mimicking worst-case setups where the attacker can also dynamically adapt using protected data and sensitive label information [28]. We note that our definition of privacy sig-

nificantly differs from differential privacy [26] since we aim to protect sensitive attributes instead of the identity of the data owner.

3 Methodology

In this section, we introduce *Censoring by Noisy Sampling (CBNS)* - a mechanism to censor point clouds for enabling utility of perception tasks while protecting leakage attack on sensitive attributes. We begin by discussing our key motivating insight and then delineate the proposed *CBNS* mechanism.

3.1 Premise Validation

State-of-the-art DNN models such as PointNet [1], PointNet++ [2], DGCNN [3] have successfully handled the irregularity of the raw point cloud and achieved remarkable progress on perception tasks such as classification, segmentation etc. Extensive empirical analysis of these networks shows that classification performance depends upon only a small subset of points (*critical points*) such that removing other non-critical points does not change prediction. Figure 1 visualizes the critical points for perceiving the category (*plant, person, bed, sofa*) and super-type (*living, non-living*) of a few ModelNet dataset samples [29] by training PointNet. The observed *localized* (i.e. depends on critical points) and *task-oriented* (different across category and super-type) saliency is a key motivating insight for censoring point clouds for privacy-utility release.

Assume a privacy-utility scenario where the super-type is task (utility) and the category is the sensitive attribute (privacy). In principle, we achieve good utility (predicting super-type) by only keeping the necessary critical points. Since critical points are visualized via post-training analysis, in practice, this presents two challenges for data release: *First,* conventional point cloud sampling methods are designed to improve compute efficiency while retaining maximal information for a specific task. Hence, we need to design methods that can jointly sample critical points for the utility task and remove information *invariant* to utility. *Second*, this invariant sampling is necessary but not sufficient, as critical points for task and sensitive attributes can overlap. For instance, the top-100 critical points for super-type and category in ModelNet have mIoU of *31%* (table in figure 1). Hence we also need to distort the sampled points to decouple the sensitive and task attributes. The venn-diagram in figure 1 helps visualize this constraint. Intuitively, we want to learn a censoring transformation that can concurrently remove $T - U_c$ and reduce $P_c \cap U_c$. With this motivation, next we describe our proposed mechanism to censor point clouds.

3.2 Censoring by Noisy Sampling

The task of censoring to mitigate information leakage involves three key entities: i) Data Owner (O), ii) User (U) and iii) Attacker (A). O censors each sample in the dataset to protect sensitive information and releases it for U, an untrusted

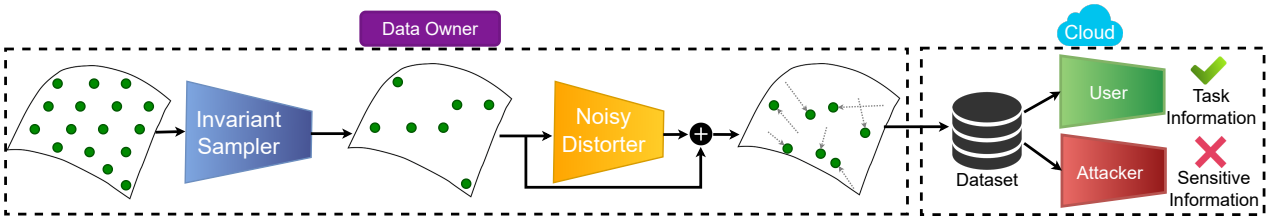


Fig. 2: **Censoring by Noisy Sampling** through a a three-player game: i) Data Owner (O), ii) User (U) and iii) Attacker (A). O censors every sample in the dataset and shares it with U to train a model on the task information. A intercepts the released dataset and attempts to leak the sensitive information. We design *CBNS*, composed of two modules: a) *Invariant Sampler* and b) *Noise Distorter*, to help O enable U's task and avert A's attack. The design of the mechanism is delineated in section 3.2.

but honest entity, to learn a model on the non-sensitive information. A intercepts the released dataset and queries it to leak the sensitive information. We design *Censoring by Noisy Sampling* (**CBNS**) to help O facilitate the task of U and prohibit the task of A. This three-player game [30] is summarized in figure 2 and described below:

a) Owner owns the point cloud dataset (P, Y) which is to be released. This entity censors the sensitive information in each sample (p, y) for $p \in P$ and $y \in Y$. CBNS is composed of two parametric modules, applied sequentially: i) *Invariant Sampler* ($f_S(\theta_S; \cdot)$), ii) *Noisy Distorter* ($f_D(\theta_D; \cdot)$).

First, $p \in R^{m \times d}$ is passed through $f_S(\theta_S; \cdot)$, differentiable DNN sampler built upon [31], which selects a subset of r points relevant for encoding task information to generate an intermediate point cloud $p_s \in R^{r \times d}$, where $r << m$. In contrast to conventional sampling methods which are aimed at improving compute efficiency while preserving all information, f_S is a lossy sampler designed to remove points *invariant* to utility. The extent and quality of censoring depends upon design of f_S, which we analyse in section 6. Releasing this p_s may still leak sensitive attribute through points which overlap with utility (mIoU table in fig 1). *Next*, p_s is passed through $f_D(\theta_D; \cdot)$ which generates task-oriented noise to distort p_s. This is done to decouple overlapping sensitive (privacy) and task (utility) information and executed in the following steps: i) $\mu, \sigma = f_D(p_s; \theta_D)$, ii) $z_s \sim \mathcal{N}(\mu, \sigma^2)$, iii) $\hat{p} = p_s + z_s$ where the sampled noise $z_s \in R^{r \times d}$. All the censored point cloud samples $(\hat{p}, \hat{y})$ are aggregated into the dataset $(\hat{P}, \hat{Y})$ and then released for use by untrusted parties.

b) User is an untrusted but honest entity that receives the released dataset $(\hat{P}, \hat{Y})$ and uses it to infer the non-sensitive task attributes. The *User* trains a state-of-the-art DNN model that can learn to directly map the raw point cloud signal to the task attribute (y_t). For training *CBNS*, we mimic the real user using a proxy user which is parameterized with $f_U(\theta_U; \cdot)$ that consumes $\hat{p} \in \hat{P}$ to predict $y_t \in \hat{Y}$.

c) Attacker is an untrusted semi-honest entity that acquires access to $(\hat{P}, Y)$ with the intention to leak sensitive information about the data owner. The attacker is parameterized with $(f_A(\theta_A; \cdot))$ which is not accessible during design of *CBNS*. Hence, for training, we use a proxy attacker that consumes $\hat{p} \in \hat{P}$ to predict the sensitive attribute (y_s). We note that f_A is a proxy attacker used for training *CBNS*, while a *distinct offline attacker* (f), not used for training, is employed for evaluation tasks.

Training: The utility loss is approximated based on the performance of the proxy user which depends upon parameters θ_T $(\theta_T = \theta_S \cup \theta_D)$ and θ_U that are learned during training. The objective function is given by:

$$L_{util}(\theta_S, \theta_D, \theta_U) \triangleq E[\ell_u(f_U(f_D(f_S(p; \theta_S)\theta_D); \theta_U), y_t)] \tag{1}$$

where, θ_S, θ_D are parameters of CBNS and θ_U are parameters of the the proxy user network; and ℓ_u is the cross entropy loss (ℓ_{cce}).

The privacy loss is approximated based on the performance of the proxy attacker which depends upon parameters θ_T, θ_A learned during training. The objective function is given by:

$$L_{priv}(\theta_S, \theta_D, \theta_A) \triangleq E[\ell_a(f_A(f_D(f_S(p; \theta_S)\theta_D); \theta_A), y_s)] \tag{2}$$

where, θ_S, θ_D are parameters of CBNS and θ_A are parameters of the proxy attacker; and ℓ_a denotes the attacker loss. For training CBNS, we define ℓ_a with the following objective function:

$$\ell_a = \alpha * (\ell_{cce}(f_A(f_D(p_s)), y_s)) + (1 - \alpha) * \ell_{aco}(f_D(p_s), y_s, y_t) \tag{3}$$

where ℓ_{cce} is categorical cross-entropy and ℓ_{aco} is an adversarial contrastive loss, inspired from [32]; and α is a scalar hyperparameter.

Adversarial Contrastive Loss (ℓ_{aco}): Our analysis in section 6 shows that using ℓ_{aco} significantly improves privacy-utility trade-offs. Consider, for instance, age (y_s) to be the sensitive attribute. In the conventional contrastive loss, we encourage to pull positive samples (same age) closer within the local neighborhood and negative samples (different age) apart. In contrast, ℓ_{aco} pulls negative samples closer (different age) and positive samples (same age) apart. The goal here is to map all different ages within a very small neighborhood of each other, to deter the attacker from learning discriminative representations of age. Intuitively, this guides *CBNS* to transform the released point cloud to introduce ambiguity in representations used by an attacker to correctly discriminate between ages, resulting in better privacy. In other words, the ℓ_{aco} forces to map the different ages to a single point in the embedding space.

The proxy attacker and proxy user have access to supervised data and attempt to minimize their losses L_{util} and L_{priv} respectively. CBNS is trained to minimize L_{util} and maximize L_{priv}, simulating an implicit min-max optimization for these two components. Furthermore, CBNS also minimizes a soft-projection loss [31] (L_{sample}) to improve stability of f_S and ensure that the sampler is

constrained to *select* points from the input set (instead of interpolating). This overall objective can be summarized as:

$$\min_{\theta_S,\theta_D}\left[\max_{\theta_A} L_{priv}(\theta_S,\theta_D,\theta_A) + \lambda \min_{\theta_S,\theta_D,\theta_U} L_{util}(\theta_S,\theta_D,\theta_U) + \min_{\theta_S} L_{sample}(\theta_S)\right] \tag{4}$$

Here, λ is a chosen hyperparameter to help regulate the trade-off between privacy and utility.

Inference: A data owner with access to a point cloud dataset (P, Y) can use CBNS to generate the censored dataset $(\hat{P}, \hat{Y})$ and release it for use by untrusted parties. This released dataset can be used for either: i) training new models or ii) running inference using pre-trained models. This is possible only because the output space of the censoring mechanism is same as the input space. In other words, censoring a point cloud using CBNS also generates a point cloud. In contrast: i) most work for censoring images requires releasing neural activations which cannot be processed by arbitrary designed networks [28,33,32] and ii) prior work for censoring point clouds releases line clouds [21,20] which, while useful for geometric tasks, are incompatible for off-the-shelf perception networks.

4 Experiments

In this section, we specify the datasets and baselines used, define the evaluation protocols and summarize implementation details for the results presented in this work. Details about the code are included in the appendix.

Datasets: a) FaceScape [34] consists of 16,940 textured 3D faces, captured from 938 subjects each with multiple categorical labels for age (100), gender (2) and expression (20). For our experiments, we sample 1024 3D points from the surface of each face mesh using Pytorch 3D [35]. To simulate privacy-utility analysis, we use the expression as the task attribute (utility) and gender as the sensitive attribute (privacy). We choose this configuration because the default critical points for the two attributes overlap (need for noisy distorter) and are also distributed across the point cloud (need for invariant sampling), but human performance motivates that they are can be inferred independently. This provides a good benchmark for testing the efficacy of CBNS. **b) ModelNet** [29] consists of 12,311 CAD-generated meshes across 40 categories (object types) of which 9,843 training and 2,468 testing data points. For our experiments, we uniformly sample 2048 3D points from the mesh surface and then project them onto a unit sphere. Since each input sample only has one attribute (object type), we adapt the strategy used by [30] to simulate our privacy-utility analysis. Specifically, to identify an additional attribute, we divide the 40 classes into two super-types: living and non-living. We anticipate living objects to have visually discriminative features instead of geometric shapes of non-living objects. For example, the task of classifying an object as living (*person, plant*) or non-living (*sofa, bed*) should not reveal any information about its underlying identity (*person, plant, sofa,*

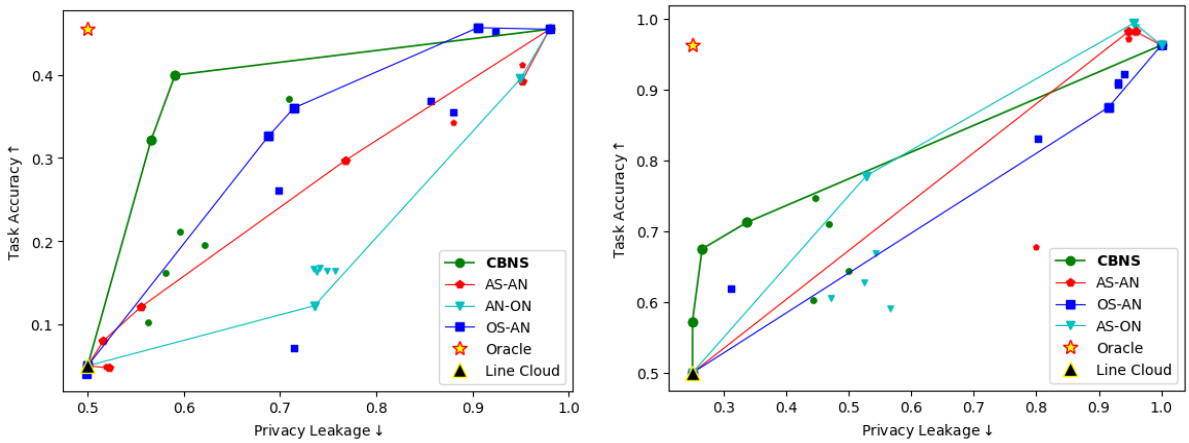
bed). We use super-types as task attribute and object type as sensitive attribute.

Baselines: Prior work in censoring point clouds has largely focused on geometric tasks (image-based localization, SLAM, SfM, etc) via line clouds [20,21,23,22]. However, our analysis (section 6) shows that line clouds are a weak baseline for perception tasks (classification, detection, etc). To ensure rigorous analysis, we define multiple baselines inspired by work in 2D vision that has focused on censoring images (and their activations) for perception tasks while mitigating attribute leakage. The baselines differ in the design of sampling and noisy distorter modules - which may be task-oriented (learned using data) or task-agnostic (deterministic). Our mechanism *CBNS* is equivalent to the *Oriented Sampling - Oriented Noise (OS-ON)* configuration. The baselines are summarized below, with more details in the appendix:

- **Agnostic Sampling - Agnostic Noise (AS-AN):** Uses farthest point sampling (FPS) with fixed gaussian distribution for noise. This is inspired from [36] which formalises differential privacy for images through Gaussian noise, without any learning. While [36] adds noise to images, we add it to a sampled point cloud obtained via FPS.
- **Agnostic Sampling - Oriented Noise (AS-ON):** Uses FPS and learns parameters of the gaussian distribution for noise (as in CBNS) using *maximum likelihood* attacker training. This is inspired from [33] which learn a noise distribution to obfuscate image activations. While [33] adds noise to image activations, we add it to a sampled point cloud obtained via FPS.
- **Oriented Sampling - Agnostic Noise (OS-AN):** Uses differentiable point cloud sampling (as in CBNS) with fixed gaussian distribution for noise. This is inspired from [28] which does channel pruning of image activations to remove sensitive information. While [28] trains a DNN to prune neural activations, we train a DNN to sample point clouds [31].

Evaluation Protocol: We evaluate different techniques by comparing the *privacy-utility trade-off*. For each technique, utility is a measure of *User*'s performance by training on the *censored* dataset and privacy is of the *Attacker*'s performance (as described in section 2). Specifically, we quantify information leakage from the dataset by comparing the performance of an attacker to correctly infer sensitive information from the censored dataset. For this analysis, we simulate a *worst case* attacker that dynamically adapts to the privatization scheme. This adaptation is modeled using a pretrained attacker model that is fine-tuned on the *censored* dataset and then evaluated on a censored test set. Inspired by [30], we quantify privacy-utility trade-offs curves by different techniques using area under the pareto-optimal curve denoted as the normalized hypervolume (**NHV**) [37]. *Higher NHV value indicates a better privacy-utility trade-off.*

Implementation Details: Unless stated otherwise, we use 3D-point clouds ($d = 3$) with the Invariant Sampler producing 64 points ($m = 64$) and with PointNet [1] as the backbone architecture for both f_U and f_A. For FaceScape, $n = 1024$ and the dataset split is 80% training and 20% testing examples. For ModelNet, $n = 2048$ and we restrict experimentation to a smaller set of 4 classes:



(a) FaceScape [34] (b) ModelNet [29]

Fig. 3: **Privacy-Utility trade-off** comparison for different techniques by interpolating their best performing points. The oracle point refers to the best possible censoring mechanism. We note that the line cloud [20,22] techniques do not yield any trade-off due to its incompatibility with perception tasks.

2 living and 2 non-living to ensure ease of analysis and avoid data imbalance issues (ModelNet has 2 living and 38 non-living objects). All experiments are implemented using Pytorch and conducted on 2 TITAN-X GPUs. We will release both our code and dataset splits for reproducibility.

5 Results

We report performance comparison with baselines on ModelNet and FaceScape, in Table 1. For completeness, apart from the baselines defined in section 4, we also benchmark with two extreme scenarios: i) **No-privacy**: default case where the data is released without any censoring and, ii) **Oracle**: best possible case with some ideal censoring mechanism which *does not exist*. Results show that **CBNS** significantly outperforms all baselines, on both datasets; as evident from higher NHV. *First,* **CBNS** consistently provides the best privacy leakage - often very close to random chance. Specifically, privacy leakage with CBNS is 0.5890 for 2-way classification in FaceScape and 0.2657 for 4-way classification in ModelNet. *Second,* the peak privacy-utility trade-off for **CBNS** is closest to the oracle, on both datasets. Specifically, when CBNS is used for data release on FaceScape, the *User* achieves a utility of *0.4013* (88% of the oracle) while the *Attacker* performance is close to random chance (0.5890). This corresponds to *13% less privacy leakage while also providing 4% more utility* than the closest baseline (OS-AN). However, a higher fall in utility is observed on ModelNet, which can be attributed to the fact that the task and sensitive attributes are more strongly coupled, than in FaceScape. *Finally,* these observations are corroborated visually by the privacy-utility trade-off curves in Figure 3 where CBNS has the highest area-under-curve (correlated with the hyper-volume).

Method	FaceScape			ModelNet		
	Privacy (↓)	Utility (↑)	NHV (↑)	Privacy (↓)	Utility (↑)	NHV (↑)
No-privacy	0.9515	0.4549	-	0.95	0.9625	-
Line Cloud [20,22]	0.5000	0.0500	-	0.2500	0.5000	-
AS-AN [36]	0.9511	0.4122	0.1524	0.9469	0.9612	0.6113
AS-ON [33]	0.9391	0.3875	0.1250	0.5281	0.7781	0.6088
OS-AN [28]	0.7143	0.3602	0.1439	0.3125	0.6187	0.6356
CBNS (Ours)	**0.5890**	**0.4013**	**0.1885**	**0.2657**	**0.6750**	**0.6492**
Oracle	0.5000	0.4549	-	0.2500	0.9625	-

Table 1: **Comparison for sensitive attribute leakage**. We compare our approach on sensitive attribute leakage with the existing works and baseline. CBNS outperforms Line Cloud [20,22], AS-AN [36], AS-ON [33], OS-AN [28] and achieves the best privacy-utility trade-off on FaceScape and ModelNet datasets. For the FaceScape, sensitive attribute is gender and task attribute is expression; and for the ModelNet, sensitive attribute is underlying object type and task attribute is super-types (living or non-living).

Technique	Privacy (↓)	Utility (↑)	NHV (↑)
AS-ON [33] (ℓ_{cce})	0.9297	0.3954	0.1322
OS-AN [28] (ℓ_{cce})	0.6942	0.3602	0.1439
CBNS (ℓ_{me})	0.6839	0.413	0.1596
CBNS (ℓ_{cce})	0.5787	0.3638	0.1615
CBNS (ℓ_a)	**0.5707**	**0.3997**	**0.1885**

Table 2: **Design of Invariant Sampler**. Privacy-utility trade-off is influenced by whether transformation is learned, and how it is learned. For FaceScape, sensitive attribute is gender and task attribute is expression. l_{cce}, l_{me} and l_a denotes cross entropy loss, max-entropy loss and CBNS loss respectively.

6 Discussion

We analyse the impact of key design choices for censoring by noisy sampling. Specifically, we study the characterization of both *CBNS* modules: i) Invariant Sampler and ii) Noisy Distorter; and impact of the perception network. For completeness, we also analyse the viability of line clouds for perception tasks. For ease of exposition, we restrict the scope of this analysis to FaceScape dataset.

– **Design of Invariant Sampler:** We study the role of two design choices: i) learning a task-oriented sampler and ii) the attacker loss that is used to optimize the parameters of the learned sampler. Results are presented in Table 2. Please note that we follow the baselines from section 4 and explicitly mention (L_{priv}) the attacker objective in parenthesis. Specifically, l_{cce} is the cross-entropy loss, l_a is our proposed loss (equation 3) and l_{me} is a max-entropy loss used in [30]. We observe the following: *First*, using a learned task-oriented sampler reduces

Technique	Privacy (↓)	Utility (↑)	NHV (↑)
OS (ℓ_{cce})	0.5787	0.3638	0.1615
OS-AN [28] (ℓ_{cce})	0.6942	0.3602	0.1439
CBNS (*shared*, ℓ_{cce})	0.9051	0.4532	0.1586
CBNS (*shared*, ℓ_a)	0.8600	0.4217	0.1625
CBNS (*pointwise*, ℓ_{cce})	0.5689	0.4013	0.1530
CBNS (*pointwise*, ℓ_a)	**0.5707**	**0.3997**	**0.1885**

Table 3: **Design of Noisy Distorter**. Privacy-utility trade-off is influenced by the learned noise and the granularity of the noise parameters (*shared v.s pointwise*). For FaceScape, sensitive attribute is gender and task attribute is expression. l_{cce} and l_a denotes cross entropy loss and CBNS loss respectively.

privacy-leakage by 35% without any loss to utility (row 1 vs 4 and 5) in contrast to a task-agnostic sampler. *Second,* using the proposed adversarial contrastive loss in L_{priv} improves privacy-utility trade-off by increasing utility by 3% without any additional privacy leakage (row 4 vs 5). *Third,* using l_{me} improves utility but with a significant increase in privacy-leakage; as evident from lower NHV. l_{me} is successful for images [30] but fails to generalize to point clouds, which can be attributed to the irregularity in the data structure.

– **Design of Noisy Distorter:** We study the role of three design choices: i) learning task-oriented noise, ii) the attacker loss used to optimize parameters of the learned noise (l_a or l_{cce}) and iii) the granularity of the noise parameters (*shared v.s pointwise*). *Shared* implies that each point is distorted using noise from the same learned distribution (i.e. $z_s \in R^1$) and *Pointwise* implies that each point is distorted from a unique independently learned distributions (i.e. $z_s \in R^r$). Results are presented in Table 3. We follow the baselines from section 4 and define *OS* (row 1) as additional baseline which only uses sampling (*without* noisy deformation). We observe the following: *First,* using noise task-agnostic (row 1 vs 2), or ii) shared task-oriented (row 1 vs 3, 4) does not provide benefit; and are infact worse than no noise baseline (*SO*). *Second, pointwise* noise distributions *significantly* improves performance. (row 3 vs 5; 4 vs 6). This increase in NHV as well as peak privacy-utility trade-offs can be attributed to improved flexibility for adapting to characteristics of sensitive, task attributes and their relationship.. *Third,* the objective function used for learning noise is also important where using l_{aco} in l_a improves privacy-utility trade-off (row 3 vs 4; row 5 vs 6).

– **Impact of Perception Network:** In our threat model, the *Owner* releases dataset for post-hoc access by the *User* and *Attacker*. Hence, the censoring mechanism should be independent of the type of perception networks used by these entities for downstream tasks. We analyse this sensitivity by comparing two different proxy attacker and user networks. Specifically, we use DGCNN [3] which is a recent state-of-the-art network with higher capacity and distinct saliency properties than PointNet [38]. Results are presented in Table 4. We observe the following: *First,* increasing capacity of proxy networks further improves learning of CBNS as evident from better privacy-utility trade-offs (row 1 vs 2; row 3 vs

Technique	Backbone	Privacy (↓)	Utility (↑)	NHV (↑)
OS-AN [28]	PointNet	0.6942	0.3602	0.1439
	DGCNN	0.7056	0.4718	0.2136
CBNS (Ours)	PointNet	0.5707	0.3997	0.1885
	DGCNN	**0.4848**	**0.4100**	**0.2361**
Line Cloud [20,22]	PointNet	0.5000	0.0500	-
	DGCNN	0.5000	0.0500	-

Table 4: **Impact of Perception Network** and **Incompatibility of Line Clouds**. For FaceScape, sensitive attribute is gender and task attribute is expression. CBNS is invariant to the type of attacker network. Using stronger perception network (DGCNN) further improves performance over PointNet and helps achieve near optimal trade-off with our proposed CBNS. Resampling line clouds provides poor (random chance) privacy-utility trade-off.

4). For instance, when CBNS is trained with DGCNN (as against PointNet), the censored dataset provides a better utility of 0.4100 (vs 0.3997) while also significantly reducing privacy leakage to 0.4848 (vs 0.5707). *Second,* importantly, we see that noisy sampling is independent of the downstream network and can generalize to multiple perception networks. Specifically, this is very encouraging since CBNS can concurrently mitigate stronger *attackers* from leaking information by improving the utility of *users* with the stronger perception backbones.

– **Incompatibility of Line Clouds:** We posit that line clouds are an incompatible baseline for perception tasks because i) they destroy semantic structure of the input point cloud which is essential for perception (see visualizations in [20]), and ii) any off-the-shelf perception network: used by both *User* and *Attacker* cannot train on line clouds. To benchmark the performance of line clouds, we generate point clouds by re-sampling line clouds and evaluate performance on our perception queries. Results in Table 4 show that: i) re-sampling line clouds provides extremely poor utility (random chance), and ii) the privacy-utility trade-off cannot be tuned (hence no NHV). Specifically, we observe that resampled line cloud obtain privacy leakage of 0.5000 (for 2-way classification) and utility of 0.05 (on a 20-way classification). Finally, we acknowledge recent work has attacked line clouds to reconstruct point clouds [25] but note that this is equivalent to our No-Privacy baseline which can improve utility but requires mechanisms like noisy sampling to provide privacy.

7 Related Work

Private Imaging. A majority of the existing works in privately sharing data focus on identifiability and anonymization [39,26,40]. In contrast to this line of work, we focus on protecting sensitive attributes. Among the techniques that focus on protecting sensitive attributes [33,28,41,32,42], their tasks are typically limited to image datasets. More recently, privacy for 3D point clouds

has emerged with a focus on geometric queries protecting privacy by releasing line clouds [20,23]. However, To the best of our knowledge, this is the first work in protecting sensitive information leakage for perception tasks in point clouds. Adjacent to research in privately sharing data, privately sharing ML model [43,44,45,46,47] has received interest recently. However, unlike protecting sensitive attributes, model sharing aims to protect the identifiability of training data.

Learning on Point Clouds. Recent advances in deep learning (DL) have allowed to learn directly on raw point clouds; enabling use in diverse perception tasks such as classification, semantic segmentation, object detection, registration etc. Various DL architectures have been proposed starting with PointNet [1], PointNet++ [2] and follow-up works in [48,8,49,50,51,52,53,5,4] that improve the performance over a given task by capturing task-oriented representations. Zheng *et al.* [38] observe that saliency of the point cloud networks is localized and network rely on a small subset of the signal for the task. This observation has led to extensive work in privacy and security [54,55,56,57] utilizing the localized saliency used to design adversarial attack (and defence) mechanisms on the trained models. We note that our setting significantly differs from adversarial attack work since we *protect the dataset* that can be used to train arbitrary models while adversarial methods focus on *attacking/protecting* the robustness of model predictions.

Sampling of Point Clouds. Processing point clouds can be computationally intensive making sampling a popular pre-processing step to alleviate this challenge. Classical methods such as random sampling and FPS [2,4] are task-agnostic and deterministic algorithms for sampling point sets. However, not utilizing task knowledge when sampling hinders performance. Recent techniques [58,31] introduce task-oriented mechanisms for sampling through differentiable approximations. The focus is to improve compute efficiency while preserving the entire signal in the sampled subset. In contrast, our goal is to censor sensitive information during the sampling process. We build upon prior work to introduce a task-oriented point-cloud sampler that censors sensitive information.

Noisy Sampling for Censoring. While not motivated for point clouds, similar intuition has been used for tabular datasets for private coresets [59,60] combines subset (coreset) selection and differentially-private noise to achieve good privacy utility trade-off. Our work is different because: i) our queries involve neural networks so computing sensitivity for DP-noise is infeasible, ii) we only want to protect sensitive attribute. We empirically validate privacy-utility trade-off using benchmark metrics, as described in section 4 and present results in section 5.

8 Conclusion

This focus of this paper is to censor point clouds to provide utility for perception tasks while mitigating attribute leakage attacks. The key motivating insight is to leverage the localized saliency of perception tasks on point clouds to provide good privacy-utility trade-offs. We achieve this through our mechanism called censoring by noisy sampling (*CBNS*), which is composed of two modules: i)

Invariant Sampling - a differentiable point-cloud sampler which learns to remove points invariant to utility and ii) Noise Distorter - which learns to distort sampled points to decouple the sensitive information from utility, and mitigate privacy leakage. We validate the effectiveness of CBNS through extensive comparisons with state-of-the-art baselines and sensitivity analyses of key design choices. Results show that CBNS achieves superior privacy-utility trade-offs.

References

1. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation (2017)
2. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems **30** (2017)
3. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic graph cnn for learning on point clouds. Acm Transactions On Graphics (tog) **38**(5) (2019) 1–12
4. Li, Y., Bu, R., Sun, M., Wu, W., Di, X., Chen, B.: Pointcnn: Convolution on x-transformed points. Advances in neural information processing systems **31** (2018)
5. Wu, W., Qi, Z., Fuxin, L.: Pointconv: Deep convolutional networks on 3d point clouds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2019) 9621–9630
6. Zhao, H., Jiang, L., Jia, J., Torr, P.H., Koltun, V.: Point transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. (2021) 16259–16268
7. Xu, C., Wu, B., Wang, Z., Zhan, W., Vajda, P., Keutzer, K., Tomizuka, M.: Squeezesegv3: Spatially-adaptive convolution for efficient point-cloud segmentation. In: European Conference on Computer Vision, Springer (2020) 1–19
8. Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., Markham, A.: Randla-net: Efficient semantic segmentation of large-scale point clouds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2020) 11108–11117
9. Zhang, Y., Zhou, Z., David, P., Yue, X., Xi, Z., Gong, B., Foroosh, H.: Polarnet: An improved grid representation for online lidar point clouds semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2020) 9601–9610
10. Chen, C., Qian, S., Fang, Q., Xu, C.: Hapgn: Hierarchical attentive pooling graph network for point cloud segmentation. IEEE Transactions on Multimedia **23** (2020) 2335–2346
11. Qi, C.R., Liu, W., Wu, C., Su, H., Guibas, L.J.: Frustum pointnets for 3d object detection from rgb-d data. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2018) 918–927
12. Qi, C.R., Litany, O., He, K., Guibas, L.J.: Deep hough voting for 3d object detection in point clouds. In: proceedings of the IEEE/CVF International Conference on Computer Vision. (2019) 9277–9286
13. Martin, S., Stefan, M., Karl, A., et al.: Complex-yolo: real-time 3d objectdetection on point clouds. In: Computer vision and pattern recognition. (2018)
14. Shi, S., Wang, X., Li, H.: Pointrcnn: 3d object proposal generation and detection from point cloud. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. (2019) 770–779

15. Sarode, V., Li, X., Goforth, H., Aoki, Y., Srivatsan, R.A., Lucey, S., Choset, H.: Pcrnet: Point cloud registration network using pointnet encoding. arXiv preprint arXiv:1908.07906 (2019)
16. Aoki, Y., Goforth, H., Srivatsan, R.A., Lucey, S.: Pointnetlk: Robust & efficient point cloud registration using pointnet. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2019) 7163–7172
17. Song, C., Shmatikov, V.: Overlearning reveals sensitive attributes. arXiv preprint arXiv:1905.11742 (2019)
18. Atchinson, B.K., Fox, D.M.: From the field: the politics of the health insurance portability and accountability act. Health affairs **16**(3) (1997) 146–150
19. Arora, S., Liang, Y., Ma, T.: Why are deep nets reversible: A simple theory, with implications for training. CoRR **abs/1511.05653** (2015)
20. Speciale, P., Kang, S.B., Pollefeys, M., Schönberger, J., Sinha, S.: Privacy preserving image-based localization. In: 2019 Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (June 2019)
21. Speciale, P., Schonberger, J.L., Sinha, S.N., Pollefeys, M.: Privacy preserving image queries for camera localization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. (2019) 1486–1496
22. Shibuya, M., Sumikura, S., Sakurada, K.: Privacy preserving visual slam. In: European Conference on Computer Vision, Springer (2020) 102–118
23. Geppert, M., Larsson, V., Speciale, P., Schönberger, J.L., Pollefeys, M.: Privacy preserving structure-from-motion. In: European Conference on Computer Vision, Springer (2020) 333–350
24. Jia, J., Gong, N.Z.: {AttriGuard}: A practical defense against attribute inference attacks via adversarial machine learning. In: 27th USENIX Security Symposium (USENIX Security 18). (2018) 513–529
25. Chelani, K., Kahl, F., Sattler, T.: How privacy-preserving are line clouds? recovering scene details from 3d lines (2021)
26. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Theory of cryptography conference, Springer (2006) 265–284
27. Hamm, J.: Minimax filter: Learning to preserve privacy from inference attacks. Journal of Machine Learning Research **18**(129) (2017) 1–31
28. Singh, A., Chopra, A., Sharma, V., Garza, E., Zhang, E., Vepakomma, P., Raskar, R.: Disco: Dynamic and invariant sensitive channel obfuscation for deep neural networks (2021)
29. Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3d shapenets: A deep representation for volumetric shapes. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2015) 1912–1920
30. Roy, P.C., Boddeti, V.N.: Mitigating information leakage in image representations: A maximum entropy approach. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2019) 2586–2594
31. Lang, I., Manor, A., Avidan, S.: Samplenet: Differentiable point cloud sampling (2020)
32. Osia, S.A., Shahin Shamsabadi, A., Sajadmanesh, S., Taheri, A., Katevas, K., Rabiee, H.R., Lane, N.D., Haddadi, H.: A hybrid deep learning architecture for privacy-preserving mobile analytics. IEEE Internet of Things Journal **7**(5) (May 2020) 4505–4518
33. Mireshghallah, F., Taram, M., Ramrakhyani, P., Jalali, A., Tullsen, D., Esmaeilzadeh, H.: Shredder: Learning noise distributions to protect inference pri-

vacy. In: Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems. (2020) 3–18
34. Yang, H., Zhu, H., Wang, Y., Huang, M., Shen, Q., Yang, R., Cao, X.: Facescape: A large-scale high quality 3d face dataset and detailed riggable 3d face prediction. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (June 2020)
35. Ravi, N., Reizenstein, J., Novotny, D., Gordon, T., Lo, W.Y., Johnson, J., Gkioxari, G.: Accelerating 3d deep learning with pytorch3d. arXiv:2007.08501 (2020)
36. Fan, L.: Image pixelization with differential privacy. In: IFIP Annual Conference on Data and Applications Security and Privacy, Springer (2018) 148–162
37. Ishibuchi, H., Imada, R., Setoguchi, Y., Nojima, Y.: How to specify a reference point in hypervolume calculation for fair performance comparison. Evolutionary computation **26**(3) (2018) 411–440
38. Zheng, T., Chen, C., Yuan, J., Li, B., Ren, K.: Pointcloud saliency maps. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. (2019) 1598–1606
39. Samarati, P., Sweeney, L.: Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. (1998)
40. Wang, T., Zhang, X., Feng, J., Yang, X.: A comprehensive survey on local differential privacy toward data statistics and analysis. Sensors **20**(24) (2020) 7030
41. Xiao, T., Tsai, Y.H., Sohn, K., Chandraker, M., Yang, M.H.: Adversarial learning of privacy-preserving and task-oriented representations. In: Proceedings of the AAAI Conference on Artificial Intelligence. Volume 34. (2020) 12434–12441
42. Liu, Z., Wu, Z., Gan, C., Zhu, L., Han, S.: Datamix: Efficient privacy-preserving edge-cloud inference. In: European Conference on Computer Vision, Springer (2020) 578–595
43. McMahan, H.B., Moore, E., Ramage, D., y Arcas, B.A.: Federated learning of deep networks using model averaging. arXiv preprint arXiv:1602.05629 (2016)
44. Gupta, O., Raskar, R.: Distributed learning of deep neural network over multiple agents. Journal of Network and Computer Applications **116** (2018) 1–8
45. Du, J., Li, S., Feng, M., Chen, S.: Dynamic differential-privacy preserving sgd. arXiv preprint arXiv:2111.00173 (2021)
46. Ho, S., Qu, Y., Gu, B., Gao, L., Li, J., Xiang, Y.: Dp-gan: Differentially private consecutive data publishing using generative adversarial nets. Journal of Network and Computer Applications **185** (2021) 103066
47. Jordon, J., Yoon, J., Van Der Schaar, M.: Pate-gan: Generating synthetic data with differential privacy guarantees. In: International conference on learning representations. (2018)
48. Liu, Z., Hu, H., Cao, Y., Zhang, Z., Tong, X.: A closer look at local aggregation operators in point cloud analysis. In: European Conference on Computer Vision, Springer (2020) 326–342
49. Yan, X., Zheng, C., Li, Z., Wang, S., Cui, S.: Pointasnl: Robust point clouds processing using nonlocal neural networks with adaptive sampling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2020) 5589–5598
50. Bytyqi, Q., Wolpert, N., Schömer, E.: Local-area-learning network: Meaningful local areas for efficient point cloud analysis. arXiv preprint arXiv:2006.07226 (2020)
51. Xu, Q., Sun, X., Wu, C.Y., Wang, P., Neumann, U.: Grid-gcn for fast and scalable point cloud learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2020) 5661–5670

52. Xiang, T., Zhang, C., Song, Y., Yu, J., Cai, W.: Walk in the cloud: Learning curves for point clouds shape analysis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. (2021) 915–924
53. Lin, C., Li, C., Liu, Y., Chen, N., Choi, Y.K., Wang, W.: Point2skeleton: Learning skeletal representations from point clouds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2021) 4277–4286
54. Lang, I., Kotlicki, U., Avidan, S.: Geometric adversarial attacks and defenses on 3d point clouds (2021)
55. Yang, J., Zhang, Q., Fang, R., Ni, B., Liu, J., Tian, Q.: Adversarial attack and defense on point sets (2021)
56. Liu, D., Yu, R., Su, H.: Extending adversarial attacks and defenses to deep 3d point cloud classifiers (2019)
57. Lee, K., Chen, Z., Yan, X., Urtasun, R., Yumer, E.: Shapeadv: Generating shape-aware adversarial 3d point clouds. arXiv preprint arXiv:2005.11626 (2020)
58. Dovrat, O., Lang, I., Avidan, S.: Learning to sample. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019) 2755–2764
59. Gupta, A., Ligett, K., McSherry, F., Roth, A., Talwar, K.: Differentially private combinatorial optimization. In: Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms, SIAM (2010) 1106–1125
60. Feldman, D., Fiat, A., Kaplan, H., Nissim, K.: Private coresets. In: Proceedings of the forty-first annual ACM symposium on Theory of computing. (2009) 361–370

Silicon
https://doi.org/10.1007/s12633-022-01716-5

ORIGINAL PAPER

Microstructural and mechanical behavior of micro-sized SiC particles reinforced friction stir processed/welded AA7075 and AA6061

Sumit Jain [1] · R. S. Mishra [1]

Received: 1 December 2021 / Accepted: 24 January 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract
In the present study, friction stir welding was employed to successfully fabricate a weld joint of dissimilar aluminum alloys AA7075 and AA6061 reinforced with SiC microparticles. The influence of dispersion pattern of different volume fractions of SiC microparticles under different processing parameters such as rotational speed and traverse speed on microstructural and mechanical behavior was examined. The outcomes demonstrated that dynamic recrystallization and pinning effect of dispersed SiC particles in the SZ, resulting in grains refinement. which significantly influenced the tensile strength, strain rate and micro-hardness. FSWed joints reinforced with higher volume fraction of SiC particles revealed higher mechanical properties. Consequently, FSWed joint produced at rotational speed of 1100 rpm and traverse speed of 40 mm/min incorporated with 10% SiC particles exhibits the higher tensile strength of 261.20 MPa with nearly 91% of joint efficiency and maximum micro-hardness of 186.83 HV, which is approximately 10% higher than the hardness of as received AA7075 base material.

Keywords FSW · SiCmp · Tensile strength · Micro-hardness

1 Introduction

Recent development in almost all industrial sectors, like automobile, naval, and aerospace has been transforming from conventional materials to lightweight materials for manufacturing light and high-strength structural components. Aluminum alloys have outstanding physical and mechanical properties including high heat conductivity, light weight, higher strength with lower weight, and high corrosion resistance. Dissimilar welding of such alloys has drawn enormous interest for industrial applications in almost all the sectors. Fusion welding processes are not appropriate for welding of low melting point metal alloys because of shortcomings like porosity, solidification cracking and high residual stresses and distortion [1–3]. Therefore, friction stir welding appears to be the best alternative for effectively joining dissimilar aluminum alloys, due to its solid-state nature of welding.

In 1991, friction stir welding (FSW) technique was developed by 'The welding institute' (TWI) [4]. FSW is commonly used to join similar and dissimilar materials [5–6]. Fabrication of high-quality weld joints of similar aluminum alloys and aluminum with other materials such as copper, magnesium and steel etc., was successfully done by friction stir welding [7–10]. The development of metal matrix composites (MMCs) based on reinforcement particles is a novel approach to improving mechanical properties (strength, stiffness, hardness, etc.), wear and corrosive resistance [11–13]. The dispersed reinforcement particles obstruct the growth of grain boundaries during dynamic recrystallization via pinning effect, resulting in tiny and equiaxed grains, which enhance the tensile strength, strain rate and micro-hardness of the processed/welded materials [14–16]. The wear resistance of reinforced joints is mainly enhanced due to reduction in surface contact between the wear samples (pin) and the disk by the distributed of reinforcing particles and the strong bonding between the metal matrix and the reinforcing particles [17]. The corrosion resistance of AZ31 reinforced with Al-SiC particles via FSP was found to be enhanced mainly due to the quantity, size and dispersion behavior of inter-metallics (β-$Al_{12}Mg_{17}$) [18]. For these MMCs, a specific size and fraction of reinforcement particles must be used for remarkable mechanical properties enhancement [19]. Enhancement of the weld

✉ Sumit Jain
sumitjain4@gmail.com

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

Published online: 16 March 2022

joint properties of aluminum alloy 6061 was achieved by MMC formation using Al_2O_3 nanoparticles [20]. The weld quality of dissimilar weld joint of aluminum alloys was also improved by the various researchers by reinforcing the nugget zone (NZ) using reinforcement particles. Reinforcing Al_2O_3 particles increased the toughness and corrosion resistance of the weld joints of AA5083 and AA7075 [21]. The tensile strength of aluminum alloy AA6063-T4 and magnesium alloy AZ31-O weld joint reinforced with TiO_2 particles was nearly two times higher than the unreinforced welded joint [22]. The SiC particles were chosen as reinforcement, because of its lower density and lower melting point in comparison to other reinforcements such as Al_2O_3, TiC, etc. The typical problem of substantial differences in coefficients of thermal expansion of aluminum and SiC causes problem during solidification in the fabrication of Al-SiC composites using traditional approaches, and this can be avoided by utilizing FSW due to solid state nature [23, 24]. Several researchers enhanced the mechanical and wear behavior via Al-SiC composite fabrication using FSW/FSP. Sound weld joints of dissimilar aluminum alloys AA2024/AA7075 was fabricated by incorporating different volume fractions of SiC particles (5, 8 and 13%). Homogeneous dispersion of SiC particles in the stir zone was observed with incorporation of 5% SiC particles and exhibited the best weld quality [25]. In comparison to SiC particles, the FSWed joint of AA6101 and AA1350 reinforced with Al_2O_3 particles had improved mechanical properties and wear resistance [26].

Both AA6061 and AA7075 show wide applications, mainly in structural components in industrial sectors like aerospace, shipbuilding and automotive. Joining of AA6061 and AA7075 in aerospace applications by riveting causes stress concentration and higher structural weight [27, 28]. The dissimilar welding of these alloys would enhance the flexibility of structural designs and broaden their applications. FSW can be employed to improve weldability without significantly reducing weld strength. The intent was to fabricate the sound FSWed joint of AA7075/AA6061 for examining the influence of the dispersion pattern of different volume fractions of SiC microparticles along with different combinations of rotational speed and traverse speed on microstructural and mechanical behavior. To the best of the author's knowledge, the FSW of AA7075 and AA6061 reinforced with SiC micro-sized particles has not been explored.

2 Materials and methods

Plates of AA6061-T6 and AA7075-T6 of dimensions 150 × 40 × 6 mm were utilized as base metals for fabricating dissimilar weld joints via FSW/SiCmp, as delineated in Fig. 1b. The optical micrographs, scanning electron microscopy (SEM) micrographs and chemical composition obtained by energy-dispersive-spectroscopy (EDS) analysis of as-received AA7075 and AA6061 are depicted in Fig. 2(a-d). The micro-sized particles of silicon carbide (SiC) of size ranging from 30 to 50 μm with 99% pureness were used as reinforcing candidates. The SEM micrograph and EDS analysis of the as-received SiC microparticles (SiCmp) are depicted in Fig. 3. To incorporate the SiCmp into the weld zone, a groove of depth 3 mm was plowed along the adjoining faces of base plates below 0.3 mm from the top, to avoid the use of the pin-less tool, as depicted in Fig. 1a. The grooves of different widths (0.24, 0.42 and 0.60 mm) were plowed to vary the volume fraction of SiCmp in the weld zone. A thick slurry of SiCmp and 99% pure ethanol was prepared for the proper filling of SiCmp into the grooves. Before filling, the grooves were properly cleaned with acetone to avoid contamination. Thereafter, SiCmp were filled into the grooves and left to dry for about an hour at room temperature.

Positioning of weld plates in the case of dissimilar welding is a very essential consideration for obtaining sound welding and improved mechanical properties. Several studies have found that local temperatures are higher on the advancing side of FSW [29]. The positioning of dissimilar materials could lead to larger asymmetry of temperature. This significant asymmetry in temperature, material flow and stress between retreating and advancing sides significantly affects the weld quality [30]. Aval et al. [31] and Park et al. [32] reported that effective material mixing was obtained when softer material (AA5xxx) was positioned on the advancing side and harder material (AA6xxx) was positioned on the retreating side. The study on FSW of AA6061 and AA7075 by Guo et al. [33] also revealed that more effective material mixing and higher weld properties were obtained when AA 6061 was positioned on the advancing side. Therefore, in the present study, the plates of aluminum alloy AA6061 were positioned on the advancing side (Fig. 1b), and a constant tool tilt angle of 2º to the normal axis was used during FSW. A numeric controlled FSW machine (HYD-40KN) was employed for FSW/SiCmp experimentations. The specialized FSW machine comprises a hydraulic fixture for proper clamping of the plates during FSW, as depicted in Fig. 1e. Experimentation was carried out using a threaded pin profiled non-consumable tool made of H13 tool steel having hardness of 468 HV. Figure 1c shows the tool's dimensional characteristics. All of the trials were carried out with a plunge depth of 0.5 mm and a tool's dwell duration of 10 s. Different combinations of process parameters were employed to achieve homogeneity in SiCmp dispersion for the betterment of mechanical characteristics. As process parameters, tool rotational speeds (RS) of 700, 900, and 1100 rpm, traverse speeds (TS) of 40, 50, and 60 mm/min, and volume fraction (VF) of SiCmp (4, 7, and 10%) were used. The range of process parameters was selected on the basis of preliminary trails and working range of FSW machine. Fifteen different parametric combinations were carried out according to design of experiment (DOE) obtained by

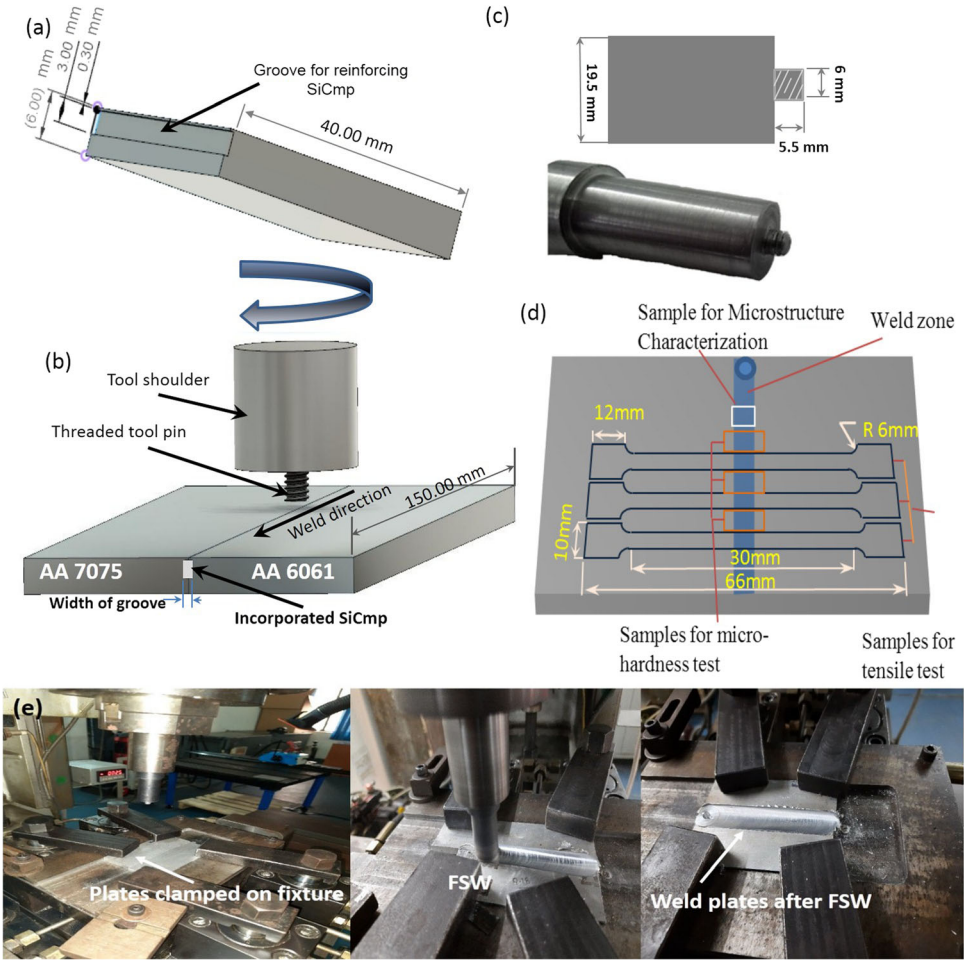


Fig. 1 (**a**) Groove dimensions for incorporation of SiC particles, (**b**) Representation of FSW setup, (**c**) Dimensions and image of FSW tool, (**d**) Dimensional features of various test samples, (**e**) Experimental set-up used for FSW

central composite design using response surface methodology, as tabulated in Table 1.

The welded specimens after implementation of FSW/SiCmp are depicted in Fig. 4. The samples for the tensile test, micro-hardness test and microstructural analysis were carved out by a wire-cut EDM machine for better dimensional accuracy, as depicted in Fig. 1d. The average result of three samples of tensile and micro-hardness tests, from each welded specimen was reported. The tensile samples employed for tensile testing were extruded as per the ASTM (E-8) standard (Fig. 1d). A Vickers micro-hardness tester (Model: XHVT-1000Z: Jinan testing equipment corp.) was employed to evaluate the hardness profile at 100 gm load for 15 s dwell time by taking 1 mm gap between the indentations. Prior to indentation, the samples were polished properly using emery papers of 600-1200 grade. The indentations were carried-out on the lateral surface at 2 mm below the top surface. Samples of dimensions of 15 × 10 × 6 mm were extruded transversely from the FSWed-joints for the microstructural examination. The examining surface of the samples was polished with emery paper 200-2000 grade. Super polishing was done by employing emery cloth comprising alumina paste. Afterward, all polished samples were etched for about 10 s with Keller's reagent (HCL: 8 ml + HNO_3: 6 ml + HF: 4 ml + H_2O: 82 ml) to investigate the microstructure and their allied characteristics using optical microscope (Model: Olympus GX-41) and scanning electron microscopy (SEM) (Model: Jeol JSM-IT 100) equipped with an EDS attachment (Model: Oxford Instrument- 51 ADD0076).

3 Results and Discussions

3.1 Microstructure observation

To investigate the influence of SiCmp on microstructural refinement, the study associated with grain structure and particle distribution in the weld zone at various processing conditions was carried out. The microstructural characterization of base materials AA7075 and AA6061 was depicted in Fig. 2 (a and c). The coarse grain structure of AA7075 and AA6061 was observed with a mean grain size of 47 and 36 μm,

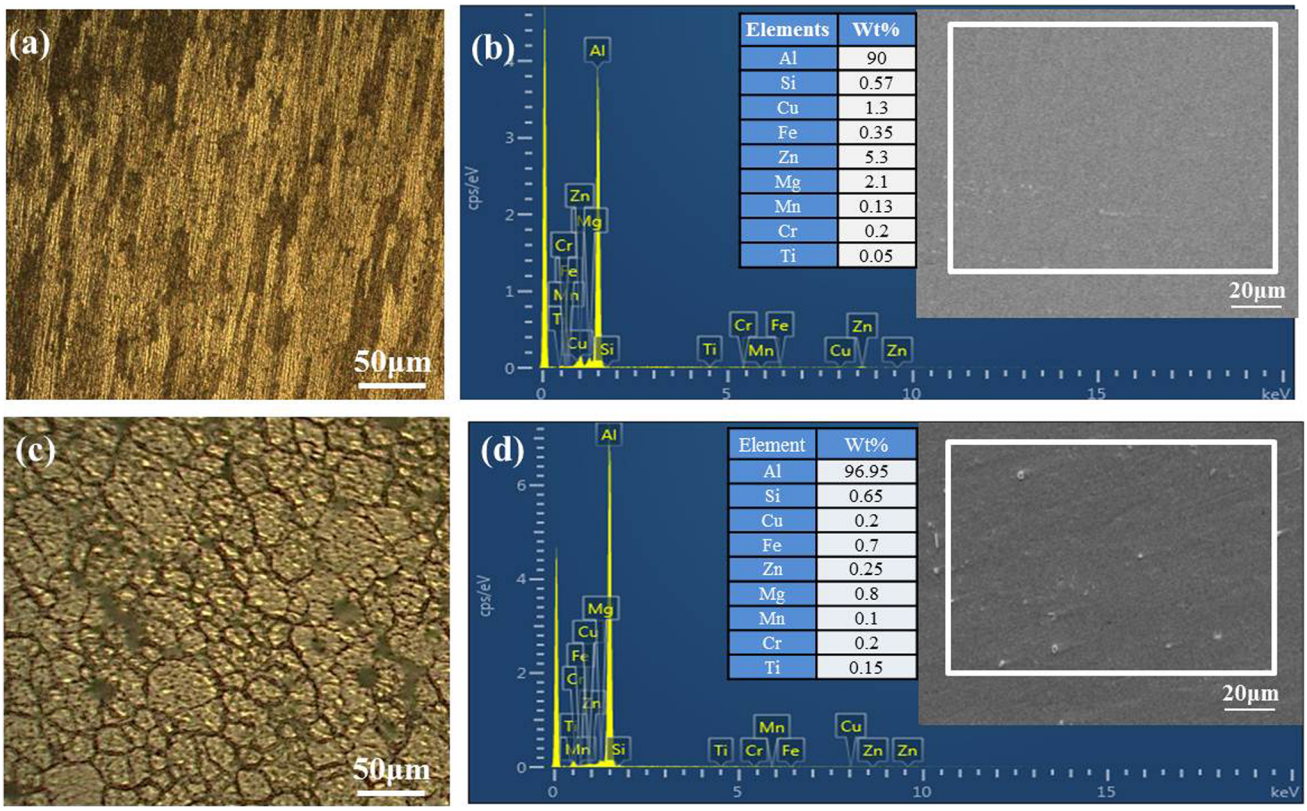


Fig. 2 Optical micrographs, SEM micrographs and EDS analysis of base metal (**a**, **b**) AA7075, (**c**, **d**) AA6061

respectively. The grain size was measured by employing ImageJ software. The grain structure of base materials was considerably refined in the processed region during FSW owing to severe plastic deformation caused by the stirring action of the tool, resulting in dynamic recrystallization (DRX) [34–36]. The processed region distinctly showed heat-affected zone (HAZ), thermo mechanical affected zone/partially recrystallization zone (TMAZ/PRZ) and shear zone/recrystallization zone (SZ/RZ) [37]. Such zones in the processed region of specimen no.- 10 were demonstrated. The grain structure refinement was noticed to be higher in the SZ than in the TMAZ and HAZ [38, 39]. Different SEM micrographs of the SZ of the FSWed joints processed under various combinations of process parameters are depicted in Fig. 5(a-d). Figure 5a demonstrates the agglomeration of SiCmp as well as poor bonding with aluminum metal matrix at RS of 700 rpm and TS of 60 mm/min with 10% SiCmp (specimen no.-4). Such agglomeration of particles represents the inadequate material mixing due to insufficient heat input at lower rotational speeds. At lower RS, enough heat is not produced for efficient material flow. As the material becomes highly viscous in the SZ, which causes unusual material mixing. Heat input is directly related to RS and TS values

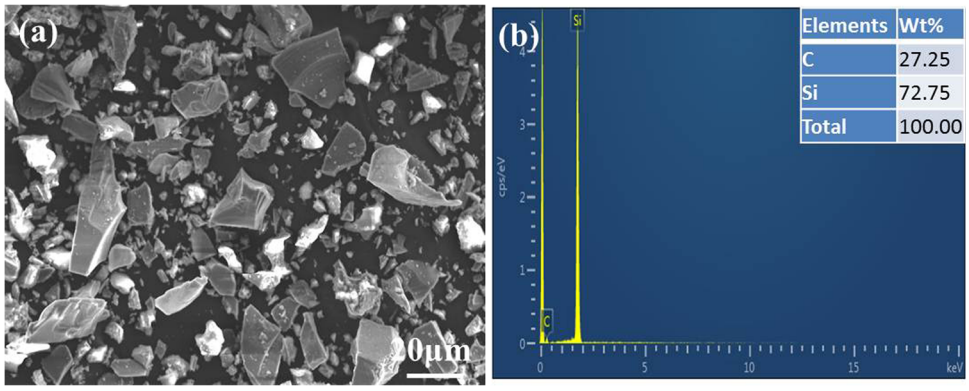


Fig. 3 SEM micrographs and EDS analysis of SiC particles

Table 1 Design of experiment

Specimens No.	Processing Parameters		
	A: Tool rotational speed (RS) (rpm)	B:Traverse speed (TS) (mm/min)	C: Volume fraction of SiCmp (%)
1	700	40	10
2	1100	50	7
3	900	40	7
4	700	60	10
5	900	50	10
6	1100	40	4
7	1100	60	10
8	900	60	7
9	900	50	7
10	1100	40	10
11	900	50	4
12	1100	60	4
13	700	60	4
14	700	40	4
15	700	50	7

during FSW. High RS and lower TS values generate high input [40]. Apart from higher heat generation at higher RS and lower TS, the stirring action and material flow were also enhanced and high plastic strain was produced in the SZ. This high plastic strain increases the stirring action and particles dispersion technique [41]. Therefore, enough heat is required for the effective material flow, which leads to uniform dispersion of reinforcement particles in the SZ by higher stirring action at higher RS. The significant dispersion of SiC particles within the plasticized SZ by enhanced stirring action at higher RS is reported by Mohammed and Subhi [42]. FSWed joint was produced at RS of 900 rpm and TS of 50 mm/min with 10% SiCmp (Specimen no.-5) characterized by rich and less-particle area (Fig. 5b). The tool's increased stirring action caused particles' fragmentation as well, although it deemed to be less significant. Hence, higher RS and lower TS values were believed to promote particles' mobility [43–45]. Homogeneous distribution of particles with significant fragmentation was observed, at the higher RS of 1100 rpm and lower TS of 40 mm/min, as depicted in Fig. 5(c, d). The absence of particles agglomeration indicates that the particles and the surrounding aluminum matrix are soundly bonded. In comparison to the FSWed joint with 4% SiCmp (Specimen no.-6) (Fig. 5c), the FSWed joint with 10% SiCmp (Specimen no.-10) was revealed to be denser and more homogeneously dispersed particles (Fig. 5d). Thus, the best dispersion of SiCmp was noticed at RS of 1100 rpm and TS of 40 mm/min with 10% SiCmp, in comparison with other parametric conditions, as illustrated in Fig. 5. It can be ascribed to the enough strain produced by tool's enhanced stirring action that improves the material mixing. The small sized SiCmp provide hindrance to the mobility of grain boundaries and restrict the grain growth via pinning effect, further reducing the grain size [46]. It is also obvious from that SiCmp are well bonded with the aluminum metal matrix Fig. 6

In comparison with base materials, reduced grain size was observed in the SZ after implementing FSW/SiCmp as depicted in optical micrographs (Fig. 7(a-d)). This can be ascribed to the, dynamic recrystallization and pinning effect of dispersed and fragmented SiCmp. Owing to unusual dispersion of SiCmp, the mean grain size of the SZ of specimen no.-4 and 5 (Fig. 7(a, b)) were found to be 14.8 μm and 10.3 μm, respectively. Whereas, the specimen no.-6 and 10 exhibited the mean grain size of 7.2 and 4.6 μm, respectively (Fig. 7(c,

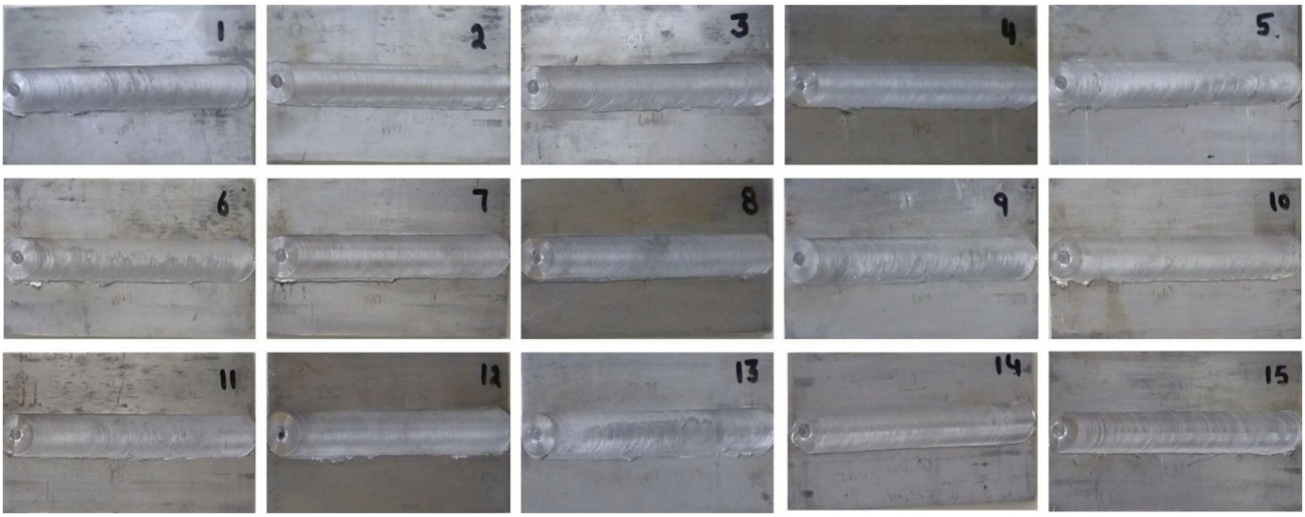


Fig. 4 Images of weld specimens after FSW/SiCmp

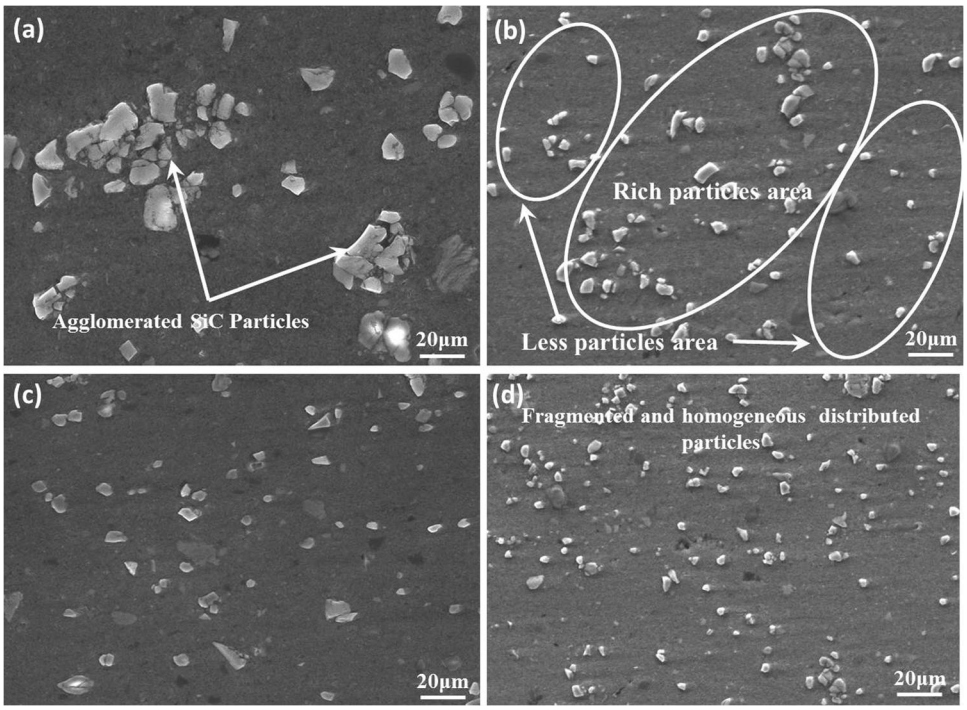


Fig. 5 SEM micrographs reveal the dispersion of SiCmp in the SZ: (**a**) at RS of 700 rpm and TS of 60 mm/min, with 10% SiCmp (Specimen no-4), (**b**) at RS of 900 rpm and TS of 50 mm/min with 10% SiCmp (Specimen no-5), (**c**, **d**) at RS of 1100 rpm and TS of 40 mm/min,(**c**) 4% SiCmp (Specimen no-6) (**d**) 10% SiCmp (Specimen no-10)

d)). The phenomenon of the reduced grain size in specimen no.-10 in comparison with specimen no.-6 can be ascribed to the incorporation of higher volume fraction of SiCmp, as the other welding parameters were constant. The grain size of the SZ was related with the Zener Holloman variable [47]. The mean grain size was determined by applying the following equation $R_c = \frac{4r}{3v_f}$, where R_c denotes critical radius. On comparing, specimen no.-10 and 6, it was noticed that incorporation of higher volume fraction of SiCmp during FSW enhanced the grain refinement. The growth rate of grains due to 2nd phase micro-particles is given by Eq. 1 [48].

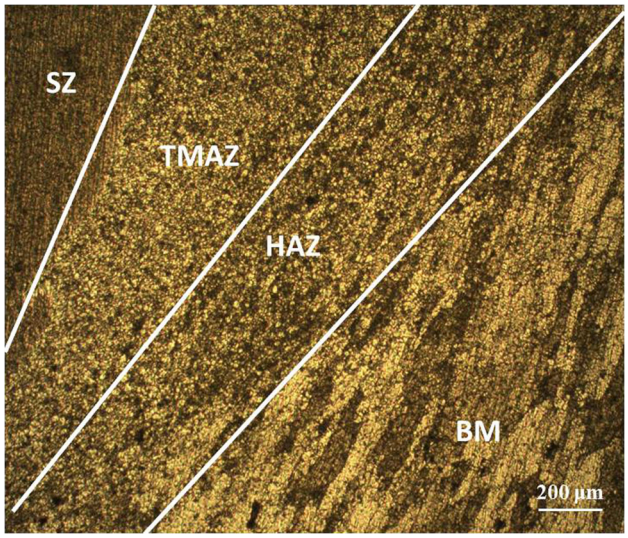


Fig. 6 Various Zones in the weld zone of specimen no.- 10

$$\frac{dR_g}{dt} = (P_d - P_{zp})M_b = \left(\frac{g_c\gamma_{be}}{R_g} - \frac{3V_f\gamma_{be}}{2r_p}\right) M_b \quad (1).$$

Where M_b denotes the mobility of grain boundaries, R_g denotes the grain's radius, V_f denotes volume fraction, P_d denotes driving pressure generated by grains boundary's curvature, P_{zp} represents Zener-pinning pressure, γ_{be} is the energy linked with grain boundary, g_c is geometrical constant, and r_p is the radius of 2nd phase particles. The growth of grains will end at $P_{zp}=P_d$.

$$\frac{g_c\gamma_{be}}{R_g} = \frac{3V_f\gamma_{be}}{2r_p} \quad (2)$$

The Zener's limiting size of grains may be observed ($g_c=1$), if the radius of curvature (Rg) equals the mean granularly radius (D), then $D_{zl} = \frac{4r_p}{3V_f}$.

Consequently, the specimen produced with higher RS, lower TS and higher volume fraction of SiCmp exhibits homogeneous dispersion of particles and reduced grain size in the SZ. These results revealed that the microstructural modification could be achieved by the incorporation of SiCmp in the SZ.

Figure 8(a-b) reveals the EDS results peaks of specimen no.-6 and specimen no.-10, produced at RS of 1100 rpm and TS of 40 mm/min but incorporated with different volume fraction of SiCmp, 4% and 10%, respectively. The EDS results revealed the existence of SiCmp in the SZ. The variation in volume fraction of SiCmp in the SZ is also indicated by EDS results. The

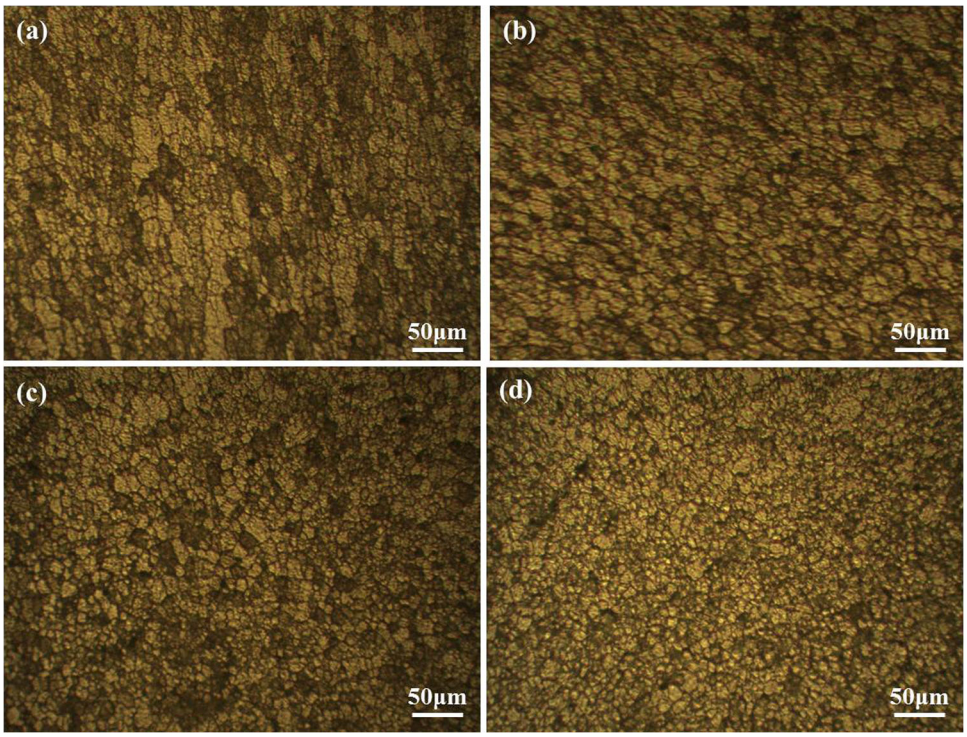


Fig. 7 Optical microstructure (OM) images at the SZ, (**a**) Specimen no.-4, (**b**) Specimen no.-5, (**c**) Specimen no.-6, (**d**) Specimen no.-10

elemental identification derived from the EDS analysis revealed the presence of silicon at the SZ of specimen no.-6 and 10 with 4.46 and 9.24%, respectively. As a consequence, the EDS data shows that SiCmp with different volume fractions were incorporated in the SZ.

3.2 Influence of SiCmp on mechanical Properties

3.2.1 Influence of SiCmp on tensile strength

The mechanical properties of the un-reinforced FSWed joints are mainly affected by the grain size. Whereas, the mechanical properties of the reinforced FSWed joints are dependent on other aspects, including the quality of bonding between reinforcement particles and matrix, size, and the dispersion pattern of reinforcement particles in the matrix [49, 50].

Table 2 summarizes the mechanical properties of all the specimens, including tensile strength, strain rate and micro-hardness. The tensile strength of base metals AA7075 and AA6061 was calculated as 496 and 288 MPa, respectively, with strain rate of 21.6% and 18.4%, respectively. The FSWed joint's tensile strength ranged from 180.50 MPa to 261.20 MPa due to different welding conditions. The stress-stain diagram of all the FSWed joints embedded with SiCmp during tensile loading was illustrated in Fig. 9. The tensile samples under tensile loading were fails at HAZ towards AA6061, where the hardness was minimal [51].

The tensile strength of composite joints is mainly influenced by grain refinement, pinning effect produced by SiCmp and bonding between SiCmp and metal-matrix. As a result of the higher grain refinement due to the pinning effect of the uniformly distributed SiCmp, specimen no.-10 exhibits the finer grains. Tensile strength is inversely proportional to grain's size, as defined by the Hall-patch relationship [52]. Consequently, specimen no.-10 exhibited the highest tensile strength of 261.20 MPa, which is approximately 91% of the tensile strength of the as received AA6061. The imparted SiCmp in the SZ during FSW yielding the increased in the tensile strength as reported [53–55]. This can also be ascribed to the higher volume fraction of SiCmp and strong particles-matrix interfacial bonding at higher RS of 1100 rpm. The incorporated SiCmp impedes the motion of dislocations, which prevents the grain growth, resulting in higher grain refinement and causing a higher strength of material [56]. Referring to Table 2, the incorporation of SiCmp were positively enhanced the tensile strength and strain rate of FSWed joints.

3.2.2 Influence of SiCmp on micro-hardness

The mean indentation micro-hardness of the base metals AA6061 and AA7075 was observed as, 96 and 170 HV, respectively. The maximum hardness value along the

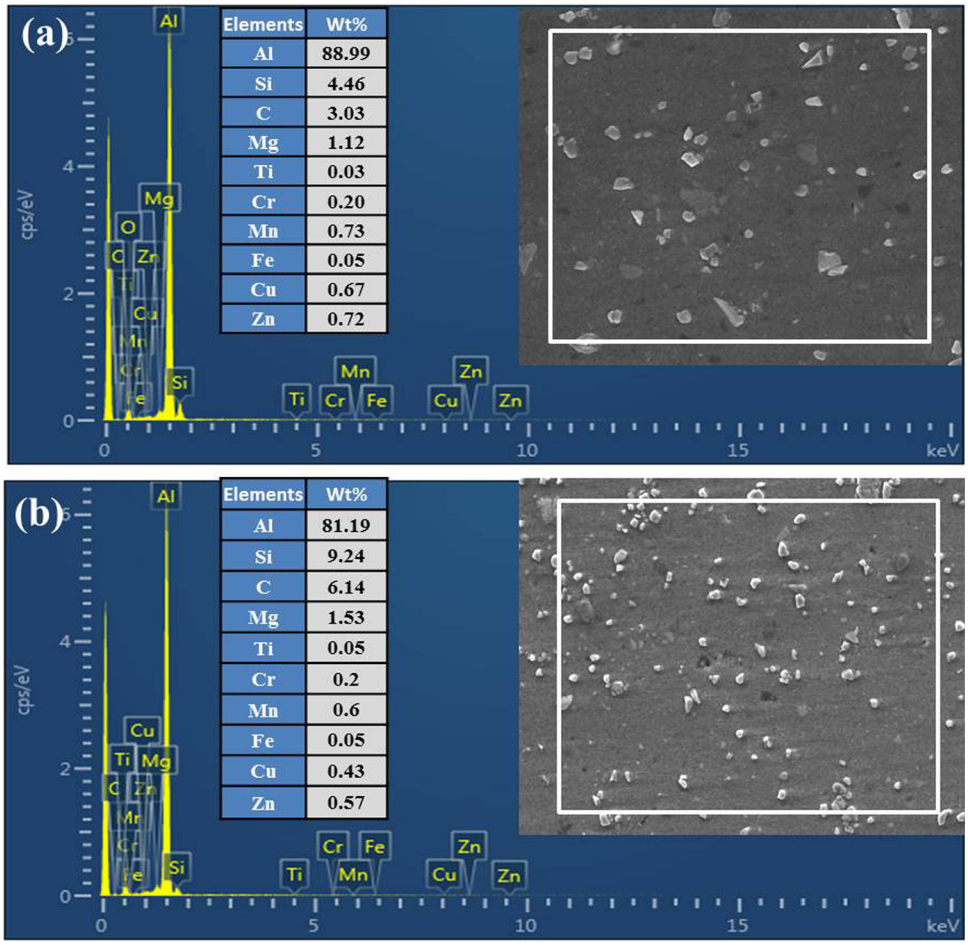


Fig. 8 EDS peaks at the SZ, (**a**) Specimen no.- 6, (**b**) Specimen no.- 10

centerline of the weld zone of specimen no.-10 was observed to be 186.83 HV at RS of 1100 rpm and TS of 40 mm/min with 10% SiCmp. Figure 10 depicts, the variation of micro-hardness of the FSWed joints embedded with SiCmp to the distance away from the center of the weld zone. The micro-hardness range of each zone is tabulated in Table 3. Grain refinement, volume fraction and reinforcement particles dispersion in the SZ, are all the factors that contribute to the hardness of the SZ [57]. A low tool traverse speed increases the heat generation during the friction stir welding/processing, which has a direct impact on the lower hardness value in the TMAZ. The same results were found in the previous studies [58–61]. The maximum micro-hardness of 186.83 HV was obtained at the center of weld joint of specimen no.-10, which is approximately 10% higher than the hardness of as received AA7075 base material. Due to the dispersion of SiCmp and the intermixing of base metal and SiCmp, a low variation in hardness was observed in the TMAZ of all FSW/SiCmp specimens. The strengthening effect of SiCmp in specimen no.-10 was greater compared to other weld specimens, because of the hard nature of the SiCmp, grain refinement and material intermixing rule.

The precipitation hardening was influenced by the size, shape as well as the dispersion of strengthening precipitates [62]. Due to smaller grains, the micro-hardness of the SZ was found to be higher than that of TMAZ and HAZ. Because of the dissolution of precipitates, the micro-hardness of the TMAZ was reduced. The HAZ's micro-hardness was diminished when the precipitates coarsened and the Guinier- Preston zone vanished [63, 64]. As a consequence, the combined impacts of grain refining, high SiCmp hardness, and dispersion strengthening can be ascribed to the improved micro-hardness of the SZ.

3.2.3 Fractured Surface Morphology

Figure 11(a-d) depicts the fracture surfaces of the FSW joints reinforced with different volume fractions of SiCmp under different processing conditions. The fractured surface was examined using scanning electron microscopy to

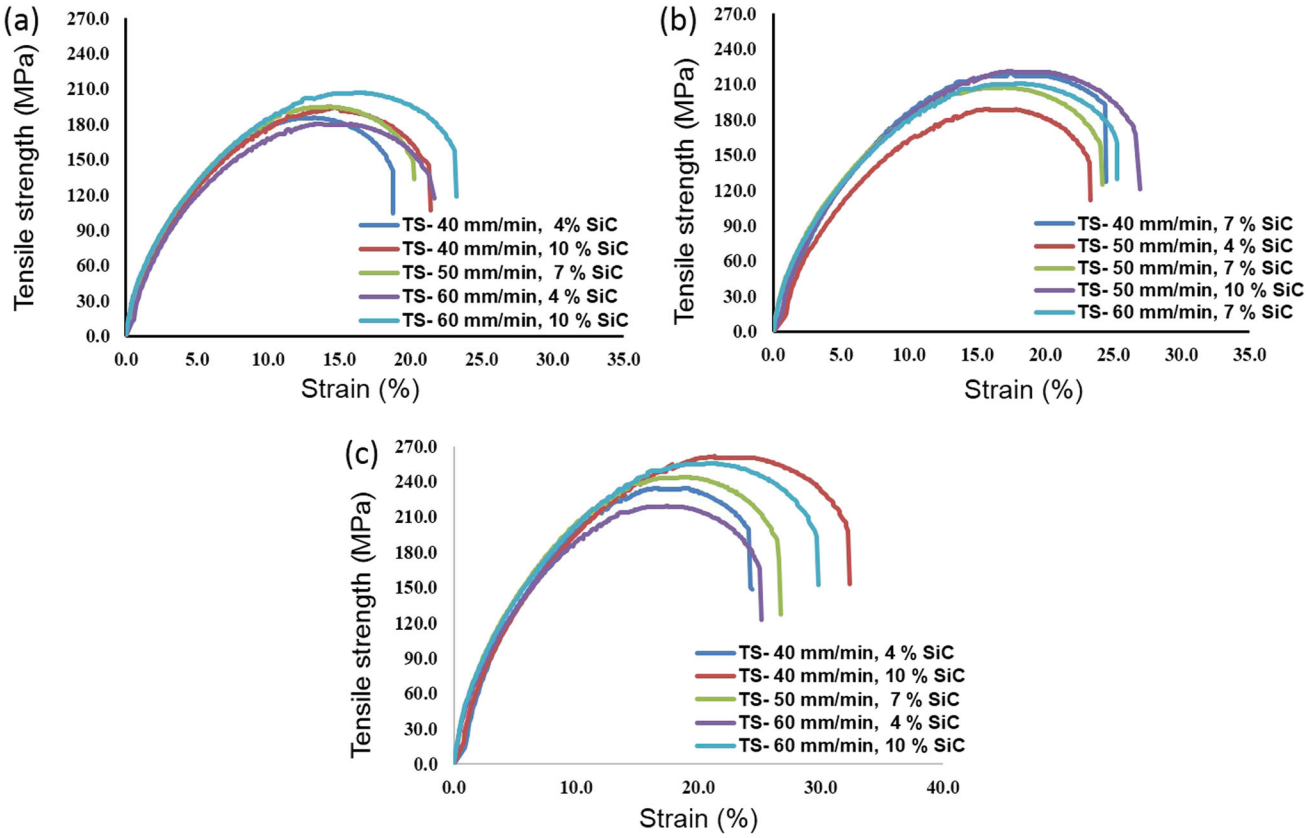


Fig. 9 Stress strain diagram of various FSWed joints, (**a**) RS- 700 rpm, (**b**) RS- 900 rpm, (**c**) RS- 1100 rpm

Table 2 Mean tensile strength, strain rate and micro-hardness at the center line of FSWed joint

FSWed specimen no.	Processing Parameters			Mean tensile Strength (MPa)	Joint's efficiency (%)	Mean strain rate (%)	Mean micro-hardness (HV)
	A: Tool rotational speed (RS) (rpm)	B:Traverse speed (TS) (mm/min)	C: Volume fraction of SiCmp (%)				
1	700	40	10	194.3	67.47	21.7	138.96
2	1100	50	7	243.6	84.58	26.7	167.78
3	900	40	7	219.4	76.18	24.3	151.94
4	700	60	10	205.3	71.28	23.2	141.24
5	900	50	10	221.3	76.84	26.8	150.94
6	1100	40	4	234.5	81.42	25.2	168.64
7	1100	60	10	254.9	88.51	29.8	181.27
8	900	60	7	210.7	73.16	25.2	146.59
9	900	50	7	207.4	72.01	24.2	144.45
10	1100	40	10	261.2	90.69	32.4	186.83
11	900	50	4	188.3	65.38	23.2	130.54
12	1100	60	4	219.7	76.28	25.1	153.94
13	700	60	4	180.5	62.67	21.8	130.05
14	700	40	4	185.5	64.41	18.8	132.40
15	700	50	7	195.3	67.81	20.2	136.61

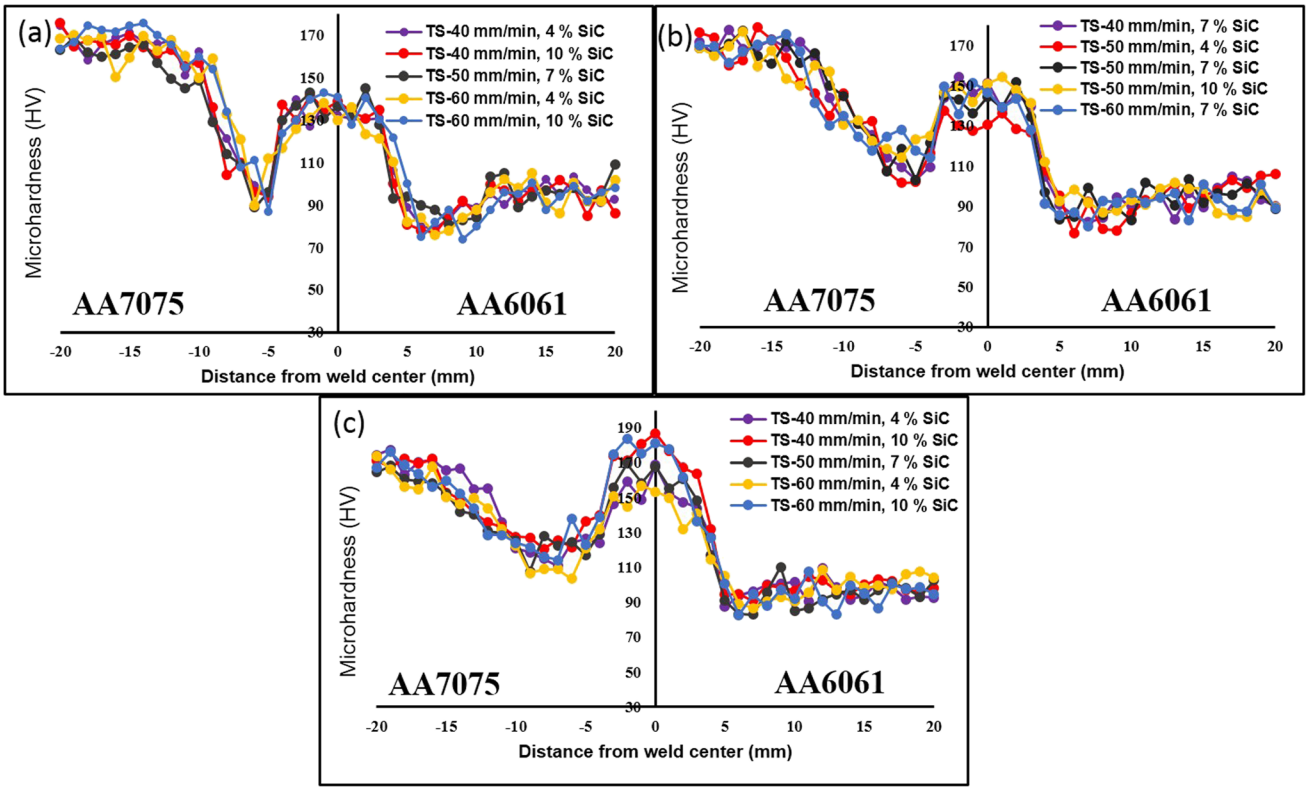


Fig. 10 micro-hardness variation of various FSWed joints, (**a**) RS- 700 rpm, (**b**) RS- 900 rpm, (**c**) RS- 1100 rpm

better understand the mode of failure of the composite FSW joint under tensile loading. After tensile testing, the fractured surfaces of all the samples showed the dimple fracture, indicating the failure of the ductile mode. The probability of strengthening in any composite matrix depends upon the transferring of stress from the base material to the reinforcement particles. Due to poor interfacial contact, the interface could fail before transmitting stress to micro-particles, resulting in no strengthening [65]. During the tensile loading, the cup-cone shaped shear plane was produced along the periphery of the test samples [66–68].

In comparison with specimen no.-4, 5 and 6 (Fig. 11(a-c)), the morphology of the fractured surface of specimen no. 10 revealed tiny and equiaxed dimples (Fig. 11d). Honeycomb dimples with a ductile fracture were observed on the fractured surface of specimen no.-4, as depicted in Fig. 11a. All the FSWed specimens were produced at the different combinations of RS and TS. As a result of effective material mixing, the homogeneous dispersion of SiCmp diminished the grain size, resulting in fine and equiaxed dimples [69, 70]. Leading to enhanced ductility, the fractured metal-matrix was found at HAZ (Fig. 12), which was validated by the variation of micro-hardness. According to the analysis of the fractured surfaces of the tensile samples, the incorporation of SiCmp significantly improved the tensile strength of FSWed joints.

4 Conclusions

In current research, the fabrication of a defect-free weld joint of AA7075/AA6061 by incorporating SiCmp was successfully carried out by FSW. The mechanical properties of the FSWed joints embedded with SiCmp were mainly inflenced due to dynamic recrystallization and the pinning effect of homogeneously dispersed SiCmp.

Table 3 Micro-hardness range of each zone

Position	AA6061	HAZ	TMAZ	SZ	TMAZ	HAZ	AA7075
Hardness range (HV)	87-92	71-80	78-86	130-186	116-138	96-111	167-175

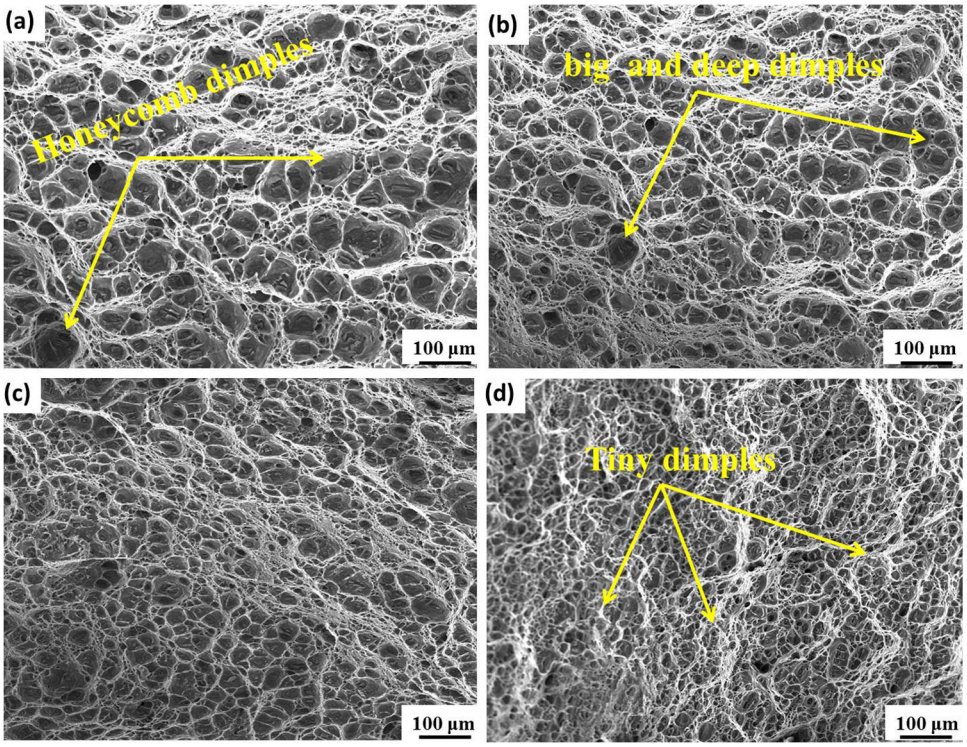


Fig. 11 Fractured surface morphology ,(**a**) Specimen no.-4, (**b**) Specimen no.-5, (**c**) Specimen no.-6, (**d**) Specimen no.-10

- Agglomeration of SiCmp as well as poor bonding of SiCmp with surrounded aluminum matrix was observed in the SZ at the RS of 700 rpm and TS of 60 mm/min.
- Homogeneous dispersion as well as adequate fragmentation of particles was observed at RS of 1100 rpm and TS of 40 mm/min. Therefore, the FSWed joint produced at RS of 1100 rpm and TS of 40 mm/min with a 10% volume fraction of SiCmp, exhibited the higher mechanical properties in comparison with other joints, due to microstructure refinement via pinning effect.



Fig. 12 Fractured samples after tensile test

- The highest tensile strength was found to be 261.20 MPa, which is approximately 91% of the tensile strength of the as received AA6061.
- Furthermore, the maximum micro-hardness of 186.83 HV was observed, which is approximately 10% higher than the base metal AA7075.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12633-022-01716-5.

Acknowledgements The authors would like to express heartfelt thanks to Delhi Technological University, Delhi, India for providing the experimental setup and their corresponding labs for supporting this work.

Author contributions: Sumit Jain designed and performed the experiments, analyzed the results, prepared the figures, and wrote the manuscript.

R. S. Mishra supervised the entire research, coordinated with author and analyzed the results.

Funding statement: Not applicable.

Availability of data and material: Not applicable.

Declarations

Conflict of interest Authors declare no conflict of interest.

Compliance with ethical standards The authors have declared that no competing interests exist.

Consent to participate: Not applicable.

Consent for Publication: Not applicable.

References

1. Mishra RS, Ma ZY (2005) Friction stir welding and processing. Material Science and Engineering R 50:1–78
2. Mehdi H, Mishra RS (2021) Microstructureand mechanical characterization of TIG-welded joint of AA6061 and AA7075 byfriction stir processing. Proc Inst Mech Eng L P I Mech Eng L-J Mater Des Appl 11:2531-2546
3. Mehdi H, Mishra RS (2019) Study of the influence of friction stir processing on tungsten inert gas welding of different aluminum alloy. SN Appl Sci 1:712
4. Gibson BT, Lammlein DH, Prater TJ et al (2014) Friction stir welding: process, automation, and control. J Manuf Process 16:56–73
5. Meng X, Huang Y, Cao J, et al (2021) Recent progress on control strategies for inherent issues in friction stir welding. Progress Mater Sci 115:100706.
6. Raja S, Muhamad MR, Jamaludin MF et al (2020) A review on nanomaterials reinforcement in friction stir welding. Journal of Materials Research and Technology 6:16459–16487
7. Sharma N, Khan ZA, Siddiquee AN (2017) Friction stir welding of aluminum to copper–An overview [J]. Trans Nonferrous MetSoc China 27(10):2113–2136
8. Bakhtiari Argesi F, Shamsipur A, Mirsalehi SE (2018) Dissimilar Joining of Pure Copper to Aluminum Alloy via Friction Stir Welding. Acta Metall Sin (Engl Lett) 31:1183–1196
9. Wang T, Komarasamy M, Liu K, Mishra RS (2018) Friction stir butt welding of strain-hardened aluminum alloy with high strength steel.Mater Sci Eng, A 737:85–89
10. Tan S, Zheng F, Chen J, Han J, Wu Y, Peng L (2017) Effects of process parameters on microstructure and mechanical properties of friction stir lap linear welded 6061 aluminum alloy to NZ30K magnesium alloy. J Magnes Alloy 5:56–63
11. Hull D, Clyne T (1996) An Introduction to Composite Materials Cambridge University Press, Great Britain
12. Gay D (2014) Composite Materials: Design and Applications CRC Press, Boca Raton
13. Chawla KK (2012) Composite Materials: Science and Engineering, Springer Science & Business Media, Heidelberg
14. Saini N, Pandey C, Dwivedi DK (2018) Ductilizing of cast hypereutectic Al–17%Si alloy by friction stir processing. Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering. 232(6):696-701
15. Essam B, Moustafa A, Melaibari M, Basha Wear and microhardness behaviors of AA7075/SiC-BN hybrid nanocomposite surfaces fabricated by friction stir processing. Ceram Int 46(10 Part B): 16938–16943
16. Saini N, Pandey C, Thapliyal S et al (2018) Mechanical Properties and Wear Behavior of Zn and MoS2 Reinforced Surface Composite Al- Si Alloys Using Friction Stir Processing. Silicon 10:1979–1990
17. Rejil CM, Dinaharan I, Vijay SJ, Murugan N (2012) Microstructure and sliding wear behavior of AA6360/(TiC+B_4C) hybrid surface composite layer synthesized by friction stir processing on aluminum substrate. Mater Sci Eng A 552:336–344
18. Liu F, Li Y, Sun Z, Ji Y (2021) Corrosion resistance and tribological behavior of particles reinforced AZ31 magnesium matrix composites developed by friction stir processing. J Mater Res Technol 11:1019-1030
19. Kaczmar J, Pietrzak K, Wlosinski W, J. Mater. Process. Technol, 106 (1) (2000) 58–67.
20. Singh T, Tiwari SK, Shukla DK (2019) Friction-stir welding of AA6061-T6: The effects of Al2O3 nanoparticles addition. Results Mater 1: 100005
21. Saeidi M, Barmouz M, Givi MKB (2015) Investigation on AA5083/AA7075+Al. Joint fabricated by friction stir welding: characterizing microstructure corrosion and toughness behavior. Mater Res 6:1156–1162
22. Asl NS, Mirsalehi SE, Dehghani K (2019) Effect of TiO2 nanoparticles addition on microstructure and mechanical properties of dissimilar friction stir welded AA6063-T4 aluminum alloy and AZ31B-O magnesium alloy. J Manuf Process 38:338–354
23. Rathee S, Maheshwari S, Siddiquee AN, Srivastava M (2017) A review of recent progress in solid state fabrication of composites and functionally graded systems via friction stir processing. Crit Rev Solid State Mater Sci 43 (4): 334-366
24. Srivastava M, Rathee SA (2021) Study on the Effect of Incorporation of SiC Particles during Friction Stir Welding of Al 5059 Alloy. Silicon 13:2209–2219
25. Kumar KS, Murigendrappa SM, kumar H (2019) Experimental investigation on effects of varying volume fractions of SiC nanoparticle reinforcement on microstructure and mechanical properties in friction stir welded dissimilar joints of AA2024- T351 and AA7075-T651. J Mater Res 34: 1229–1247
26. Ashok Kumar R, Ragav GR, Nagarajan KJ et al (2019) Effect of hybrid reinforcement at stirred zone of dissimilar aluminum alloys during friction stir welding. Metall Res Technol 116:631
27. Bahemmat P, Haghpanahi M, Besharati MK et al (2010) Study on mechanical , micro- , and macrostructural characteristics of dissimilar friction stir welding of AA6061-T6 and AA7075-T6. Proc Inst Mech Eng Part B J Eng Manuf 224(12):1854–1865
28. Çömez N, Durmuş H (2019) "Cold Metal Transfer Welding of AA6061 to AA7075: Mechanical Properties and Corrosion." ASME. J. Eng. Mater. Technol 141(3): 031002
29. Nandan R, DebRoy T, Bhadeshia HKDH (2008) Recent advances in friction-stir welding – process, weldment structure and properties. Prog Mater Sci 53:980–1023
30. Kumar K, Kailas SV (2010) Positional dependence of material flow in friction stir welding: analysis of joint line remnant and its relevance to dissimilar metal welding. Sci Technol Weld Joining 15: 305–311
31. Aval HJ, Serajzadeh S, Kokabi AH (2011) Thermo-mechanical and microstructural issues in dissimilar friction stir welding of AA5086–AA6061. J Mater Sci 46:3258–3268
32. Park S-K, Hong S-T, Park J-H, Park K-Y, Kwon Y-J, Son H-J (2010) Effect of material locations on properties of friction stir welding joints of dissimilar aluminium alloys. Sci Technol Weld Joining 15:331–336
33. Guo JF, Chen HC, Sun CN et al (2014) Friction stir welding of dissimilar materials between AA6061 and AA7075 Al alloys effects of process parameters. Mater Design 56:185–192
34. Mehdi H, Mishra RS (2021) Effect of Friction Stir Processing on Mechanical Properties and Wear Resistance of Tungsten Inert Gas Welded Joint of Dissimilar Aluminum Alloys. Journal of Material Engineering and Performance 30:1926–1937
35. Ghangas G, Singhal S (2018) Effect of tool pin profile and dimensions on mechanical properties and microstructure of friction stir welded armor alloy. Materials Research Express 5:066555
36. Sharma VM, Sharma S, Mewar SK et al (2018) Friction stir processing of Al6061-SiC-graphite hybrid surface composites. Mater Manuf Process 33:795–804
37. Ghangas G, Singhal S (2018) Investigations of multi-pass friction stir welding for Al-Zn-Mg alloy. Mater Today Proc 5 (9): 17107-17113
38. Mehdi H, Mishra RS (2020) Effect of Friction Stir Processing on Microstructure and Mechanical Properties of TIG Welded Joint of

AA6061 and AA7075. Metallography Microstructure and Analysis 9:403–418
39. Mehdi H, Mishra RS (2021) Consequence of reinforced SiC particles on microstructural and mechanical properties of AA6061 surface composites by multi-pass FSP. J Adhes Sci Technol. https://doi.org/10.1080/01694243.2021.1964846
40. Cartigueyen S, Mahadevan K (2016) Effects of heat input and particulate deposition on Cu/SiCp surface composite processed by friction stir processing. Indian J Eng Mater Sci (IJEMS) 23(2&3): 145-151.
41. Acharya U, Roy BS, Saha SC Effect of tool rotational speed on the particle distribution in friction stir welding of AA6092/17.5 SiCp-T6 composite plates and its consequences on the mechanical property of the joint. Def Technol 16(2):381-391
42. Mohammed MH, Subhi AD (2021) Exploring the influence of process parameters on the properties of SiC/A380 Al alloy surface composite fabricated by friction stir processing. Eng Sci Technol Int J 24(5):2215–0986
43. Patel V, Li W, Xu Y (2019) Stationary shoulder tool in friction stir processing: a novel low heat input tooling system for magnesium alloy. Mater Manuf Process 34:177–182
44. Palanivel R, Dinaharan I, Laubscher RF et al (2016) Influence of boron nitride nanoparticles on microstructure and wear behavior of AA6082/TiB2 hybrid aluminum composites synthesized by friction stir processing. Mater Des 106:195–204
45. Salih OS, Ou H, Sun W, McCartney DG (2015) A review of friction stir welding of aluminum matrix composites. Mater Des 86:61–71
46. Pandey C, Mahapatra MM, Kumar P, Daniel F, Adhithan B (2019) Softening mechanism of P91 steel weldments using heat treatments. Arch Civil Mech Eng 19(2):297–310
47. Srivastava M, Rathee S, Siddiquee AN et al (2019) Investigation on the Effects of Silicon Carbide and Cooling Medium during Multi-Pass FSP of Al-Mg/ SiC Surface Composites. Silicon 11:2149–2157
48. Chang CI, Lee CJ, Huang JC (2004) Relationship between grain size and Zener–Holloman parameter during friction stir processing in AZ31 Mg alloys. Scripta Mater 51:509–514
49. Humphreys FJ, Hatherly M (2004) Chapter 11 - grain growth following recrystallization, in: Recrystallization and Related Annealing Phenomena.Elsevier Ltd., 333–378
50. Ghangas G, Kundu J, Rattan N, Sharma SK (2017) Friction Stir Welding: Merits over other Joining Processes. Int J Curr Eng Technol 7(3):1170–1174
51. Zhang Z, Chen DL (2006) Consideration of Orowan strengthening effect in particulate-reinforced metal matrix nanocomposites: A model for predicting their yield strength Author links open overlay panel. Scripta Mater 54:1321–1326
52. Mehdi H, Mishra RS (2020) Influence of friction stir processing on weld temperature distribution and mechanical properties of TIG welded joint of AA6061 and AA7075. Trans Indian Inst Met 73: 1773–1788
53. Jamalian HM, Ramezani H, Ghobadi H, Ansari M, Yari S (2016) Givi MKB Processing–structure–property correlation in nano-SiC-reinforced friction stir welded aluminum joints. J Manuf Process 21:180-189
54. Khodabakhshi F, Gerlich AP (2018) Potentials and strategies of solid-state additive friction-stir manufacturing technology: a critical review. J Manuf Process 36:77–92
55. Sharma V, Prakash U, Kumar BVM (2015) Surface composites by friction stir processing: a review. J Mater Process Technol 224:117–134
56. Sharma DK, Patel V, Badheka V et al (2019) Fabrication of hybrid surface composites AA6061/(B 4 C+MoS) via friction stir processing. J Tribology 141:52201–52210
57. Ebrahimzadeh V, Paidar M, Safarkhanian MA et al (2018) Orbital friction stir lap welding of AA5456-H321/AA5456-O aluminum alloys under varied parameters. Int J Adv Manuf Technol 96: 1237–1254
58. Xue P, Xie GM, Xiao BL, Ma ZY, Geng L (2010) Effect of heat input conditions on micro-structure and mechanical properties of friction-stir-welded pure copper. Metall Mater Trans A 41:2010–2021
59. Mehdi H, Mishra RS (2021) Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075. Def Technol 17:715–727
60. Khodaverdizadeh H, Mahmoudi A, Heidarzadeh A, Nazari E (2012) Effect of friction stir welding (FSW) parameters on strain hardening behaviour of pure copper joints. Mater Des 35:330–334
61. Husain Mehdi RS, Mishra (2016) Mechanical properties and microstructure studies in Friction Stir Welding (FSW) joints of dissimilar alloy – a review. J Achiev Mater Manuf Eng 77(1):31–40
62. Fuller CB, Mahoney MW, Calabrese M et al (2010) Evolution of microstructure and mechanical properties in naturally aged 7050 and 7075 Al friction stir welds. Mater Sci Eng A 527(9):2233–2240
63. Azimzadegan T, Serajzadeh S (2010) An Investigation into microstructures and mechanical properties of AA7075-T6 during friction stir welding at relatively high rotational speeds. J Mat Eng Perf 19: 1256–1263
64. Husain Mehdi RS Mishra. An experimental analysis and optimization of process parameters of AA6061 and AA7075 welded joint by TIG+FSP welding using RSM. Adv Mater Process Technol. https://doi.org/10.1080/2374068X.2020.1829952
65. Lloyd DJ (1994) Particle reinforced aluminum and magnesium matrix composites. Int Mater Rev 39:1
66. Pandey C, Saini N, Mahapatra MM et al (2017) Study of the fracture surface morphology of impact and tensile tested cast and forged (C&F) Grade 91 steel at room temperature for different heat treatment regimes. Eng Fail Analy 71:131–147
67. Pandey C, Saini N, Mahapatra MM, Kumar P (2017) Study of the fracture surface morphology of impact and tensile tested cast and forged (C&F) Grade 91 steel at room temperature for different heat treatment regimes. Eng Fail Anal 71:131–147
68. Pandey C, Mahapatra MM, Kumar P, Kumar P, Saini N, Thakare JG, Kumar S (2019) Study on effect of double austenitization treatment on fracture morphology tensile tested nuclear grade P92 steel. Eng Fail Anal 96:158–167
69. Heidarzadeh A, Barenji RV, Esmaily M et al (2015) Tensile Properties of Friction Stir Welds of AA 7020 Aluminum Alloy. Trans Indian Inst Met 68:757–767
70. Liu HJ, Fujii H, Maeda M et al (2013) Tensile properties and fracture locations of friction-stir-welded joints of 2017-T351 aluminum alloy. J Mat Proc Technol 142: 692–696

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

3 Biotech (2022) 12:92
https://doi.org/10.1007/s13205-022-03152-z

REVIEW ARTICLE

Natural protein-based electrospun nanofibers for advanced healthcare applications: progress and challenges

Anushka Agarwal[1] · Gyaneshwar K. Rao[1] · Sudip Majumder[1] · Manish Shandilya[1] · Varun Rawat[1] · Roli Purwar[2] · Monu Verma[3] · Chandra Mohan Srivastava[1,4]

Received: 17 March 2021 / Accepted: 16 February 2022
© King Abdulaziz City for Science and Technology 2022

Abstract

Electrospinning is an electrostatic fiber fabrication technique that operates by the application of a strong electric field on polymer solution or melts. It is used to fabricate fibers whose size lies in the range of few microns to the nanometer range. Historic development of electrospinning has evinced attention due to its outstanding attributes such as small diameter, excellent pore inter-connectivity, high porosity, and high surface-to-volume ratio. This review aims to highlight the theory behind electrospinning and the machine setup with a detailed discussion about the processing parameters. It discusses the latest innovations in natural protein-based electrospun nanofibers for health care applications. Various plant- and animal-based proteins have been discussed with detailed sample preparation and corresponding processing parameters. The usage of these electrospun nanofibers in regenerative medicine and drug delivery has also been discussed. Some technical innovations in electrospinning techniques such as emulsion electrospinning and coaxial electrospinning have been highlighted. Coaxial electrospun core–shell nanofibers have the potential to be utilized as an advanced nano-architecture for sustained release targeted delivery as well as for regenerative medicine. Healthcare applications of nanofibers formed via emulsion and coaxial electrospinning have been discussed briefly. Electrospun nanofibers have still much scope for commercialization on large scale. Some of the available wound-dressing materials have been discussed in brief.

Keywords Proteins · Electrospinning · Nanofibers · Nanocomposites · Healthcare applications

Introduction

Electrospinning is a voltage-driven spinning process that utilizes polymer solution or melts to fabricate fibers range from few microns to the nanoscale (Tucker et al. 2012). Electrospinning, which generates nanoscale fibers, has gained enormous popularity over the other methods due to its low cost and easy to operate feature. Some technological advancements have also been taken place to fabricate two-dimensional (2D) (Yang et al. 2007a, b) and three-dimensional (3D) electrospun fibers (Yuan et al. 2012a, b). The 2D fibers are found to be present in either random sheet or aligned sheet form whereas three-dimensional fibers are found in form of bundles. The 3D fibers demonstrate greater mechanical and tensile strength over two-dimensional fibers (Bosworth et al. 2013). The 2D and 3D fibers have been used in various fields such as energy harvesting, filtration protective materials, sensors, storage, and biomedical to produce synthetic grafts (Sun et al. 2014).

The development of electrospinning is old and has a vast history. William Gilbert in 1600 observed the electrostatic attraction of a liquid. CV Boys described the process of electrospinning in the 1890s. However, the first patent in electrospinning was filed by John Francis Cooley in 1900 that led to the use of electrospinning in industries. In 1934, Anton Formhals filed his patent for the process and apparatus of electrospinning in which he described the synthesis of

✉ Chandra Mohan Srivastava
cmsrivastava@ggn.amity.edu

1 Department of Chemistry, Biochemistry and Forensic Science, Amity School of Applied Sciences, Amity University Haryana, Gurugram 122413, India

2 Department of Applied Chemistry, Delhi Technological University, New Delhi, Delhi 110042, India

3 Department of Environmental Engineering, University of Seoul, Seoul 130743, South Korea

4 Centre for Polymer Technology, Amity School of Applied Sciences, Amity University Haryana, Gurugram 122413, India

Published online: 14 March 2022

cellulose acetate nanofibers. The spinning setup developed by Formhals failed as the tip-to-collector distance was not appropriate which resulted in a less integrated web structure attributed to the fact that the fibers could not be dried properly before reaching the collector plate. Formhals worked on the drawbacks and made a great contribution in this field by filing more than 22 patents (Subbiah et al. 2005). In older days, electrospinning was only applicable to soluble polymers which was a major drawback. Most of the synthetic polymers do not easily dissolve in organic solvents. Hence, it was very difficult to electrospun such kinds of high-performance synthetic polymers such as polyetheretherketone (PEEK).

In 1936, Charles Norton filed his first patent on melt electrospinning. In this technique, he electrified the polymer melt to prepare nanofibers. The electrified molten jet formed in melt electrospinning has a more predictable path which resulted in the deposition on the collector accurately. In 1969, Sir Geoffrey Ingram Taylor proposed the mathematical aspects (Eq. 1) of the shape of the cone formed by polymer solution as it comes out of the syringe:

$$V_C^2 = 4\left(\frac{H}{h}\right)^2 (\ln(-1.5) - 1.5)(1.3\,\pi R\gamma)(0.09), \quad (1)$$

where V_c is the critical voltage, H is the tip to the collector distance (in cm), R is the inner radius of the spinneret (in cm), h is the length of the liquid column (in cm), and g is the surface tension of the spinning solution (in dyne/cm). The factor 0.09 is inserted to predict the voltage (Taylor 1969).

Setup

The basic electrospinning setup includes a syringe pump with a capillary, a conducting collector, and a high voltage source. The syringe pump takes the solution from the syringe to the spinneret. The solution forms a cone shape during ejection from the syringe known as the Taylor cone. Taylor cone is electrified by applying high voltage. The Taylor cone deforms and the solution proceeds towards the collector when the electrostatic repulsion exceeds the surface tension. As the jet leaves from the needle tip, stretching of the jet takes place during the flight. This results in evaporation of the solvent and thinning of the jet up to the nano-range. These nanofibers were finally deposited at the collector plate (Bonadies 2019) (Fig. 1).

Several parameters together decide the morphology and properties of electrospun fibers. These include:

a) *Flow rate:* The flow rate controls the volume of solution suspended at the tip of the spinneret, which manages the shape of Taylor Cone, a crucial step of this process. The formation of fiber does not take place when the feed rate is slow. This is due to the unsuccessful formation of the Taylor cone. Increasing the feed rate results in the production of fibers. However, if the feed rate is substantially higher, beads would be formed due to insufficient time available for solvent evaporation. The feed rate should be intermediate to result in the formation of uniform fibers (Yuan et al. 2004; Zuo et al. 2005).
b) *Concentration*: At a very low concentration, jet breaks during the flight before reaching the collector due to insufficient molecular polymer entanglement. This results in the formation of spheres or beads rather than nanofiber. (McKee et al. 2004) The fiber diameter increases with the increase in concentration. (Casasola et al. 2014) At very high concentrations, the polymer solution becomes viscous. This results in the difficult ejection of the polymer from the needle and the formation of ribbon-like fibers take place (Yang et al. 2004).
c) *Voltage:* Voltage plays a significantly important role in electrifying Taylor Cone. Electrification of Taylor cone does not take place at low voltage and jet forma-

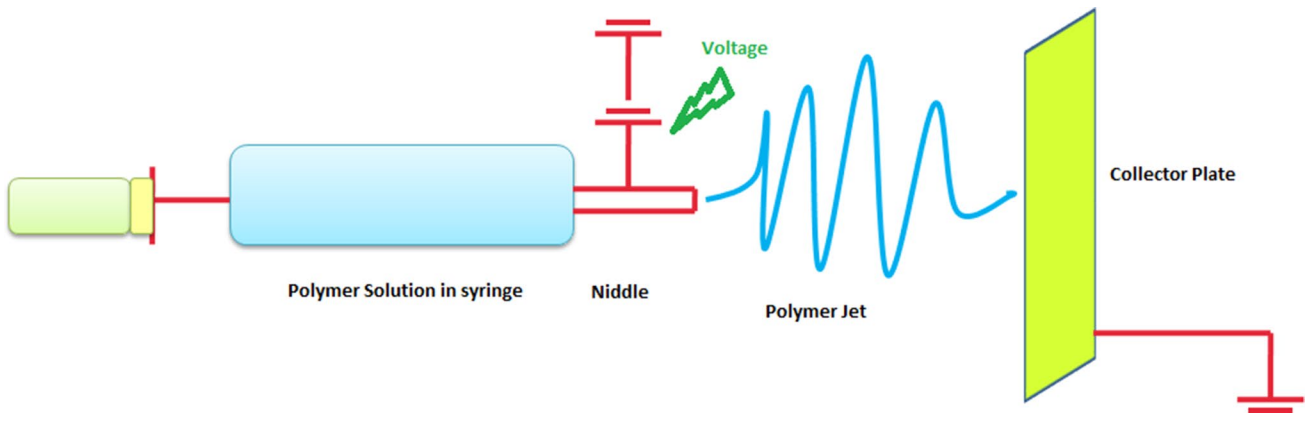


Fig. 1 Electrospinning setup

tion cannot take place formed which results in no fiber formation. Fiber formation occurs only when the voltage surpasses a minimum threshold value. (Yördem et al. 2008) If voltage is increased to a higher value, stretching of the jet during the flight increases substantially leading to thinner electrospun fibers (Baumgarten 1971).

d) *Tip to collector distance (TCD)*: TCD should have appropriate value to obtain different kinds of fibers. If the TCD is very less, it may result in the formation of an agglomerate fiber (Bhardwaj and Kundu 2010; Lee et al. 2004). When TCD is substantially high, the diameter of the fiber would decrease as the polymer jet have to cover more distance which resulted in an increase in the stretching of the jet during flight (Tabe 2014).

e) *Molecular weight:* The molecular weight of the polymer affects its viscosity by entangling the polymer chains. A polymer with less molecular weight will lead to the formation of beads. Increasing the molecular weight will result in the formation of the smooth fiber. The formation of micro-ribbon would take place when the molecular weight of the polymer is very high (Koski et al. 2004).

f) *Viscosity:* Viscosity should be appropriate to obtain smooth fibers. Solution with very little viscosity produces smooth and continuous fibers. However, the solutions with extremely high concentrations make it difficult to eject the solution out of the syringe to form a jet. Therefore, the viscosity should be tuned to a suitable value by changing the concentration of the solution (Larrondo and Manley 1981).

g) *Surface tension:* Surface tension determines the morphology and diameter of the fiber when all other parameters are kept constant. Beaded nanofibers can be converted into smooth fibers by reducing the surface tension and keeping all other conditions constant. (Yang et al. 2004)

h) *Conductivity:* The conductivity of a solution is determined by the polymer and the solvent used for electrospinning. If the polymer is a poly-electrolyte, it will face higher tension during the flight and will result in poor fibers. (Zong et al. 2002) However, smaller diameter fibers can be obtained by increasing the conductivity of a solution by adding salts to it (Huang et al. 2006).

Electrospun fibers for healthcare applications

Several methods are available for the production of nanofibers like phase separation (Zhang and Lu 2014), template synthesis (Tao and Desai 2007), an island in the sea (Zhang et al. 2017a, b), drawing (Suzuki and Tanizawa 2009), gas jet technique (Reneker et al. 2002), self-assembly (He et al. 2010), nano-lithography (Noy et al. 2002), melt blowing (Zuo et al. 2013), and electrospinning (Kalaoglu-Altan et al. 2016). Among these techniques, electrospinning has gained enormous popularity in recent years due to its low cost, easy setup, high aspect ratio, enhanced mechanical properties, better pore connectivity, and high surface-to-volume ratio. It also has the potential to electrospun almost any polymer solution of high molecular weight (Shi et al. 2015). Electrospinning has been successfully used to electrospun nanofibers of natural polymers, synthetic polymers, polymer blends, nanoparticle or drug-impregnated polymers, and ceramic polymers. Such polymers have huge application in various fields such as health care including tissue engineering (Casper et al. 2007), drug delivery (Stack et al. 2019), wound healing (Rho et al. 2006), dental materials (Zafar et al. 2016), solar cells (Nagata et al. 2013), nano-generators (Kim et al. 2020), lithium-ion batteries (Luo et al. 2014), supercapacitors (Kim et al. 2007), filtration membranes (Yoon et al. 2006), self-cleaning materials (Cooley 1902), reinforcement fillers (Neppalli et al. 2013), sensors (Yang et al. 2010), and many more. Modification in the process parameters discussed above can result in nanofibers with different structures and morphologies such as beaded, porous, ribbon, hollow, yarn, and core-shells (Shi et al. 2015).

The emergence of nanotechnology and biocompatibility of natural materials has led to the advancement of the use of nano-forms of various biomaterials to be used in biomedical science. A wide range of biomaterials' natural protein nanofibers, synthetic protein nanofibers, blends of these polymers with other organic polymers have been found to show interesting properties. These include their biodegradability, biocompatibility, ease in availability and ability to mimic the extracellular matrix (ECM). Nano-materials would help to early diagnose the disease and prevent it before its manifestations which makes the treatment easily affordable for all (Ramakrishna et al. 2006). Nanofiber mats and films have significant importance in biomedical science. They have been explored for generating scaffolds for tissue engineering, wound dressing, and drug delivery systems. Nanofiber mats and films can mimic the ECM and the amount of drug loaded in them to be released can be varied (Wen et al. 2017). Moreover, the cells seeded on tissue mimic nanofibers mats show better adherence, growth and proliferation corresponding to its high surface area, porous architecture. This allows ease in permeation of water vapors, gases, nutrients, and metabolites which are essential for the growth of neo-tissues.

Electrospun nanofibers scaffolds can be arranged in two orientations; aligned and random fibers (Fig. 2). The orientations of fibers play a crucial role in determining the rate of cell proliferation and adhesion. Poly(lactide-co-glycolide) (PLGA) and poly(ϵ-caprolactone) (PCL) were co-electrospun to generate a nanofibrous scaffold having both types of

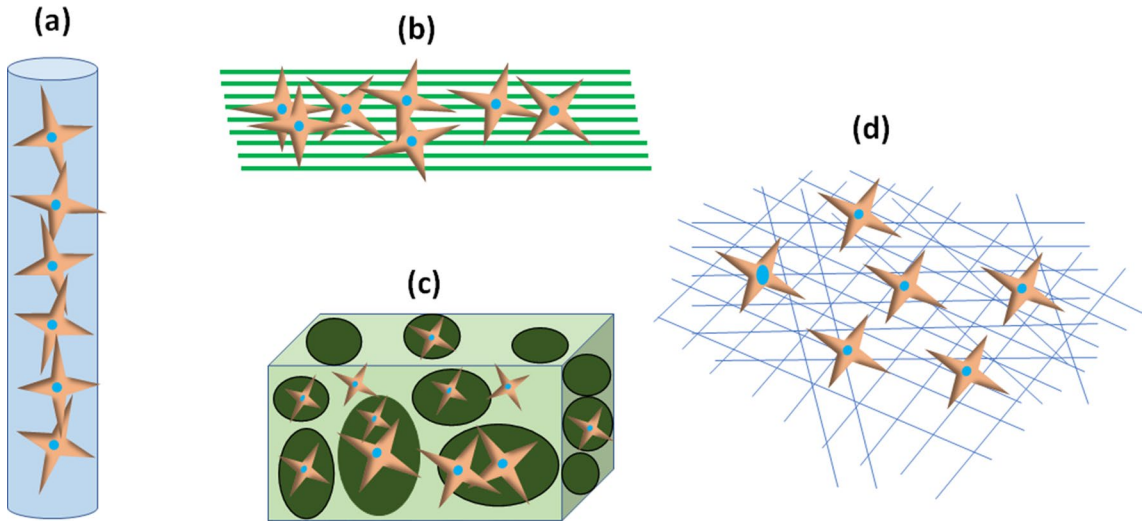


Fig. 2 Effect of the morphology of scaffolds on adherence, growth and proliferation of cells. Cells seeded on: **a** microfibers, **b** aligned nanofibers, **c** porous hydrogels, **d** random nanofibers

alignment. Human vascular smooth muscle cells (HVSMCs) were cultured on both scaffolds. The aligned nanofibrous scaffold structure maintained the cell shape and structure during the growth phase as compared to the random orientation nanofibrous scaffold (Zhong et al. 2015).

Lee et al. (2005) studied the effects of extracellular matrix (ECM) production of human ACL fibroblast on aligned nanofiber sheets, aligned nanofiber bundles, and randomly oriented nanofiber sheets. It was found that a higher amount of collagen was synthesized on aligned nanofiber sheets over the others. In addition, human ligament fibroblast (HLF) cells maintained their spindle shape on aligned nanofiber sheets and oriented themselves in the direction of nanofibers. The above two studies showed that aligned orientation is a promising candidate in the field of tissue engineering.

Electrospun nanofiber has several advantages such as (i) high surface-to-volume ratio, (ii) porosity, (iii) ease to fabrication, (iv) low cost, and (v) excellent biocompatibility. However, there are several challenges also associated with nanofibers in the field of tissue engineering and regenerative medicine. The electrospun nanofibers possess inadequate mechanical strength, especially for the regeneration of hard tissue. During the electrospinning, we must control several parameters such as concentration, voltage, tip-to-collector distance, pH, and temperature (Jin et al. 2004). Variation in any parameter can alter the morphology of nanofibers. It is very difficult to electrospun the branched biopolymer such as pectin and xanthan. Several attempts have been made by blending these branched biopolymers with other polymers having high electro-spinnability. Electro-spinnability of a charge polymer like the protonated chitosan is extremely difficult. The positive charged chitosan interferes electric field applied to syringe. It is difficult to control pore size, and polydispersity in pore size is very high. Small pore size (usually 400–1000 nm) results in limited in-filtration and migration of cells. Most of the natural proteins have solubility in buffer solution that makes difficult the spinning of buffer protein solution having high concentration of ions due to charge repulsion (Ghosal et al 2014 and Rather et al 2018). In our previous research work, we have also found that electrospinning of non-mulberry silk fibroin protein ionic liquid solution is very difficult due to charge repulsion. Before the spinning of these solution, it is important to remove the charged ions by dialysis (Srivastava and Purwar 2016; Srivastava et al 2019).

Protein-based electrospun fibers

Proteins are complex nitrogen-containing molecules that are made up of several smaller units called amino acids, that are linked to each other through a peptide bond (Hoffman and Falvo 2004). Natural proteins are either animal-based or plant-based. Several animal proteins including casein (Xie and Hsieh 2003), collagen (Rho et al. 2006), keratin (Thompson et al. 2016), silk (Xu et al. 2017), gelatin (Tonsomboon and Oyen 2013), elastin (Sell et al. 2006), and fibrinogen (McManus et al. 2008) have been electrospun (either alone or as blends with other polymers). Similarly, various plant proteins have also been electrospun which includes soy protein (Xu et al. 2014), zein (Qureshi et al. 2017), and pea protein (Aguilar-Vázquez et al. 2018) (either alone or as blends with other polymers).

The formation and morphology of various nanofibers depend on the solvent system as well as the process parameters. We have discussed the effect of electrospinning

parameters (flow rate, voltage, and tip-to-collector distance) and the solvent system on the morphology and diameter for various plant and animal-based proteins.

Process parameters for different animal-derived protein nanofibers

Collagen-based nanofibers have been successfully electrospun for healthcare applications. The selection of solvent is a crucial component for electrospinning collagen protein to avoid its denaturation and loss of identity. 1,1,1,3,3,3-Hexafluoro-2-propanol (HFP) has been used for the electrospinning of collagen fiber due to its high volatility and rapid evaporation from the solution during the jet flight (Law et al. 2017).

Trifluoroethanol (TFE) is another highly volatile solvent that promotes collagen fiber formation. However, it is expensive, toxic and usually denatures collagen protein. Hence, the alternative to TFE has been explored and some other eco-friendly solvents were standardized for collagen electrospinning (Law et al. 2017). Collagen has been electrospun with Phosphate-Buffered Saline (PBS) and ethanol (1:1 wt/wt) at a variable voltage of 15–20 kV, flow rate 0.1–0.2 mL/h, and tip-to-collector distance of 16 cm. This led to the production of both random and aligned fibers with diameters range 250–720 nm (average 467 ± 76 nm) and 340–860 nm (average 496 ± 97 nm), respectively (Baek et al. 2016).

Fibrinogen is a natural, globular, plasma protein that contributes an important role in clot formation (McManus et al. 2006a, b). Several methods have been proposed to electrospun fibrinogen nanofibers for their use as tissue engineering scaffold, wound healing, and hemostasis products. This is because fibrinogen nanofiber scaffolds provide better cellular interaction over other scaffolds based on synthetic polymers. Fibrinogen scaffolds generated from traditional techniques had less mechanical strength as compared to electrospinning. Wnek et al. in 2003 electrospun fibrinogen fiber with average diameter 80–700 nm at a voltage of 22 kV, feed rate 1.85 mL/h and tip-to-collector distance of 12.5 cm.

Gelatin is a naturally abundant protein derived from the hydrolysis of collagen (Choktaweesap et al. 2007). Gelatin has been electrospun with several solvent systems. However, electrospinning of gelatin solely in aqueous media is quite difficult. This is because the aqueous solution of gelatin forms gel below 298 K which results in water-soluble nanofibers. To avoid such formation, gelatin is either electrospun with an organic solvent or a solvent system containing water as one of the components. One such solvent system (ethanol/formic acid/water) has been reported by Chen et al. (2008a, b) where gelatin was electrospun with glutaraldehyde (GTA), an across-linker. It was observed that gelatin cannot be electrospun with formic acid at a concentration below 7 wt%. Fibers of different morphology and diameter were obtained which depends on sample concentration. When 10 wt% solution was electrospun, beads were formed without the formation of nanofibers. However, at a higher gelatin concentration (15 wt%), the formation of nanofibers was observed with an average diameter of 77 ± 23 nm along with beads. Further, increase in the concentration (up to 20 wt%), selective formation of fibers has been observed with a diameter of 119 ± 77 nm. It was also observed that if the concentration of gelatin was increased to 25 wt% and 30 wt%, the diameter of fibers was increased to 169 ± 125 nm and 132 ± 70 nm, respectively. Gelatin fibers generally transform from helical to random coil conformation when electrospun in formic acid (Chen et al. 2008a, b; Ki et al. 2005).

Silk is a natural fibrous protein that has been extensively explored for biomedical applications. India is rich in four different varieties of silk namely mulberry, muga, tasar and eri. The muga, tasar and eri varieties of silk are known as non-mulberry silk. Mulberry silk fibroin proteins have been successfully electrospun using different solvents such as 1,1,1,3,3,3-hexafluoro-2-propanol (Zhang et al. 2009), formic acid (Amiraliyan et al. 2009), water (Meinel et al. 2009), chloroform (Paşcu et al. 2013), dichloromethane (Li et al. 2011), acetic acid (Zhou et al. 2010), and trifluoro ethylene (Biazar et al. 2015). Solvent plays a very vital role in electro-spinnability and controls morphology and diameter of the fiber. To enhance the processability of mulberry silk fibroin, it is blended with another polymer such as poly(ethylene oxide) (Jin et al. 2004), poly(vinyl alcohol) (Sayed et al. 2019), poly(L-lactic acid-*co*-ε-caprolactone) (Yao et al. 2018), poly(D,L-lactide-co-glycolide) (Zhou et al. 2015), and poly(ε-caprolactone) (Lee and Kim 2010). The concentration of silk fibroin in formic acid plays a crucial role to decide the morphology and diameter of the fiber. Kim and coworkers (2003) electrospun silk fibroin in formic acid by varying concentration between 6 and 15 wt% at voltage 15 kV, flow rate 1.5 mL/h and air gap distance 7 cm. Nanofibers with cylindrical morphology having a diameter ranging from 30 to 120 nm (average 80 nm) have been obtained.

Silk fibroin self assembles itself in aqueous media to form different nanostructures such as nano-spheres and nano-filaments depending upon chain folding and hydrophobic interactions of the silk with solvent (Zhang et al. 2012). Concentration plays a crucial role in deciding the diameter and morphology of the silk fiber. Wang and coworkers (Wang et al. 2005) have reported that if the concentration of silk fibroin is below the critical concentration of 17 wt% in aqueous media, no fiber could be fabricated. This is because no entanglement between silk will take place at a low concentration which results in the jet break and leads to the formation of beads or droplets. On the other hand, increasing the concentration to 28 wt%, the formation of circular smooth fibers with a diameter of 400–800 nm takes

place. On further increasing the concentration to 39 wt%, ribbon-shaped uneven fibers were obtained. At higher concentrations and viscosity, the solution became unstable and heterogeneous. Thus, the solution left with a good amount of solvent which led to the solidification of fibers and the fabrication of ribbon-shaped uneven fibers takes place.

A low amount of silk fibroin is needed to make a solution in formic acid than water. This is due to the high viscosity of the formic acid (Kishimoto et al. 2017; Wang et al. 2011). Silk fibers electrospun in formic acid show a higher diameter than in an aqueous solution at the same concentration since formic acid is highly volatile. It is rapidly removed before excessive stretching of flight whereas the water being low volatile is slowly vaporized and experience excessive stretching (Kishimoto et al. 2017).

Recently, non-mulberry silk fibroin proteins are reported to have higher biocompatibility over the mulberry silk fibroin proteins due to the presence of tripeptide sequence; Gla(R)-Ala(G)-Arg(D) in the former. Higher adherence, growth and proliferation of cells seeded on non-mulberry silk are due to the presence of arginine residue that attracts negative charge cell surface. Non-mulberry silk fibroin proteins have not been fully explored due to their insolubility in most of the solvents and most of the studies have been carried out on silkworm gland protein.

Process parameters for different plant-derived protein nanofibers

Soy protein is an abundantly available cheap plant protein extracted from soybean. Soy protein has been blended with several polymers to electrospun novel biocomposite mats and fibers. Soy protein blended with poly(ethylene oxide) in 1% NaOH at voltage 15–27 kV, feed rate 0.004 mL/min and tip-to-collector distance of 18 cm can be fabricated as fiber of diameter ranging from 50 to 270 nm. The strength of fibers was increased by cross-linking it with carbodiimide. The fibers were found to be viable, and increased human mesenchymal stem cell adhesion and proliferation have been observed (Ramji and Shah 2014).

Soy protein has been electrospun solely in NaOH at voltage 45 kV and tip-to-collector distance 25 cm to fabricate 2D and 3D fibers. Human mesenchymal cells showed maximum cell adhesion and proliferation in 15 days on both fibers. 3D soy protein fibers showed a uniform distribution of cells and almost double the rate of cell proliferation in comparison to both 2D and commercial 3D porous PCPU (polycarbonate polyurethane-urea) scaffold (Xu et al. 2014).

Zein is a yellowish prolamin protein obtained from maize (corn). Zein was electrospun in ethanol/water (70:30 v/v ratio) with a flow rate of 6 mL/h, tip-to-collector distance 10 cm and the voltage between 12 and 20 kV. When 10% w/v zein was electrospun, the fibers obtained had microbeads' morphology. On increasing the concentration to 20% w/v, fibers with a diameter of around 500 nm were obtained with fewer beads. On further increasing the concentration to 30%, 40% and 50% w/v, thick fibers having ribbon-like morphology with a diameter around 1000, 2000, and 7000 nm, respectively, have been obtained. It can be concluded that the concentration of zein in the solvent plays a significant role in deciding the morphology and diameter of the fiber obtained. The higher concentration of zein in the solvent increases the fiber diameter (Yao et al. 2007). At 50% w/v zein concentration, ribbon-like morphology was obtained due to the collapsing of fibers at a higher concentration (Koombhongse et al. 2001). Zein nanofibers synthesized via electrospun were renewable, biodegradable, flexible, and more lustrous but have lower mechanical strength. Thus, the fibers were cross-linked with hexamethylene diisocyanate (HDI) which successfully increased the tensile strength of the zein fibers to a great extent (Yao et al. 2007).

Role of protein-based electrospun nanofibers in health care

A wide range of synthetic as well as natural polymers has been electrospun for healthcare applications. Natural polymers are biocompatible, promote better cell adhesion, proliferation and differentiation but lose their native structure and biological properties when dissolved in the organic solvent. On the other hand, synthetic polymers are less compatible with the body due to the presence of hydrophobic parts. However, synthetic fibers possess better mechanical properties and are easier to synthesize, process and modify (Gunn and Zhang 2010). To prepare hybrid materials having advantages of both synthetic and natural polymers, researchers blended both types of polymers and then electrospun them into nanofibers (Kai et al. 2013).

The nanofibers obtained from proteins are the potential candidate in the medical field especially for regenerative medicine, wound healing and drug delivery. Protein nanofibers are better than other fibers because of their biocompatibility, tunable biodegradability and better mechanical integrity. Several proteins electrospun for tissue engineering and wound healing are shown in Tables 1 and 2, respectively.

Tissue engineering

The human body consists of several different kinds of tissues: muscle, epithelial, nervous, connective tissues, etc. Not all these tissues can regenerate. When a non-regenerable tissue is damaged either due to an injury, disease or congenital defects, the tissue has to be replaced or transplanted to recover the body function. Tissue engineering is the branch that deals with fabricating the scaffold for the formation of

Table 1 List of proteins or peptides electrospun for tissue engineering

Protein	Other polymer or material	Solvent	Electrospinning parameters	Diameter of fiber	Applications	References
Casein	PEO/PVA	TEA	Voltage = 25 kV TCD = 30 cm	100–500 nm	Tissue engineering	Xie and Hsieh (2003)
Casein	PEO/AgNPs	Water	Flow rate = 0.3 mL/h Voltage = 20 kV TCD = 15 cm	Around 200 nm	Tissue engineering	Selvaraj et al. (2018)
Collagen	-	HFP	Flow rate = 2–8 mL/h Voltage = 20–25 kV TCD = 10 cm	0.10 ± 0.04 to 4.6 ± 2.1 µm	Vascular tissue engineering	Boland et al. (2004)
Collagen	-	AA/HFP	Flow rate = 9 mL/h Voltage = 14 kV TCD = 22 cm	3–6 µm	Tissue engineering	Casper et al. (2007)
Collagen	-	HFP	Flow rate = 0.02 mL/min Voltage = 15–20 kV TCD = 21.5 cm	100–1200 nm	Tissue engineering	Rho et al. (2006)
Collagen	PCL	Methanol/chloroform/HFP	Flow rate = 1 mL/h Voltage = 13 kV TCD = 13 cm	300–700 nm	Blood vessel engineering	Venugopal et al. (2005)
Collagen	P(LLA-CL)	HFP	Flow rate = 1.2 mL/h Voltage = 12 kV TCD = 12 cm	100–200 nm	Blood vessel substitutes	He et al. (2005)
Collagen/elastin	PEO	HCl	Flow rate = 30–69 µL/min Voltage = 10–25 kV TCD = 20–30 cm	220–600 nm	Blood vessel substitutes	Buttafoco et al. (2006)
Collagen/elastin	PLGA	HFP	Flow rate = 3 mL/h Voltage = 25 kV TCD = 15 cm	0.720 ± 0.35 mm	Cardiac tissue engineering	Stitzel et al. (2006)
Collagen	Gelatin	HFP	Flow rate = 1–10 mL/h Voltage = 10 kV TCD = 15 cm	200–500 nm	Tissue engineering	Li et al. (2005)
Collagen	-	AA	Flow rate = 0.05–0.30 mL/h Voltage = 4–9 kV	50–451 nm	Corneal tissue replacement	Phu et al. (2011)
Collagen	-	AA	Flow rate = 0.05–0.30 mL/h Voltage = 4–9 kV	160–240 nm	Corneal tissue replacement	Wray and Orwin (2009)
Collagen	-	HFP/DMF	Flow rate = 4 µL/min Voltage = 13 kV TCD = 10 cm		Cornea tissue replacement	Acun and Hasirci (2014)
Collagen	-	AA/DMSO	Flow rate = 0.60 mL/h Voltage = 17 kV TCD = 23 cm	200–1100 nm	Cardiac tissue engineering	Elamparithi et al. (2015)
Collagen	-	AA/HFP/PLLA	Flow rate = 2.5 mL/h Voltage = 16 kV TCD = 23 cm	1.82 µm	Liver regeneration	Grant et al. (2019)
Collagen	HA	HFP	Flow rate = 1.5 mL/h Voltage = 13 kV TCD = 13 cm	293 ± 1.45 nm	Bone tissue engineering	Venugopal et al. (2008)
Elastin	PDO	HFP	Flow rate = 4 and 8 mL/h Voltage = 22 kV TCD = 12 cm	0.4–1.2 microns	Cardiovascular graft	Sell et al. (2006)

Table 1 (continued)

Protein	Other polymer or material	Solvent	Electrospinning parameters	Diameter of fiber	Applications	References
Elastin	PDO	HFP	Flow rate = 5.25–6 mL/h Voltage = 24 kV TCD = 12.7 cm	–	Vascular tissue engineering	Smith et al. (2008)
Eri fibroin	n-Hap	TFA	Voltage = 20 kV TCD = 15 cm	700–1500 nm	Bone tissue engineering	Andiappan et al. (2013)
Eri-tasar silk fibroin	-	FA/chloroform	Flow rate = 0.5 mL/h Voltage = 20 kV TCD = 15 cm	< 300 nm	Bone tissue engineering	Panda et al. (2014)
Tropoelastin	Collagen	HFP	Flow rate = 3 mL/h Voltage = 20 kV TCD = 20 cm	21–6.5 µm	Skin tissue engineering	Rnjak-Kovacina et al. (2012)
Fibrinogen	PDO	HFP	Flow rate = 1.8 mL/h Voltage = 22 kV TCD = 10 cm		Urologic tissue engineering	McManus et al. (2008)
Fibrinogen	-	HFP	Flow rate = 1.8 mL/h Voltage = 22 kV TCD = 10 cm	120–610 nm	Tissue engineering	McManus et al. (2006a, b)
Fibrinogen	-	HFP	Flow rate = 1.8 mL/h Voltage = 22 kV TCD = 10 cm	80–500 nm	Tissue engineering	McManus et al. (2006a, b)
Fibrinogen	-	HFP	Flow rate = 3.5 mL/h Voltage = 25 kV TCD = 12 cm	710 ± 220 nm	Tissue engineering	Sell et al. (2008)
Fibrinogen	P(LLA-CL)	HFP	Flow rate = 0.8 mL/h Voltage = 20 kV TCD = 13 cm	145–800 nm	Soft tissue engineering	He et al. (2011)
Gelatin	-	AA/FA	Flow rate = 0.8 mL/h Voltage = 15 kV TCD = 15 cm	109–761 nm	Tissue engineering	Songchotikunpan et al. (2008)
Gelatin	-	AA/ethyl acetate/water	Flow rate = 0.5 mL/h Voltage = 12 kV TCD =	110 ± 40 nm and 600 ± 110 nm	Bone tissue engineering	Sisson et al. (2010)
Gelatin	PCL	HFP	Flow rate = 1 mL/h Voltage = 12 kV TCD =	113 ± 33, 189 ± 56, and 431 ± 118 nm	Nerve tissue engineering	Ghasemi-Mubarakeh et al. (2008)
Gelatin	PANi	HFP	Flow rate = Voltage = 10 kV TCD = 10 cm	61 ± 13 nm	Tissue engineering	Li et al. (2006a, b)
Gelatin	PLLA	Methylene chloride/DMF	Flow rate = 2 mL/h Voltage = 26 kV TCD = 15 cm	190–250 nm	Cartilage tissue engineering	Chen and Su (2011)
Gelatin	PLA	HFP	Flow rate = 1.0 mL/h Voltage = 15 kV TCD = 15 cm	200 nm–2 µm	Tissue engineering	Yuan et al. (2008)
Gelatin		AA/ethyl acetate/water	Flow rate = 0.003 mL/min Voltage = 12 kV TCD = 10 cm	60–148 nm	Corneal tissue replacement	Tonsomboon and Oyen (2013)
Gelatin	PLLA	HFP/DMF	Flow rate = 30 µL/min Voltage = 15 kV TCD = 18 cm	750–1000 nm	Corneal tissue replacement	Zhang et al. (2015)

Table 1 (continued)

Protein	Other polymer or material	Solvent	Electrospinning parameters	Diameter of fiber	Applications	References
Gelatin	PLLA	HFP/DMF	Flow rate = 30 µL/min Voltage = 15 kV TCD = 18 cm	800–1000 nm	Cornea tissue replacement	Yan et al. (2011)
Hemoglobin	Gelatin/fibrinogen	TFE/HFP	Flow rate = 1 mL/h Voltage = 15 kV TCD = 14 cm	290 ± 130 nm	Cardiac tissue engineering	Ravichandran et al. (2013a, b)
Keratin	PLA	HFP	Flow rate = 1 mL/h Voltage = 15 kV TCD = 15 cm	500 nm	Tissue engineering	Yuan et al. (2008)
Keratin wool	PLGA	TCM/DMF	Flow rate = 0.2 mL/min Voltage = 20 kV TCD = 13 cm	–	Bone tissue engineering	Zhang and Liu (2013)
Keratin	Fibroin	FA	Voltage = 15 kV TCD = 10 cm	4.20 ± 1.19 µm, 2.94 ± 0.81 µm, and 2.27 ± 0.65 µm	Vascular tissue engineering	Yen et al. (2016)
Silk fibroin	PEO	Water	Flow rate = 1.2 mL/h Voltage = 12–15 kV TCD = 19 cm	530 ± 100 nm	Tissue engineering	Meinel et al. (2009)
Silk fibroin	PEO	Water	Flow rate = 0.03 mL/h Voltage = 12.5 kV TCD = 21.5 cm	700 ± 50 nm	Bone tissue engineering	Jin et al. (2004)
Silk fibroin	Collagen	AA	Flow rate = 0.015 mL/min Voltage = 20 kV TCD = 20 cm	1.11 ± 0.20 µm, 1.29 ± 0.28 µm, 1.40 ± 0.28 µm and 2.45 ± 0.80 µm	Vascular tissue engineering	Zhou et al. (2010)
Silk fibroin	PHBV, nHAp	Chloroform	Flow rate = 5 mL/h Voltage = 10 kV TCD = 15 cm	13.04 ± 1.14 µm, and 18.03 ± 1.36 µm	Bone tissue engineering	Paşcu et al. (2013)
Silk fibroin	PLGA-collagen	HFP	Flow rate = 4 mL/h Voltage = 10 kV TCD = 10 cm	280 ± 26 to 168 ± 21 nm	Nerve tissue engineering	Wang et al. (2011)
Silk fibroin	PEO/nHAP/BMP-2	Water	Flow rate = 0.02 mL/min Voltage = 12 kV TCD = 21.5 cm	520 ± 55 nm	Bone tissue engineering	Li et al. (2006a, b)
Silk fibroin	PCL	DCM/HFP	Flow rate = 1.5 mL/h Voltage = 15 kV TCD = 12–15 cm	343 ± 131 nm 274 ± 92 nm	Tissue engineering	Li et al. (2011)
Silk fibroin	P(LLA-CL)	HFP	Flow rate = 1.2 mL/h Voltage = 12 kV TCD = 12–15 cm	201–237 nm	Tissue engineering	Zhang et al. (2009)
Silk fibroin	-	Water	Flow rate = 1.2 mL/h Voltage = 20 kV TCD = 10 cm	–	Liver regeneration	Xu et al. (2017)
Silk fibroin	-	TFE	Voltage = 24 kV	–	Cornea tissue replacement	Biazar et al. (2015)
Silk fibroin	P(LLA-CL)	HFP	Flow rate = 1 mL/h Voltage = 12 kV TCD = 11 cm	123–649 nm	Cornea tissue replacement	Chen et al. (2015a, b)
Soy protein	-	Sodium hydroxide	Voltage = 45 kV TCD = 25 cm	–	soft tissue engineering	Xu et al. (2014)

Table 1 (continued)

Protein	Other polymer or material	Solvent	Electrospinning parameters	Diameter of fiber	Applications	References
Soy protein	PEO	Sodium hydroxide	Flow rate = 0.004 mL/min Voltage = 15 to 27 kV TCD = 18 cm	50–270 nm	Tissue engineering	Ramji and Shah (2014)
Tasar silk	PCL	Chloroform	Flow rate = 0.1 mL/h Voltage = 11 kV TCD = 15 cm	150 and 500 nm	Bone tissue engineering	Bhattacharjee et al. (2015)
Tasar silk	AgNPs	BMIMAc	Flow rate = 0.5 mL/h Voltage = 15, 20 and 25 kV TCD = 8 cm	300–1400 nm	Skin tissue engineering	Srivastava et al. (2019)
Tussah silk fibroin	nHAP	Water/ammonia	Flow rate = 0.1–0.3 mL/h Voltage = 20 kV TCD = 18 cm	510–840 nm and 860–950 nm	Bone tissue engineering	Shao et al. (2016)
Tussah silk fibroin	PCL	HFP	Flow rate = 0.6 mL/h Voltage = 15 kV TCD = 15 cm	226 ± 24 nm 237 ± 43 nm	Tissue engineering	Li et al. (2017)
Zein		Citric acid, ethanol	Flow rate = 2 mL/h Voltage = 19.5 kV TCD = 20 cm		Tissue engineering	Jiang et al. (2010)
Zein	PGS	AA	Flow rate = 0.3 mL/h Voltage = 20 kV TCD = 15 cm	0.09 ± 0.03 µm	Soft tissue engineering	Vogt et al. (2018)

In solvents: *HFP* 1,1,1,3,3,3-hexafluoro-2-propanol, *AA* acetic acid, *HCl* hydrochloric acid, *FA* formic acid, *TFA* trifluoroacetic acid, *BMIMAc* 1-butyl-3-methyl imidazolium acetate, *TCM* trichloromethane, *DCM* dichloromethane, *TFE* tetrafluoroethylene, *TEA* triethanol amine, *DMF* dimethylformamide

In other polymers: *PEO* poly(ethylene oxide), *PVA* poly(vinyl alcohol), *P(LLA-CL)* poly(L-lactic acid)-*co*-poly(-caprolactone), *PLGA* poly(D,L-lactide-co-glycolide), *PDO* polydioxanone, *PCL* poly(ε-caprolactone), *PANi* polyaniline, *PLLA* poly(lactic acid), *PHBV* polyhydroxybutyrate-co-(3-hydroxyvalerate) with valerate fraction, *nHAp* nano-hydroxyapatite, *PLA* polylactide, *PGS* poly(glycerol sebacate), *BMP*-2 bone morphogenetic protein 2, *HA* hydroxyapatite

Parameters: *TCD* tip-to-collector distance

new cells. For the treatment of damaged tissues, materials are explored to design and fabricate suitable scaffolds which show biocompatibility, mechanical strength, high porosity (enhance the growth of cells), biodegradability (scaffold can be replaced with the newly generated cells) (Liu et al. 2012a, b). Electrospun nanofibers and loosely connected 3D porous mats make suitable scaffolds as they can imitate the biological environment. This is due to the same scale, high porosity and high surface area which makes them a potential candidate for regenerative medicine (Agarwal et al. 2008; Gerardo-Nava et al. 2009; Jiang et al. 2008) (Fig. 3).

Based on different tissues present in the body, tissue engineering scaffolds have been designed for the heart, cornea, liver, blood vessel substitutes, bone, nerve, and skin. Such systems are discussed in the following sub-sections.

Bone tissue engineering

Bone grafting is done if an injured bone cannot be treated (Andiappan et al. 2013). Bone grafting is a technique to replace the damaged bone with either a bone harvested from one anatomic site and transplanted to the defective site within the patient's body (autologous bone graft). In the second method, bone can also be harvested from a donor, taken from the bone bank, and transplanted into the patient's body (allograft) (Paşcu et al. 2013).

Autologous bone grafting is one of the preferred therapy paths since it is an immune privilege. This grafting has certain drawbacks such as donor site pain, increased operative time, excess blood loss, chances of infection at the donor site, and limited supply of graft. While allografts can be arranged easily from a corpse but are less biocompatible and possess involves the risk of pathogen transmission (Roberts and Rosenbaum 2012). A new alternative to this method was

Table 2 List of proteins or peptides electrospun for wound healing

Protein	Other polymer or material	Solvent	Electrospinning parameters	Diameter of fiber	Applications	References
Collagen	Zein	AA	Flow rate = 0.75 mL/h Voltage = 20 kV TCD = 15 cm	910 nm	Wound dressing	Lin et al. (2012)
Collagen	PLGA	HFP	Flow rate = 4 mL/h Voltage = 15–20 kV TCD = 15 cm	150–650 nm	Wound dressing	Liu et al. (2010)
Collagen	Chitosan/PEO	Water	Flow rate = 0.3 to 1.5 mL/h Voltage = 20 to 34 kV TCD = 10 to 25 cm	398 ± 76 nm	Wound dressing	Chen et al. (2008a, b)
Collagen	-	HFP	Flow rate = 0.02 mL/min Voltage = 15–20 kV TCD = 21.5 cm	100–1200 nm	Wound dressing	Rho et al. (2006)
Eri silk fibroin	PLA	TFA/chloroform	Flow rate = 1 mL/h Voltage = 20 kV TCD = 15 cm	502 nm	Wound dressing	Shanmugam and Sundaramoorthy (2015)
Fibrinogen	-	HFP	Flow rate = 1.85 mL/h Voltage = 22 kV TCD = 12.5 cm	80–700 nm	Wound dressing	Wnek et al. (2003)
Fibrinogen	-	HFP	Flow rate = 1.8 mL/h Voltage = 22 kV TCD = 7.5 cm	120–610 nm	Wound dressing	McManus et al. (2006a, b)
Gelatin	PCL	TFE	Flow rate = 0.7 mL/h Voltage = 10.5 kV TCD = 15 cm	470 ± 120 nm	Wound dressing	Chong et al. 2007
Gelatin	PU	HFP	Flow rate = 1 mL/h Voltage = 18 kV TCD = 15 cm	0.4–2.1 µm	Wound dressing	Kim et al. (2009)
Gelatin	Keratin	FA/PVA/water	Flow rate = 0.1 mL/h Voltage = 20 kV TCD = 10 cm	160.4 nm	Wound dressing	Yao et al. (2017)
Keratin	PU/AgNPs	HFP	Flow rate = 1.5 mL/h Voltage = 20 kV TCD = 15 cm	637 ± 211 nm, 616 ± 219 nm and 593 ± 187 nm	Wound dressing	Wang et al. (2015)
Keratin	PHBV	HFP	Flow rate = 1 mL/h Voltage = 10 kV TCD = 15 cm	720 ± 124 nm	Wound dressing	Yuan et al. (2012a, b)
Silk fibroin	PEO	Water	Flow-rate = 0.02 mL/min Voltage = 10–11 kV TCD = 15–21 cm	1.06 µm ± 0.03	Wound dressing	Schneider et al. (2009)

Table 2 (continued)

Protein	Other polymer or material	Solvent	Electrospinning parameters	Diameter of fiber	Applications	References
Silk fibroin	PEO	Water	Flow rate = 1.2 mL/h Voltage = 20 kV TCD = 15 cm	–	Burnt wound treatment	Ju et al. (2016)
Silk Fibroin	AgNPs	HFP/FA	Flow rate = 0.8 mL/h Voltage = 12 kV TCD = 10 cm	–	Wound dressing	Uttayarat et al. (2012)
Silk Sericin	PVA	Water	Flow rate = 1 mL/h Voltage = 12–15 kV TCD = 10 cm	130–160 nm	Chronic wounds' treatment	Gilotra et al. (2018)
Soy protein isolate	PEO	HFP	Flow rate = 0.8 mL/h Voltage = 12 kV TCD = 15 cm		Wound dressing	Har-el et al. (2014)
Soy protein isolate	PEO	NaOH/ethanol	Flow rate = 1 mL/h Voltage = 15 kV TCD = 15 cm	151 nm and 415 nm	Wound dressing	Thirugnanaselvam et al. (2013)
Tasar fibroin		BMIMAc/FA/$CaCl_2$	Flow rate = 0.5 mL/h Voltage = 15 and 25 kV TCD = 8 cm	259–700 nm and 310–800 nm	Wound dressing	Srivastava and Purwar (2018)
Tasar silk fibroin	Chitosan	PVA	TCD = 15 cm	80–400 nm	Wound dressing	Sapru et al. (2018)
Zein	AgNPs	Citric acid/ethanol/NaOH	Flow rate = 2 mL/h Voltage = 20 kV TCD = 20 cm	350–500 nm	Wound dressing	Dashdorj et al. (2015)
Zein	PVA-SbQ	AA	Flow rate = 1 mL/h Voltage = 25 kV TCD = 15 cm	150–550 nm	Wound dressing	Cui et al. (2015)
Zein	Cellulose acetate/PU	DMF/MEK	Flow rate = 0.5 mL/h Voltage = 18 kV TCD = 15 cm	400–700 nm	Wound dressing	Unnithan et al. (2013)

In other polymers: *PVA* polyvinyl alcohol, *PLGA* poly (lactic-co-glycolic acid), *PEU* polyester urethane, *PU* polyurethane, *PEO* polyethylene oxide, *EDTA* ethylenediaminetetra acetic acid, *PLGA* polylactidepolyglycolide, *PCL* poly(ε-caprolactone), *PVA-SbQ* poly(vinyl alcohol)-stilbazol quaternized, *PHBV* poly(hydroxybutylate-co-hydroxyvalerate), *PLA* poly-L-lactic acid

In solvents: *TFE* trifluoroethanol, *DMF* *N*,*N*-dimethylformamide, *THF* tetrahydrofuran, *MEK* methyl ethyl ketone, *AA* acetic acid, HOBt hydroxybenzotriazole, *TPP* thiamine pyrophosphate, *HFP* 1,1,1,3,3,3-hexafluoro-2-propanol, *TFA* Trifluoroacetic acid, *TFE* tetrafluoroethylene, *BMIMAc* 1-butyl 3-methyl imidazolium acetate, *FA* formic acid, *PVA* polyvinyl alcohol

In parameters *TCD* tip-to-collector distance

explored, and scaffolds of several proteins such as collagen (Venugopal et al. 2008), eri silk fibroin (Andiappan et al. 2013), eri-tasar fibroin (Panda et al. 2014), gelatin (Sisson et al. 2010), keratin wool (Zhang and Liu 2013), mulberry silk fibroin (Jin et al. 2004), tasar silk fibroin (Bhattacharjee et al. 2015), and tussah silk fibroin (Shao et al. 2016) have been designed by electrospinning. Such materials have been found suitable for bone tissue replacement and explored as potential graft materials.

Andiappan and coworkers fabricated Eri, a non-mulberry silk fibroin, (ESF) scaffold with hydroxyapatite coating for comparing its property with pure ESF (Andiappan et al. 2013). MSCs were cultured on both the scaffolds and the growth of cells was monitored. Initially, with an increase in incubation time up to 5 days, the percentage of cells increased and remained almost constant till the 7th day. In addition, it was observed that the cells adhere and proliferate more readily on ESF-nHAP scaffold than on the ESF scaffold, indicating the role of nHAP in promoting cell growth.

Panda et al. (2014) compared the growth of hMSCs on eri-tasar (ET), Bombyx mori and gelatin nanofibers' scaffolds. Cells adhered and proliferated more on ET and

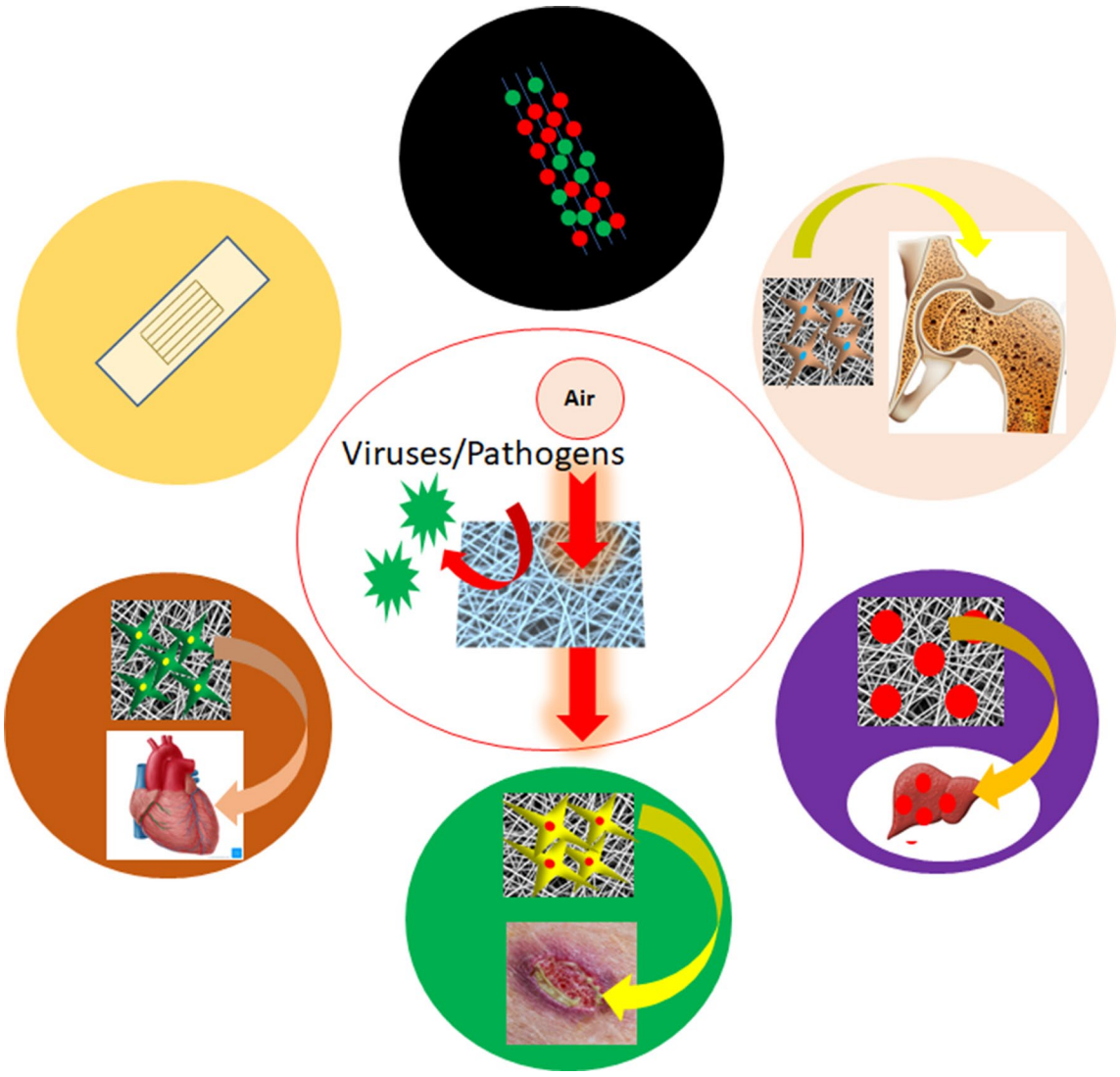


Fig. 3 Application areas of nanofibers in regenerative medicine

SF scaffold in comparison to gelatin. ET and SF scaffold showed almost comparable cell growth.

Sisson et al. (2010) prepared two different types of nanofiber mats with an average diameter of 600 ± 110 nm and 110 ± 40 nm. The MG63 cells were cultured on both the nanofibers mats. It was observed that the cell proliferation was comparable for both nanofiber mats up to 3 days of incubation. However, after 5 and 7 days of incubation, the growth of the cells was higher with the fiber having a larger diameter which suggests the potential of cells to distinguish the fibers of different diameters.

Zhang and coworkers (2013) reported that the adherence and proliferation of bone mesenchymal stem cells on keratin/PLGA nanofibers scaffold was significantly higher than that of pure PLGA nanofibers scaffolds after 3 and 5 days of incubation.

Shao and coworkers (2016) reported that osteoblast cells (MG63) cultured on scaffold fabricated from a composite of hydroxyapatite and tussah silk fibroin showed higher biocompatibility. It promotes better cell attachment and proliferation than that of pure Tussah silk fibroin after days 4 and 7 days of incubation.

Heart

Cardiovascular diseases remain at the top for the higher mortality rate and morbidity globally. Initially, few drugs were synthesized to cure such diseases but failed. Organ transplantation emerged as a new hope in the medical field. This could be done either autologous or harvested from certain donors and implanted into patients. Lack of autologous cardiac tissue was a problem that increased the demand for

donors (Nemeno-Guanzon et al. 2012). Even implantation was not easy due to the scarcity of heart donors and the high risk of rejection of graft by the patient's body. Many patients have died while remaining on the waiting list before the organ could be made available. Thus, it was very urgent to find an alternative to prepare cardiac tissues in laboratories to meet the demand. Cardiovascular tissue engineering made this possible by synthesizing tissues artificially in the laboratory (Yang et al. 2005).

Stitzel and coworkers (2006) electrospun nanofibers scaffold of the polymer blend, type I collagen, elastin and PLGA. It was co-cultured with endothelial cells and smooth muscle cells. This scaffold showed both in vitro and in vivo biocompatibility.

Elamparithi and coworkers (2015) electrospun collagen fiber and checked its biocompatibility and cytotoxicity on rat skeletal myoblasts L6 cell line. Primary neonatal rat ventricular cardiomyocytes (NRVCM) was cultured on a collagen scaffold. The scaffold was found to retain its contractile function and released a higher level of desmin. This study showed the potential role of collagen nanofibers in cardiovascular tissue engineering. Similarly, Sell et al. (2006) fabricated a scaffold of elastin composed with PDO and cultured human dermal fibroblasts on it. Within 7 days, cells migrated and showed full-thickness penetration growth compared to no penetration on pure PDO scaffold which suggests the inclusion of elastin in PDO enhances cell adhesion and proliferation.

Cornea

Approximately, 10 million people are suffering from vision impairment and are diagnosed to have damaged cornea (Ghezzi et al. 2015; Kong and Mi 2016; McLaughlin et al. 2009). Like other organs, allogeneic cornea grafting is a preferred therapeutic treatment. However, its rejection results in a high risk for grafting. Over 10% of cases of corneal tissue transplantation have rejected cornea within 1 year of surgery (Ghezzi et al. 2015). Acute shortages and rejection posed were serious problems. In this regard, other materials were explored to prepare three-dimensional bioengineered cornea in vitro having high biocompatibility and superior performance (McLaughlin et al. 2009).

Yan and coworkers (2011) fabricated gelatin and PLLA-based nanofibers scaffolds with both random and aligned orientation. They found that rate of cell adhesion and proliferation of different cells changes with change in the orientation of fiber. Corneal epithelial cells favored randomly oriented scaffolds where as keratinocytes favored alignment scaffolds.

Silk fibroin and P(LLA-CL) transparent scaffold has been prepared and cultured on a human corneal endothelial B4G12 cell line (Chen et al. 2015a, b). It has been revealed that the scaffold was biocompatible and suitable for monolayer formation.

Liver regeneration

As per the WHO report, about 2.5% of the total deaths reported all over the globe are caused due to liver diseases. Liver disease will rank among the 14th cause for mortality by 2030 (Palakkan et al. 2013) (Mathers and Loncar 2006). The liver is the only organ that shows high regenerative capacity. It could regenerate itself even after 70% of its loss. Despite this exceptional behavior, a patient suffering from end-stage liver cirrhosis, liver cancer, or chronic liver failure is left with Liver Transplantation. However, it suffers from the shortage of donors, immune rejection, and high-cost create problems in the procedure (Lee and Cho 2012). A variety of other methods were developed among which one consists of the use of stem cells such as embryonic stem cells (ESCs), induced pluripotent stem cells (iPSCs), and mesenchymal stem cells (MSCs). They can differentiate into hepatocyte-like cells in vitro. However, providing a suitable environment for differentiation in laboratories is very difficult (Lee and Cho 2012; Xu et al. 2017). Liver tissue engineering has raised new possibilities by designing biocompatible scaffolds in vitro or in vivo. This can provide the optimal environment similar to the natural extracellular matrix for cell growth and cell differentiation (Griffith and Swartz 2006).

Xu et al. (2017) seeded mesenchymal stem cells (MSCs) on regenerated silk fibroin (RSF) matric and then placed it on the liver surface of mice. This system showed great biocompatibility of MSCs and RSF. MSCs differentiated hepatocyte-like cells in 10 days in vitro. Furthermore, MSCs seeded on RSF demonstrated better liver functions indicating that it can be a potential candidate for liver tissue engineering (Xu et al. 2017).

Grant et al. (2019) electrospun collagen scaffold in the environment of human liver ECM and cultured THLE-3 hepatocytes on it. Several tests have been performed to monitor the growth rate of cells. It was found that hepatocytes grew along with the production of albumin. Furthermore, the immunohisto-chemistry of the scaffold confirmed the retention of protein in it. This study was useful to depict protein–polymer scaffold potential in the environment of ECM.

Nerve

The nervous system is a highly complex part that can be divided into two categories:

1. The central nervous system (CNS) includes the brain, spinal cord, optic, olfactory, and auditory systems.
2. The peripheral nervous system (PNS) consists of main nerves and is subdivided into three parts (a) somatic, (b)

autonomic, and (c) enteric. The somatic nervous system deals with voluntary movements. The autonomic and enteric nervous system deals with involuntary movements. The autonomic nervous system is further divided into sympathetic (mobilize energy in the emergency) and parasympathetic (activates in a relaxed state). The enteric nervous system controls the digestive system (Schmidt and Leach 2003).

Approximately $150 billion is spent annually on the treatment of nerve injuries (Taylor et al. 2008). Nerve regeneration is a perplexing phenomenon. When a nerve is injured, it has to be treated surgically either by coaptation or by nerve grafting. The former can be done only in case of small injuries where as the later is performed for major injuries. Autografts are considered as golden standards when they are more histo-compatible. However, due to their lower availability, secondary deformity, a requirement of more than one surgery, size mismatch between donor and receptor site limits its use (Xie et al. 2009). On the other hand, there are no restrictions on the availability of human cadaveric nerve allografts. However, it is more complex, highly expensive and comes with a high risk of immune- and disease-related complications (Moore et al. 2009).

Tissue-engineered constructs for nerve regeneration have proven to be helpful. Natural and synthetic materials are available to construct scaffolds. The natural materials are more biocompatible, produce less toxic effects, and enhance migration of support cells like Schwann cells. Certain variations in synthetic materials provide great mechanical strength and porosity (Evans 2001).

Ghasemi-Mobarakeh et al. (2008) fabricated and studied a biocomposite scaffold of PCL/Gelatin by varying its concentration. The potential of synthesized aligned and random fibers has been used in nerve tissue engineering. PCL/Gelatin biocomposite scaffold having a 70:30 concentration ratio was suitable for the promotion of nerve differentiation and proliferation. Furthermore, the inclusion of gelatin in PCL enhanced cell differentiation and proliferation as compared to pure PCL scaffold.

The effect of varying concentrations of PLGA, collagen and silk fibroin individually has been carried out to examine their potential in nerve tissue engineering. Wang et al. (2011) designed a novel biocomposite scaffold of PLGA-silk fibroin-collagen by varying concentrations. It has been found that the 50:25:25 concentration scaffold perform better than the rest of the concentrations when Schwann cells were seeded onto it and the cells showed remarkable growth in 11 days.

Skin

Skin deficiencies were treated by auto-grafts and allografts (Boyce and Lalley 2018). Autologous keratinocytes obtained from epithelium are supplanted on the insufficient site. Another technique to repair and rebuild the epidermal layer is to use the stem cells which can differentiate tissues (Vig et al. 2017). Recent advances led to the use of engineered tissues for repairing and rebuilding the skin. Several proteins have been electrospun to design suitable scaffolds for skin tissue engineering.

Wound healing

Skin is the largest organ of the human body. The tissues under the skin repair themselves when the skin is damaged through a complex process called wound healing. It is a highly dynamic process consisting of five stages (Murphy and Evans 2012; Reiger et al. 2013) (Fig. 4).

The first step known as wounding is to get tissue damage. Once the damage has occurred, the blood flows out as an inflammatory response (Gonzalez et al. 2016). It is essential to restrict local hemorrhage to prevent blood loss. Leukocytes are influxes in minutes of injury at the wounded site, to cleanse cell debris, to kill the invading microorganism (Shaw and Martin 2009), for tissue degradation, and tissue formation (Eming et al. 2007). Platelets and thrombocytes activate and then accumulate at the site of injury. Platelets and thrombocytes are plugs together with the help of collagen and thrombin through a series of reactions. This process is commonly known as hemostasis. This results in the formation of a fibrin network consisting of a large number of insoluble fibrin fibers responsible for restoring hemostasis. It organizes an interim matrix to stimulate the migration of cells and for the proliferation of fibroblasts (Gonzalez et al. 2016).

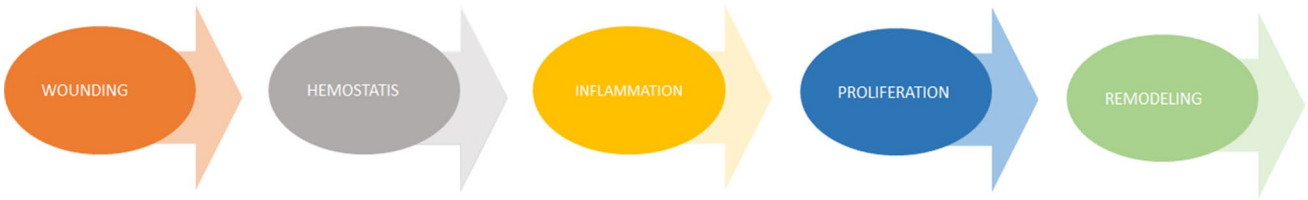


Fig. 4 Five stages of wound-healing process

The wounded area swells on blood clotting and reddens, a process known as inflammation, appears. During this phase, the pathogens damaged cells, bacteria are removed from the wounded area, and the essential supply of white blood cells, enzymes, nutrients, growth factors is simulated. In the proliferation phase, the wound has to be rebuilt with collagen and extracellular matrix. A new network of blood vessels that supplies essential nutrients and oxygen to the cells, known as angiogenesis, is triggered. Remodeling is the final phase in which collagen is remodeled from type III to type I. The disorganized collagen formed during the proliferative phase is rearranged and water is reabsorbed from them to stick collagen fibers closely. This phase occurs when the wound has been completely closed. In this phase, the cells which are no longer needed in the area are migrated or killed by the process of apoptosis, and the number of blood vessels decreases (Gonzalez et al. 2016). It has been reported that tropoelastin and collagen-based nanofibers scaffolds support adherence, growth and migration of cells in vitro (Rnjakkovacina et al. 2012).

A nanofibrous mat of MUGA fibroin has been prepared and seeded L929 fibroblast cell onto it (Srivastava et al. 2019). After 5 days of incubation, cells proliferated to a higher extent and covered the entire scaffold with monolayer formation whereas it does not happen in muga cast film. This showed that muga fibroin nanofibrous mat is more cytocompatible and promotes cell attachment and proliferation as compared to muga cast film. The performance of non-mulberry tasar silk fibroin-based nanofibers scaffolds was also enhanced by functionalization with silver nanoparticles (Srivastava and Purwar 2016).

Wound dressings were fabricated from materials of natural origin in the early times. Several researchers have used artificial materials to fabricate wound dressings that are capable of rapid hemostasis and are biocompatible with antimicrobial properties (Liu et al. 2017). Electrospun nanofibers dressings have significant importance in this field. These materials can be incorporated with nanomaterials having antimicrobial properties to treat and heal wounds (Chen and Liu 2015). Such composites are better than the conventional dressings being capable of high absorption of exudates from the wound site. They are also porous which promotes cell proliferation by exchanging nutrients and gases efficiently and protects the injured tissue by covering it (Zhang et al. 2017a, b). Several materials have been electrospun for wound dressing. They are mentioned (along with their electrospinning parameters and diameters of the nanofibers obtained) in Table 2.

Collagen nanofibrous mats provide a large surface area to volume ratio, are antibacterial, biodegradable, biocompatible, and show weak antigenicity. Owing to these reasons, collagen serves as a potential therapeutic agent for drug delivery, nanoparticles for gene delivery, matric for cell culture system, and transdermal delivery (Lee et al. 2001). Several methodologies have been proposed to electrospun collagen fiber and nanofibrous mats. Nanofibrous mat and fiber have been electrospun by blending collagen and zein in aqueous acetic acid at voltage 20 kV, air distance 15 cm, and feed rate 0.75 mL/h (Lin et al. 2012). The nanofiber diameter could be altered by varying the blending ratio of zein and collagen protein polymer. The nanofibrous mat fabricated was found to be biodegradable, biocompatible, surface wettable, and antibacterial indicating that it can be a potential substitute for wound-dressing material.

PLGA and collagen have been co-electrospun using HFP solvent to form a biomimetic nanofibrous mat with a diameter range of 150–650 nm (Liu et al. 2010). It has been reported that human fibroblast cultured on the mat shows comparable adhesion and proliferation in the same incubation time than that of a tissue culture plate. Moreover, 50:50 blending concentration of PLGA and collagen have shown to increase the rate of cell proliferation and adhesion in comparison to other blending ratios. The nanofibrous mat showed no sign of cytotoxicity. Furthermore, an open wound-healing test was performed on a rat with one wound covered with the synthesized PLGA/collagen nanofibrous mat and with a commercial dressing. It was found that the PLGA/collagen nanofibrous mat-covered wound healed faster than the commercial dressing.

Eri silk fibroin (ESF) has been blended with PLA to form a nanofibrous mat with an average diameter of 502 nm (Shanmugam et al. 2015). Tetracycline hydrochloride (TCH) was added before electrospinning. An open wound-healing test was performed and it was observed that there was a remarkable growth of epidermal skin with sebaceous glands, hair follicles, and granulation tissue on the 3rd day after surgery.

Bio-mimic fibrous fibrinogen nanofibers scaffold has been constructed with fiber having a diameter in the range of 120–610 nm for wound dressing (McManus et al. 2006a, b). Fibers were electrospun with a flow rate of 1.8 mL/h and applied voltage 22 kV. The scaffold mimicked the same mechanical strength as that of native tissue.

Bilayer novel wound-dressing material was constructed with commercial polyurethane on the outer side and electrospun gelatin/keratin nanofibrous mat on the inner side (Yao et al. 2017). The gelatin/keratin nanofibrous mat electrospun had a homogeneous fibrous structure with bead-free morphology and a diameter range of 160.4 ± 45.8 nm. L929 fibroblasts were cultured on the bilayer membrane. The cells adhered better and the residues secreted by the nanofibrous mat have been found to increase the rate of proliferation of

fibroblast cells. The bilayer showed no sign of cytotoxicity. Wound-healing ability was scrutinized without keratin.

Uniform fibers (size: 0.4–2.1 μm) have been synthesized by blending polyurethane with gelatin (Kim et al. 2009). NIH-3T3 fibroblast cells were cultured over fibers that adhered and proliferated well with an increase in gelatin content. The blended fibers showed good hydrophilicity and surface cell recognition sites over synthetic materials.

PCL/Gelatin nanofibrous mat of smooth morphology having a diameter range of 300–600 nm has electrospun directly onto polyurethane dressing (Chong et al. 2007). Fibroblast cells were cultured on both sides of the dressing material unlike the traditional way of culturing on only one side. It was observed that the cells adhered and proliferated in 7 days span forming a three-dimensional analog.

Keratin-based biomaterials are being used in abundance for biomedical applications due to their biodegradability, biocompatibility, high mechanical strength, and natural availability. Yuan et al. (2012a, b) constructed one such novel wound dressing mat of PHBV/m-keratin (ratio 7:3) (modified keratin). The average diameter of the mat was found to be at 720 ± 124 nm. NIH-3T3 fibroblast cells showed higher viability on PHBV/keratin on day 5 than on pure PHBV mat. It was found that the keratin blended in PHBV nanofibers accelerates the growth of cells and increased the efficiency of wound healing.

Biocomposite of polyurethane (PU) and keratin have co-electrospun and coated with silver nanoparticles (AgNPs) (Wang et al. 2015). Keratin was chosen to be blended with PU because of its biodegradability, mechanical durability and easy availability. Keratin was found to increase the rate of fibroblast cell proliferation and adhesion over pristine PU. The dressing of PU/K/AgNPs reduced the size of the wound in 9 days compared to other spongy dressings. AgNPs inclusion in the biocomposite made it antibacterial and mild inflammatory.

Epidermal growth factor (EGF) has been incorporated in electrospun silk mats and tested for wound-healing ability in vitro which heals the same as the wound in vivo (Schneider et al. 2009). Biofunctionalized silk mats containing growth factors were found to be biocompatible, biodegradable and reduced the wound by 90%.

Ju and coworkers reported the reduction in the amount of interleukin IL-1α (a pro-inflammatory cytokine) in 7 days when the burn wound was treated with silk fibroin (Ju et al. 2016). Medical gauze on the other hand showed higher expression of Interleukin 1 alpha (IL-1α), Interleukin6 (IL-6), and Interleukin 10 (IL-10). This demonstrated that the silk mat is effective in reducing the wound size. An antibacterial biocomposite and biodegradable nanofibrous scaffold have been fabricated by incorporating silver nanoparticles in silk fibroin mat for applications in wound dressing and healing (Uttayarat et al. 2012).

Biocomposite nanofibrous scaffold has been electrospun by incorporating silver nanoparticles in zein solution which resulted in a nanofibrous mat with a diameter of 350–500 nm (Dashdorj et al. 2015). NIH-3T3 fibroblasts cells were cultured onto the biocomposite Zein/Ag mat and found to be proliferated well with time. They showed maximum growth after 6 days of incubation. The cells' growth was higher on the biocomposite Zein/Ag mat in comparison to pure zein. This showed the biocomposite nanofibrous mat is non-toxic for the cells. The incorporation of silver nanoparticles led to the induction of antibacterial properties in the biocomposite scaffold and made it potentially viable to be used as a wound dressing.

Poly(vinylalcohol)-stilbazol quaternized (PVA-SbQ) blended with zein in acetic acid solution to design a biocompatible hydrophilic scaffold membrane for wound dressing (Cui et al. 2015). PVA-SbQ was added to enhance cross-linking, tensile strength, surface wettability and in vitro degradation. A drug vaccarin was incorporated into the scaffold membrane to heal the wound faster due to its medicinal properties. PVA-SbQ/zein and PVA-SbQ/zein/vaccarin showed cell viability and cell attachment. The addition of vaccarin into the scaffold membrane enhanced the rate of cell proliferation, an essential requirement of a wound-healing material.

Biocomposite nanofibrous having a diameter ranging 400–700 nm scaffold was electrospun by blending polyurethane (PU) with cellulose acetate (CA) and zein (Unnithan et al. 2013). A drug, streptomycin sulfate was incorporated to induce antibacterial properties in the nanofibrous scaffold matric. The resulting formulation enhanced the blood clotting and activated platelets in comparison to pristine polyurethane fiber. 3T3-L1 fibroblasts cells were cultured on the biocomposite mat PU/CA/zein/drug which showed significantly higher proliferation as compared to control. Zein and Cellulose acetate addition to polyurethane increased its hydrophilicity and bio-activity. This formulation has been found to create a moist environment near the wound resulting in an increase in cell adhesion and proliferation on the nanofibrous scaffold PU/CA/zein/drug as compared to pristine PU fibers.

Drug delivery

Electrospun proteins nanofibers are making advances in medical science because of their exceptional therapeutic uses. Several natural proteins, their corresponding blends and composites have been reported and nano-formulated via electrospinning for drug delivery. Small drug molecules

have also been incorporated into these protein fibers for drug delivery. For drug delivery, drug molecules are first loaded into protein feedstock solutions and then the drug-loaded fibers are electrospun (Zhang and Qian 2011). The high voltage applied for electrospinning for fibers does not alter the activity of the drug (Geert et al. 2003). Gelatin/PVA nanofibers have been fabricated for delivering Raspberry Ketone (RK) (Yang et al. 2007a, b). It was found that the drug showed burst release in the first hour and then no significant change in release rate was observed after 2 h. The drug release rate was found to be altered by either changing ratio of Gelatin/PVA or altering the amount of drug loaded in the nanofiber.

Nanofibers of cod (Gadus morhua) have been synthesized using sarcoplasmic proteins (Stephansen et al. 2014). The nanofibers showed variation in morphology with varying concentrations of protein. When the concentration was as low as 50 mg/mL, particle-like structures were formed. Increasing the concentration of protein to 75 mg/mL showed fiber formation, although particles and beads-like structures were persisted. At this concentration, the diameter of the nanofiber was between 100 and 700 nm. Increasing the concentration to 125 mg/mL, the fiber diameter was increased to 1000 nm. The beads and particle-like structures were found to be absent at the concentration of 125 mg/mL. This concentration was then chosen for encapsulation of a dipeptide of Alanine and Tryptophan (Ala-Trp). The release rates were determined in both gastric buffer and intestinal buffer. It was found that burst release of 30% occurred in the first minute, followed by the next 40% release in 30 min for gastric buffer and 15 min for intestinal buffer.

Nanofibers having smooth morphology were prepared using a blend solution of soy protein isolate (SPI)/Poly(ethylene oxide) (PEO) and poly(lactic acid) (PLA) (Vega-Lugo and Lim 2009). The SPI and PLA polymer solution were electrospun at the feed rate of 0.04 mL/min, tip-to-collector distance of 26 cm and the applied voltage of 20–30 kV. The drug allyl isothiocyanate (AITC) was incorporated into a fiber-forming solution. AITC is an antimicrobial compound naturally found to be present in plants of the *Cruciferae* family. Drug release of AITC was observed to be negligible in dry experimental conditions but showed an increment in release rate with an increase of water content under experimental conditions.

Water-resistant zein nanofibers have been synthesized and modified by citric acid and cross-linked using sodium hydroxide. (Jiang and Yang 2011) These cross-linked zein protein based nanofibers exhibits much higher dry as well as wet tensile strength of 183% and 448% respectively. Moreover, cross-linking of zein nanofibers increased its drug loading efficiency to 58% with a sustained drug release rate. Non-mulberry muga silk fibroin nanofiber mat of average diameter 160 nm has been fabricated using electrospinning. These mats were further loaded with antibiotic (gentamycin sulfate) and their drug-releasing behavior was studied (Srivastava and Purwar in 2016) (Table 3).

Coaxial electrospinning of proteins

Electrospinning has revealed itself to be useful to synthesize hollow tubular or core–shell-like structures at the submicron and nano-range. A slight modification in the conventional electrospinning process results in the formation of cable-like fiber known as coaxial electrospinning (Qin 2017). In this process, two or more polymer solutions are fed together by coaxial capillaries sharing the same axis. This allows the injection of one solution (core solution) to get drawn into the outer solution (sheath solution) to form a continuous coated or hollow tubular nanofiber (Díaz et al. 2008). These coaxial nanofibers have more advantages over conventional round fibers in biomedical science summarized as follows:

a) The structure of these hollow tubular coaxial fibers matches exactly that of the natural extracellular matrix.
b) The coaxial fibers properties can be easily tuned by selecting appropriate polymers to impart desired properties. A more biocompatible polymer can be chosen for the shell and a less biocompatible polymer can be fabricated as a core.
c) By tuning the thickness of both core and sheath, drug delivery can be made sustained and controlled irrespective of normal nanofibers which fail to perform better in drug delivery due to the burst release of drugs before reaching the target site.
d) The drugs remain bioactive inside these coaxial nanofibers as they remain protected inside the sheath (Lu et al. 2016).

Because of the above advantages, several proteins have been electrospun coaxially to be used in biomedical science for fabricating scaffolds for tissue engineering, drug delivery, and wound healing. These are listed in Table 4.

Table 3 Protein electrospun nanofibers for drug release

Protein	Other polymer or material	Solvent	Electrospinning parameters	Diameter of fiber	Drug loaded	References
Gelatin	PVA	Water	Voltage = 30 kV TCD = 10 cm	Between 177 and 340 nm	Raspberry ketone	Yang et al. (2007a, b)
FSP	–	HFIP	Flow rate = 0.02 mL/min Voltage = 20 kV TCD = 15 cm	Around 1000 nm	Ala-Trp	Stephansen et al. (2014)
Muga silk fibroin	–	BMIMAc ionic liquid	Flow rate = 0.5 mL/h TCD = 8 cm	160 nm	Gentamycin sulfate	(Srivastava and Purwar 2016)
Soy protein	PEO and PLA	Chloroform, DMF	Flow rate = 0.04 mL/min Voltage = 20–30 kV TCD = 26 cm	From 200 nm to 2 µm	AITC	(Vega-Lugo and Lim 2009)
Silk fibroin	–	FA	Flow rate = 1.5 mL/h Voltage = 14.1 kV TCD = 15 cm	98.2–400.56 nm	Lafutidine	Nangare et al. (2020)
Zein	CA	Ethanol	Flow rate = 2 mL/h Voltage = 19.5 kV TCD = 20 cm	Around 400–500 nm	Diclofenac	Jiang and Yang (2011)

In Proteins: *FSP* Fish Sarcoplasmic Protein

In other polymer or material: *PVA* polyvinyl alcohol, *PEO* poly(ethylene oxide), *PLA* poly(lactic acid), *CA* citric acid

In solvent: *DMF* dimethylformamide, *HCl* hydrochloric acid, *HFIP* 1,1,1,3,3,3-hexafluoro-2-propanol, *FA* formic acid, *BMIMAc* 1-butyl-3-methyl imidazolium acetate

In electrospinning parameter: *TCD* tip-to-collector distance

In drug loaded: *AITC* allyl isothiocyanate, *Ala-Trp* alanine-tryptophan

Emulsion electrospinning of proteins

Emulsions are the dispersion of minute droplets of one solvent into another where both of them are immiscible. The different morphology of fibers can be obtained by electrospinning emulsions. To electrospun core–shell nanofibers using emulsions, the basic apparatus is the same as used in blend electrospinning. In this process, the polymer is dissolved in organic solvent making a continuous phase and the precursor is dissolved in water. Organic solvent forming a continuous phase evaporates more rapidly increasing the viscosity of the solution. This change in viscosity introduces the viscosity gradient which further acts as the driving force for the precursor in the water phase to migrate to the center of the jet to form the core and the polymer in an organic solvent as the shell (Xu et al. 2005) (Fig. 5 and Table 5).

Nanofibers based wound-dressing products available in the market

In vitro tests of several electrospun polymers have been performed to evaluate their usability in the biomedical field. Only a few materials have been tested for biocompatibility in vivo. Despite significant efforts in this field, only a few materials have been developed. Commercialization is required for such nanofibers to be used on a large scale to meet the demands in medical science. The trade name of products commercially available and the manufacturer name (enclosed in brackets) for wound dressing are Integra™ (Integra Life Sciences), Apligraf® (Novartis), Nanocell™ (Thai Nano Cellulose), Kerlix™ AMD™ (Kendall), DermaFuse™ (Mo-Sci Corporation, USA), Tegaderm™ (3 M Company), TransCyte® (Advanced Tissue Sciences), Dermagraft-TC® (Advanced Tissue Sciences, Inc.), Permacol™ (Covidien), AlloDerm® (Candavers acellular matrix),

Table 4 Coaxially electrospun protein for biomedical science

Core	Core solvent	Electrospinning parameters for core	Shell	Shell solvent	Electrospinning parameters for shell	Drug loaded (if any)	Application	References
BSA-dextran	Water	Flow rate = 0.1–0.8 mL/h Voltage = 16 kV TCD = 20 cm	PCL/PEG	DMF/chloroform	Flow rate = 1 mL/h Voltage = 16 kV TCD = 20 cm	BSA	Tissue Engineering	Jiang et al. (2006)
DEX	Plants oil	–	SF/PEO	Water	–	DEX	Endothelial tissue engineering	Chen et al. (2015a, b)
EGF, insulin, hydrocortisone, and retinoic acid	5% BSA	Flow rate = 0.3 mL/h Voltage = 15 kV TCD = 12 cm	Gelatin/PLLCL	HFP	Flow rate = 1 mL/h Voltage = 15 kV TCD = 12 cm	EGF, insulin, hydrocortisone, and retinoic acid	Skin tissue engineering	Jin et al. (2013)
Gelatin	AA	Flow rate = 0.04 mL/h Voltage = 10–16 kV TCD = 12 cm	PCL	AA	Flow rate = 0.04 mL/h Voltage = 10–16 kV TCD = 12 cm		Tissue engineering	Gulfam et al. (2011)
PCL	HFP	Flow rate = 0.4 mL/h Voltage = 18 kV TCD = 15 cm	Gelatin	HFP	Flow rate = 1 mL/h Voltage = 18 kV TCD = 15 cm		Bone tissue engineering	Kai et al. (2012)
PGS	HFP	Flow rate = 0.1 mL/h Voltage = 20 kV TCD = 12 cm	Fibrinogen	*HFP*	Flow rate = 0.5 mL/h Voltage = 20 kV TCD = 12 cm		Cardiac tissue engineering	Ravichandran et al. (2013a, b)

In core: *BSA* bovine serum albumin, *PGS* poly(glycerol sebacate), *DEX* dexamethasone

In core solvent: *HFP* 1,1,1,3,3,3-hexafluoro-2-propanol, *BSA* bovine serum albumin, *AA* acetic acid

In shell: *PCL* poly(ε-caprolactone), *PLLCL* poly(L-lactic acid)-co-poly-(ε-caprolactone), *SF* silk fibroin, *PEO* polyethylene oxide, *PEG* poly(ethylene glycol), *PLGA* poly(lactic-co-glycolic acid)

In shell solvent: *HFP* 1,1,1,3,3,3-hexafluoro-2-propanol, *DMF* dimethylformamide

In parameters: *TCD* tip-to-collector distance

ChitoFlex® (HemCon Med. Tech. Inc.), and Biobrane™ (Smith & Nephew) (Advanced wound care world market Prospects 2011–2021).

Conclusions and future perspective

The present review focused on the parameters to electrospun plant proteins such as soy protein, zein and various animal proteins such as casein, collagen, elastin, fibrinogen, keratin, gelatin, mulberry silk fibroin, and non-mulberry silk such as muga, eri, tasar and tussah. It also covers their use in biomedical applications of regenerative medicine (tissue engineering and wound dressing) and drug delivery. Immense

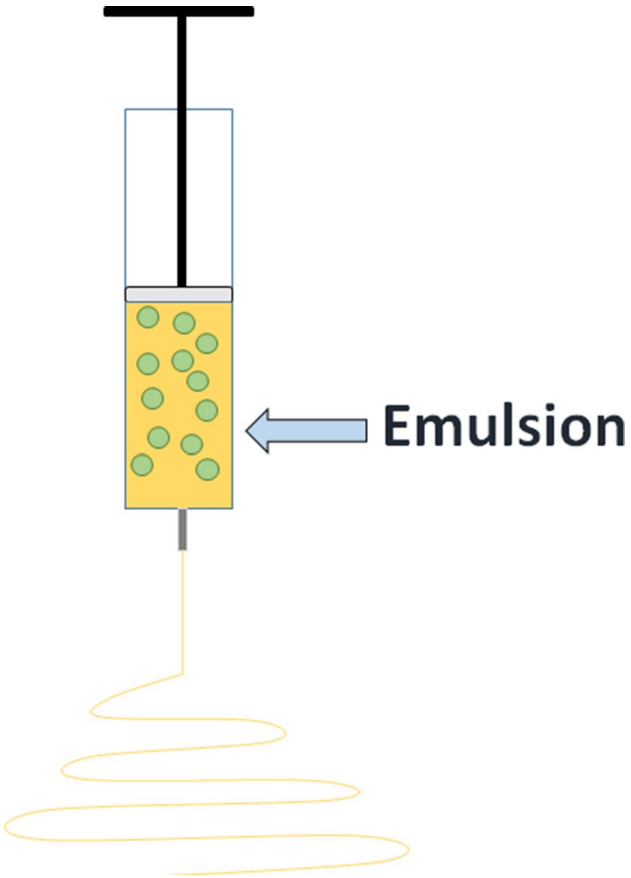


Fig. 5 Emulsion electrospinning setup

growth and research of electrospinning techniques have been applied to generate fibers at the nanoscale range in the last decade. These fibers having a large surface-to-volume ratio and superior mechanical properties. Various polymers and their blends have been electrospun for their use in diverse fields. Protein polymer (both plant-derived and animal-derived) and protein composites have been explored for use in biomedical applications such as tissue engineering and regenerative medicine (TERM), drug delivery, wound dressing, and enzyme immobilization. This is due to their biocompatibility, biodegradability, ease in availability, renewability. The size, morphology, and weight of such fibers can be easily tuned to attain desired specifications. The natural degradation of these proteins decreases the accumulation of by-products which is yet another advantage for human health over synthetic polymer systems.

The present review also covers the technical advancements which have been taken place in recent times in electrospinning techniques. Coaxial and emulsion electrospinning has enabled us to fabricate nanofibers with desired properties and can be used for better results in healthcare applications. The review includes the commercially available wound-dressing materials. However, there is still scope for more commercialization of electrospun nanofibers for tissue engineering, regenerative medicine and drug delivery. Electrospun nanofiber of recombinant proteins and peptides has better advantages as a drug delivery vehicle and regenerative medicine over nanofibers of natural proteins. Multifunctional scaffolds can be designed to work in more than one area. Researchers have successfully performed cephalosomatic anastomosis in a human model and head transplant on human cadavers, but there is still scope to achieve such transplantations in living humans.

Drug release and tissue engineering are inter-related areas of biomedical. The porous structure of proteins has made a tactile approach to be used in tissue engineering and regenerative medicine (TERM). However, research has to be done on controlled release of the drug, oxygen release and growth factors that are significant in drug delivery. The burst release of drugs with nanofibrous mats is still unexplored. To overcome this, long-lasting slow degrading biomimetic nanofibrous materials have to be explored to be effectively used in drug delivery. Moreover, many studies of using electrospun fibers and composite mats in biomedical applications have been done in vitro and very little research has been performed on animals in vivo. Scientists and biologists have to pay considerable attention to focus more on in vivo experiments to explore these electrospun nanomaterials and find a way to generate them on large scale to be used commercially.

Table 5 Proteins' emulsions electrospun for biomedical applications

Organic	Aqueous	Solvent	Electrospinning parameters	Diameter of fiber	Applications	References
Gelatin/PLGA	EGF	Chloroform/acetone and AA	Flow rate = 0.4 mL/h Voltage = 16 kV TCD = 10 cm	175 ± 45 nm	Wound healing	Norouzi et al. (2014)
PLGA	Collagen	Chloroform/toluene/water	Flow rate = 0.012 mL/min Voltage = 17 kV TCD = 15 cm	600–2000 nm	Bone tissue engineering	Wei et al. (2011)
PLGA	Gelatin/cefradine	Chloroform/DMF/water	Flow rate = 15 µL/min Voltage = 17.5 kV TCD = 25–27 cm		Drug delivery	Hu et al. (2012)
PCL/ketoprofen	Gelatin	Chloroform/methanol	Flow rate = 18 µL/min Voltage = 18 kV TCD = 13 cm		Wound healing	Basar et al. (2017)
PCL	Silk fibroin	Chloroform and formic acid	Flow rate = 4.0–6.0 µL/min Voltage = 22–23 kV TCD = 10–12 cm	336.3 ± 60.4 nm; 212.6 ± 43.4 nm and 370.1 ± 111.43 nm	Tissue engineering	Roy et al. (2018)

In organic: *PLGA* poly(lactic-co-glycolic acid), *PCL* poly(ε-caprolactone)

In aqueous: *EGF* epidermal growth factor

In solvent: *DMF N,N*-dimethyl formamide, *AA* acetic acid

Acknowledgements The authors acknowledge the support provided under the DST-FIST Grant No.SR/FST/PS-I/2018/48 of Government of India.

References

Acun A, Hasirci V (2014) Construction of a collagen-based, split-thickness cornea substitute. J Biomater Sci Polym Ed 25(11):1110–1132

Advanced wound care world market Prospects 2011–2021. https://www.visiongain.com

Agarwal S, Wendorff JH, Greiner A (2008) Use of electrospinning technique for biomedical applications. Polymer 49:5603–5621

Aguilar-Vázquez G, Loarca-Pina G, Figueroa-Cárdenas JD, Mendoza S (2018) Electrospun fibers from blends of pea (*Pisum sativum*) protein and pullulan. Food Hydrocoll 83:173–181

Amiraliyan N, Nouri M, Kish MH (2009) Electrospinning of silk nanofibers. I. An investigation of nanofiber morphology and process optimization using response surface methodology. Fibers Polym 10(2):167–176

Andiappan M, Sundaramoorthy S, Panda N et al (2013) Electrospun eri silk fibroin scaffold coated with hydroxyapatite for bone tissue engineering applications. Prog Biomater 2(1):1–11

Baek J, Sovani S, Glembotski NE et al (2016) Repair of avascular meniscus tears with electrospun collagen scaffolds seeded with human cells. Tissue Eng Part A 22(5–6):436–448

Basar AO, Castro S, Torres-Giner S, Lagaron JM, Sasmazel HT (2017) Novel poly(ε-caprolactone)/gelatin wound dressings prepared by emulsion electrospinning with controlled release capacity of Ketoprofen anti-inflammatory drug. Mater Sci Eng: C 81:459–468. https://doi.org/10.1016/j.msec.2017.08.025

Baumgarten PK (1971) Electrostatic spinning of acrylic microfibers. J Colloid Interface Sci 36(1):71–79

Bhardwaj N, Kundu SC (2010) Electrospinning: a fascinating fiber fabrication technique. Biotechnol Adv 28(3):325–347

Bhattacharjee P, Naskar D, Kim HW, Maiti TK, Bhattacharya D, Kundu SC (2015) Non-mulberry silk fibroin grafted PCL nanofibrous scaffold: promising ECM for bone tissue engineering. Eur Polym J 71:490–509

Biazar E, Baradaran-Rafii A, Heidari-keshel S, Tavakolifard S (2015) Oriented nanofibrous silk as a natural scaffold for ocular epithelial regeneration. J Biomater Sci Polym Ed 26(16):1–13

Boland ED, Matthews JA, Pawlowski KJ, Simpson DG, Wnek GE, Bowlin GL (2004) Electrospinning collagen and elastin: preliminary vascular tissue engineering. Front Biosci 9:1422–1432

Bonadies I (2019) Electrospun nanofibrous filtration membranes. In: Thomas S, Pasquini D, Leu SY, Gopakumar D (ed) Nanoscale materials in water purification, pp 231–46

Bosworth LA, Alam N, Wong JK (2013) Investigation of 2D and 3D electrospun scaffolds intended for tendon repair. J Mater Sci Mater Med 24(6):1605–1614

Boyce ST, Lalley AL (2018) Tissue engineering of skin and regenerative medicine for wound care. Boyce Lalley Burns Trauma 6(4):1–10

Buttafoco L, Kolkman NG, Engbers-Buijtenhuijs P et al (2006) Electrospinning of collagen and elastin for tissue engineering applications. Biomaterials 27:724–734

Casasola R, Thomas NL, Trybala A, Georgiadou S (2014) Electrospun poly lactic acid (PLA) fibres: effect of different solvent systems on fibre morphology and diameter. Polymer 55(18):4728–4737

Casper CL, Yang W, Farach-Carson MC, Rabolt JF (2007) Coating electrospun collagen and gelatin fibers with perlecan domain I for increased growth factor binding. Biomacromol 8:1116–1123
Chen DWC, Liu SJ (2015) Nanofibers used for delivery of antimicrobial agents. Nanomedicine (london) 10(12):1959–1971
Chen JP, Su CH (2011) Surface modification of electrospun PLLA nanofibers by plasma treatment and cationized gelatin immobilization for cartilage tissue engineering. Acta Biomater 7:234–243
Chen HC, Jao WC, Yang MC (2008a) Characterization of gelatin nanofibers electrospun using ethanol/formic acid/water as a solvent. Polym Adv Technol 20(2):98–103
Chen JP, Chang GY, Chen JK (2008b) Electrospun collagen/chitosan nanofibrous membrane as wound dressing. Colloids Surf A Physicochem Eng Aspects 313–314:183–188
Chen J, Yan C, Zhu M (2015a) Electrospun nanofibrous SF/P(LLA-CL) membrane: a potential substratum for endothelial keratoplasty. Int J Nanomed 10:3337–3350
Chen W, Li D, El-Shanshory A et al (2015b) Dexamethasone loaded core–shell SF/PEO nanofibers via green electrospinning reduced endothelial cells inflammatory damage. Colloids Surf, B 126:561–568
Choktaweesap N, Arayanarakul K, Aht-Ong D, Meechaisue C, Supaphol P (2007) Electrospun gelatin fibers: effect of solvent system on morphology and fiber diameters. Polym J 39(6):622–631
Chong EJ, Phan TT, Lim IJ et al (2007) Evaluation of electrospun PCL/gelatin nanofibrous scaffold for wound healing and layered dermal reconstitution. Acta Biomater 3:321–330
Cooley JF (1902) Apparatus for electrically dispersing fluids. US Patent No: US692631A
Cui J, Qiu L, Qiu Y, Wang Q, Wei Q (2015) Co-electrospun nanofibers of PVA-SbQ and Zein for wound healing. J Appl Polym Sci 132:1–9
Dashdorj U, Reyes MK, Unnithan AR et al (2015) Fabrication and characterization of electrospun zein/Ag nanocomposite mats for wound dressing applications. Int J Biol Macromol 80:1–7
Díaz JE, Fernández-Nieves A, Barrero A, Márquez M, Loscertales IG (2008) Fabrication of structured micro and nanofibers by coaxial electrospinning. J Phys: Conf Ser 127:1–9
Elamparithi A, Punnoose AM, Kuruvilla S (2015) Electrospun type 1 collagen matrices preserving native ultrastructure using benign binary solvent for cardiac tissue engineering. Artif Cells Nanomed Biotechnol 44(5):1–8
Eming SA, Krieg T, Davidson JM (2007) Inflammation in wound repair: molecular and cellular mechanisms. J Investig Dermatol 127:514–525
Evans GRD (2001) Peripheral nerve injury: a review and approach to tissue engineered constructs. Anat Rec 263:396–404
Geert V, Iksoo C, Jef P et al (2003) Preparation and characterization of nanofibers containing amorphous drug dispersions generated by electrostatic spinning. Pharm Res 20:810–817
Gerardo-Nava J, Führmann T, Klinkhammer K et al (2009) Human neural cell interactions with orientated electrospun nanofibers in vitro. Nanomedicine 4(1):11–30
Ghasemi-Mobarakeh L, Prabhakaran MP, Morshed M, Nasr-Esfahani MH, Ramakrishna S (2008) Electrospun poly(ε-caprolactone)/gelatin nanofibrous scaffolds for nerve tissue engineering. Biomaterials 29:4532–4539
Ghezzi CE, Rnjak-Kovacina J, Kaplan DL (2015) Corneal tissue engineering: recent advances and future perspectives. Tissue Eng Part B 21(3):278–287
Ghosal K, Thomas S, Kalarikkal N, Gnanamani A (2014) Collagen coated electrospun polycaprolactone (PCL) with titanium dioxide (TiO2) from an environmentally benign solvent: preliminary physico-chemical studies for skin substitute. J Polym Res 21(5):410–417
Gilotra S, Chouhan D, Bhardwaj N, Nandi SK, Mandal BB (2018) Potential of silk sericin based nanofibrous mats for wound dressing applications. Mater Sci Eng, C 90:420–432
Gonzalez ACDO, Costa TF, Andrade ZDA, Medrado ARAP (2016) Wound healing—a literature review. An Bras Dermatol 91(5):614–620
Grant R, Hallett J, Forbes S, Hay D, Callanan A (2019) Blended electrospinning with human liver extracellular matrix for engineering new hepatic microenvironments. Sci Rep 9:1–12
Griffith LG, Swartz MA (2006) Capturing complex 3D tissue physiology in vitro. Nat Rev Mol Cell Biol 7(3):211–224
Gulfam M, Lee JM, Kim J, Lim DW, Lee EK, Chung BG (2011) Highly porous core-shell polymeric fiber network. Langmuir 27(17):10993–10999
Gunn J, Zhang M (2010) Polyblend nanofibers for biomedical applications: perspectives and challenges. Trends Biotechnol 28(4):189–197
Har-el Y-e, Gerstenhaber J, Brodsky R, Huneke RB, Lelkes PI (2014) Electrospun soy protein scaffolds as wound dressings: enhanced reepithelialization in a porcine model of wound healing. Wound Med 5:1–24
He W, Yong T, Teo WE, Ma Z, Ramakrishna S (2005) Fabrication and endothelialization of collagen-blended biodegradable polymer nanofibers: potential vascular graft for blood vessel tissue engineering. Tissue Eng 11(9/10):1574–1588
He T, Abbineni G, Cao B, Mao C (2010) Nanofibrous bio-inorganic hybrid structures formed through self-assembly and oriented mineralization of genetically engineered phage nanofibers. Small 6(20):2230–2235
He C, Xu X, Zhang F et al (2011) Fabrication of fibrinogen/P(LLA-CL) hybrid nanofibrous scaffold for potential soft tissue engineering applications. J Biomed Mater Res, Part A 97(3):339–347
Hoffman JR, Falvo MJ (2004) Protein—which is best? J Sports Sci Med 3(3):118–130
Hu J, Wei J, Liu W, Chen Y (2012) Preparation and characterization of electrospun PLGA/gelatin nanofibers as a drug delivery system by emulsion electrospinning. J Biomater Sci Polym Ed 24(8):972–985
Huang C, Chen S, Lai C, Reneker DH, Qiu H, Ye Y, Hou H (2006) Electrospun polymer nanofibres with small diameters. Nanotechnology 17(6):1558–1563
Jiang Q, Yang Y (2011) Water-stable electrospun zein fibers for potential drug delivery. J Biomater Sci Polym Ed 22(10):1393–1408
Jiang H, Hu Y, Zhao P, Li Y, Zhu K (2006) Modulation of protein release from biodegradable core-shell structured fibers prepared by coaxial electrospinning. J Biomed Mater Res B Appl Biomater 79(1):50–57
Jiang W, Kim BYS, Rutka JT, Chan WCW (2008) Nanoparticle-mediated cellular response is size-dependent. Nat Nanotechnol 3:145–150
Jiang Q, Reddy N, Yang Y (2010) Cytocompatible cross-linking of electrospun zein fibers for the development of water-stable tissue engineering scaffolds. Acta Biomater 6:4042–4051
Jin HJ, Chen J, Karageorgiou V, Altman GH, Kaplan DL (2004) Human bone marrow stromal cell responses on electrospun silk fibroin mats. Biomaterials 25:1039–1047
Jin G, Prabhakaran MP, Kai D, Ramakrishna S (2013) Controlled release of multiple epidermal induction factors through core–shell nanofibers for skin regeneration. Eur J Pharm Biopharm 85(3):689–698
Ju HW, Lee OJ, Lee JM et al (2016) Wound healing effect of electrospun silk fibroin nanomatrix in burn-model. Int J Biol Macromol 85:29–39
Kai D, Prabhakaran MP, Stahl B, Eblenkamp M, Wintermantel E, Ramakrishna S (2012) Mechanical properties and in vitro behavior of

nanofiber–hydrogel composites for tissue engineering applications. Nanotechnology 23(9):1–11
Kai D, Jin G, Prabhakaran MP, Ramakrishna S (2013) Electrospun synthetic and natural nanofibers for regenerative medicine and stem cells. Biotechnol J 8(1):59–72
Kalaoglu-Altan OI, Verbraeken B, Lava K et al (2016) Multireactive poly(2-oxazoline) nanofibers through electrospinning with crosslinking on the Fly. ACS Macro Lett 5(6):676–681
Ki CS, Baek DH, Gang KD, Lee KH, Um IC, Park YH (2005) Characterization of gelatin nanofiber prepared from gelatin–formic acid solution. Polymer 46(14):5094–5102
Kim SH, Nam YS, Lee TS, Park WH (2003) Silk fibroin nanofiber electrospinning, properties, and structure. Polym J 35(2):185–190
Kim C, Ngoc BTN, Yang KS et al (2007) Self-sustained thin webs consisting of porous carbon nanofibers for supercapacitors via the electrospinning of polyacrylonitrile solutions containing zinc chloride. Adv Mater 19:2341–2346
Kim SE, Heo DN, Lee JB et al (2009) Electrospun gelatin/polyurethane blended nanofibers for wound healing. Biomed Mater 4:1–12
Kim Y, Wu X, Oh JH (2020) Fabrication of triboelectric nanogenerators based on electrospun polyimide nanofibers membrane. Sci Rep 10:1–9
Kishimoto Y, Kobashi T, Yamanaka S, Morikawa H, Tamada Y (2017) Comparisons between silk fibroin non-woven fabrics electrospun using aqueous and formic acid solution. Int J Polym Mater Polym Biomater 67(7):462–467
Kong B, Mi S (2016) Electrospun scaffolds for corneal tissue engineering: a review. Materials 9(8):1–20
Koombhongse S, Liu W, Reneker DH (2001) Flat polymer ribbons and other shapes by electrospinning. J Polym Sci, Part b: Polym Phys 39(21):2598–2606
Koski A, Yim K, Shivkumar S (2004) Effect of molecular weight on fibrous PVA produced by electrospinning. Mater Lett 58(3–4):493–497
Larrondo L, St-John-Manley R (1981) Electrostatic fiber spinning from polymer melts. I. Experimental observations on fiber formation and properties. J Polym Sci Polym Phys Ed 19(6):909–920
Law JX, Liau LL, Saim A, Yang Y, Idrus R (2017) Electrospun collagen nanofibers and their applications in skin tissue engineering. Tissue Eng Regener Med 14:699–718
Lee JS, Cho SW (2012) Liver tissue engineering: recent advances in the development of a bio-artificial liver. Biotechnol Bioprocess Eng 17:427–438
Lee H, Kim G (2010) Biocomposites electrospun with poly(ε-caprolactone) and silk fibroin powder for biomedical applications. J Biomater Sci Polym Ed 21(13):1687–1699
Lee CH, Singla A, Lee Y (2001) Biomedical applications of collagen. Int J Pharm 221:1–22
Lee JS, Choi KH, Ghim HD, Kim SS, Chun DH, Kim HY, Lyoo WS (2004) Role of molecular weight of atactic poly(vinyl alcohol) (PVA) in the structure and properties of PVA nanofabric prepared by electrospinning. J Appl Polym Sci 93(4):1638–1646
Lee CH, Shin JH, Cho IH, Kang YM, Kim IA, Park KD, Shin JW (2005) Nanofiber alignment and direction of mechanical strain affect the ECM production of human ACL fibroblast. Biomaterials 26(11):1261–1270
Li M, Mondrinos MJ, Gandhi MR, Ko FK, Weiss AS, Lelkes PI (2005) Electrospun protein fibers as matrices for tissue engineering. Biomaterials 26:5999–6008
Li C, Vepari C, Jin HJ, Kim HJ, Kaplan DL (2006a) Electrospun silk-BMP-2 scaffolds for bone tissue engineering. Biomaterials 27:3115–3124
Li M, Guo Y, Wei Y, MacDiarmid AG, Lelkes PI (2006b) Electrospinning polyaniline-contained gelatin nanofibers for tissue engineering applications. Biomaterials 27:2705–2715
Li L, Li H, Qian Y et al (2011) Electrospun poly (ε-caprolactone)/silk fibroin core-sheath nanofibers and their potential applications in tissue engineering and drug release. Int J Biol Macromol 49:223–232
Li X, Zhang Q, Ye D et al (2017) Fabrication and characterization of electrospun PCL/*Antheraea pernyi* silk fibroin nanofibrous scaffolds. Polym Eng Sci 57(2):1–8
Lin J, Li C, Zhao Y, Hu J, Zhang LM (2012) Co-electrospun nanofibrous membranes of collagen and zein for wound healing. ACS Appl Mater Interfaces 4:1050–1057
Liu SJ, Kau YC, Chou CY, Chen JK, Wu RC, Yeh WL (2010) Electrospun PLGA/collagen nanofibrous membrane as early-stage wound dressing. J Membr Sci 355:53–59
Liu W, Thomopoulos S, Xia Y (2012a) Electrospun nanofibers for regenerative medicine. Adv Healthc Mater 1(1):10–25
Liu X, Lin T, Gao Y et al (2012b) Antimicrobial electrospun nanofibers of cellulose acetate and polyester urethane composite for wound dressing. J Biomed Mater Res Part b, Appl Biomater 100(6):1556–1565
Liu M, Duan XP, Li YM, Yang DP, Long YZ (2017) Electrospun nanofibers for wound healing. Mater Sci Eng, C 76:1413–1423
Lu Y, Huang J, Yu G et al (2016) Coaxial electrospun fibers: applications in drug delivery and tissue engineering. Wiley Interdiscip Rev Nanomed Nanobiotechnol 8(5):654–677
Luo L, Cui R, Qiao H et al (2014) High lithium electroactivity of electrospun $CuFe_2O_4$ nanofibers as anode material for lithium-ion batteries. Electrochim Acta 144:85–91
Mathers CD, Loncar D (2006) Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 3(11):1–20
McKee MG, Wilkes GL, Colby RH, Long TE (2004) Correlations of solution rheology with electrospun fiber formation of linear and branched polyesters. Macromolecules 37(5):1760–1767
McLaughlin CR, Tsai RJF, Latorre MA, Griffith M (2009) Bioengineered corneas for transplantation and in vitro toxicology. Front Biosci 14:3326–3337
McManus MC, Boland ED, Koo HP et al (2006a) Mechanical properties of electrospun fibrinogen structures. Acta Biomater 2(1):19–28
McManus MC, Boland ED, Simpson DG, Barnes CP, Bowlin GL (2006b) Electrospun fibrinogen: feasibility as a tissue engineering scaffold in a rat cell culture model. J Biomed Mater Res, Part A 81(2):299–309
McManus MC, Sell SA, Bowen WC, Koo HP, Simpson DG, Bowlin GL (2008) Electrospun fibrinogen-polydioxanone composite matrix: potential for in situ urologic tissue engineering. J Eng Fibers Fabr 3(2):1–10
Meinel AJ, Kubow KE, Klotzsch E et al (2009) Optimization strategies for electrospun silk fibroin tissue engineering scaffolds. Biomaterials 30:3058–3067
Moore AM, Ray WZ, Chenard KE, Tung T, Mackinnon SE (2009) Nerve allotransplantation as it pertains to composite tissue transplantation. Hand 4(3):239–244
Murphy PS, Evans GRD (2012) Advances in wound healing: a review of current wound healing products. Plast Surg Int 2012:1–8
Nagata S, Atkinson GM, Pestov D, Tepper G, Mcleskey JT Jr (2013) Electrospun polymer-fiber solar cell. Adv Mater Sci Eng 2013(3):1–6
Nangare S, Dugam S, Patil P et al (2020) Silk industry waste protein: isolation, purification and fabrication of electrospun silk protein nanofibers as a possible nanocarrier for floating drug delivery. Nanotechnology 32:035101
Nemeno-Guanzon JG, Lee S, Berg JR et al (2012) Trends in tissue engineering for blood vessels. J Biomed Biotechnol 2012:1–14
Neppalli R, Causin V, Marigo A, Meincken M, Hartmann P, van Reenen AJ (2013) Effect of electrospun ethylene vinyl alcohol

copolymer (EVOH) fibres on the structure, morphology, and properties of poly (lactic acid)(PLA). Polymer 54:5909–5919
Norouzi M, Shabani I, Ahvaz HH, Soleimani M (2014) PLGA/gelatin hybrid nanofibrous scaffolds encapsulating EGF for skin regeneration. J Biomed Mater Res, Part A 103(7):2225–2235
Noy A, Miller AE, Klare JE, Weeks BL, Woods BW, DeYoreo JJ (2002) Fabrication of luminescent nanostructures and polymer nanowires using dip-pen nanolithography. Nano Lett 2(2):109–112
Palakkan AA, Hay DC, Anil-Kumar PK, Kumar TV, Ross JA (2013) Liver tissue engineering and cell sources: issues and challenges. Liver Int 33(5):666–676
Panda NN, Biswas A, Pramanik K, Jonnalagadda S (2014) Enhanced osteogenic potential of human mesenchymal stem cells on electrospun nanofibrous scaffolds prepared from eri-tasar silk fibroin. J Biomed Mater Res B Appl Biomater 103(5):1–12
Paşcu EI, Stokes J, McGuinness GB (2013) Electrospun composites of PHBV, silk fibroin and nano-hydroxyapatite for bone tissue engineering. Mater Sci Eng C 33(8):4905–4916
Phu D, Wray LS, Warren RV, Haskell RC, Orwin EJ (2011) Effect of substrate composition and alignment on corneal cell phenotype. Tissue Eng Part A 17(5–6):799–807
Qin X (2017) Coaxial electrospinning of nanofibers. In: Afshari M (ed) Electrospun nanofibers, pp 41–71
Qureshi UA, Khatri Z, Ahmed F, Khatri M, Kim IS (2017) Electrospun zein nanofiber as a green and recyclable adsorbent for the removal of reactive black 5 from the aqueous phase. ACS Sustain Chem Eng 5(5):4340–4351
Ramakrishna S, Fujihara K, Teo WE, Yong T, Ma Z, Ramaseshan R (2006) Electrospun nanofibers: solving global issues. Materialstoday 9(3):40–50
Ramji K, Shah RN (2014) Electrospun soy protein nanofiber scaffolds for tissue regeneration. J Biomater Appl 29(3):411–422
Rather HA, Thakore R, Singh R, Jhala D, Singh S, Vasita R (2018) Antioxidative study of cerium oxide nanoparticle functionalised PCL-gelatin electrospun fibers for wound healing application. Bioact Mater 3(2):201–211
Ravichandran R, Seitz V, Venugopal JR et al (2013a) Mimicking native extracellular matrix with phytic acid-crosslinked protein nanofibers for cardiac tissue engineering. Macromol Biosci 13:366–375
Ravichandran R, Venugopal JR, Sundarrajan S, Mukherjee S, Sridhar R, Ramakrishna S (2013b) Expression of cardiac proteins in neonatal cardiomyocytes on PGS/fibrinogen core/shell substrate for cardiac tissue engineering. Int J Cardiol 167:1461–1468
Reiger KA, Birch NP, Schiffman JD (2013) Designing electrospun nanofiber mats to promote wound healing—a review. J Mater Chem B 1(36):4531–4541
Reneker DH, Chun I, Ertley D (2002) US Patent No. US6382526B1
Rho KS, Jeong L, Lee G et al (2006) Electrospinning of collagen nanofibers: effects on the behavior of normal human keratinocytes and early-stage wound healing. Biomaterials 27:1452–1461
Rnjak-Kovacina J, Wise SG, Li Z et al (2012) Electrospun synthetic human elastin: collagen composite scaffolds for dermal tissue engineering. Acta Biomater 8:3714–3722
Roberts TT, Rosenbaum AJ (2012) Bone grafts, bone substitutes and orthobiologics—the bridge between basic science and clinical advancements in fracture healing. Organogenesis 8(4):114–124
Roy T, Maity PP, Rameshbabu AP, Das B, John A, Dutta A, Ghorai SK, Chattopadhyay S, Dhara S (2018) Core-Shell Nanofibrous Scaffold Based on Polycaprolactone-Silk Fibroin Emulsion Electrospinning for Tissue Engineering Applications. Bioeng 5(3):68. https://doi.org/10.3390/bioengineering5030068
Sapru S, Das S, Mandal M, Ghosh AK, Kundu SC (2018) Prospects of nonmulberry silk protein sericin-based nanofibrous matrices for wound healing—in vitro and in vivo investigations. Acta Biomater 78:137–150
Sayed MM, Mousa HM, El-Aassar MR et al (2019) Enhancing mechanical and biodegradation properties of polyvinyl alcohol/silk fibroin nanofibers composite patches for cardiac tissue engineering. Mater Lett 255:1–4
Schmidt CE, Leach JE (2003) Neural tissue engineering**:** strategies for repair and regeneration. Annu Rev Biomed Eng 5(1):293–347
Schneider A, Wang XY, Kaplan DL, Garlick JA, Egles C (2009) Biofunctionalized electrospun silk mats as a topical bioactive dressing for accelerated wound healing. Acta Biomater 5:2570–2578
Sell SA, McClure MJ, Barnes CP et al (2006) Electrospun polydioxanone–elastin blends: potential for bioresorbable vascular grafts. Biomed Mater 1(2):72–80
Sell SA, Francis MP, Garg K, McClure MJ, Simpson DG, Bowlin GL (2008) Cross-linking methods of electrospun fibrinogen scaffolds for tissue engineering applications. Biomed Mater 3(4):1–12
Selvaraj S, Thangam R, Fathima NN (2018) Electrospinning of casein nanofibers with silver nanoparticles for potential biomedical applications. Int J Biol Macromol 120:1–25
Shanmugam K, Sundaramoorthy S (2015) Development and characterization of an electrospun mat from Eri silk fibroin and PLA blends for wound dressing application. RSC Adv 5(40):1–13
Shao W, He J, Sang F et al (2016) Coaxial electrospun aligned tussah silk fibroin nanostructured fiber scaffolds embedded with hydroxyapatite–tussah silk fibroin nanoparticles for bone tissue engineering. Mater Sci Eng C 58:342–351
Shaw TJ, Martin P (2009) Wound repair at a glance. J Cell Sci 122:3209–3213
Shi X, Zhou W, Ma D, Bridges D, Ma Y, Hu A (2015) Electrospinning of nanofibers and their applications for energy devices. J Nanomater 2015:1–20
Sisson K, Zhang C, Farach-Carson MC, Chase DB, Rabolt JF (2010) Fiber diameters control osteoblastic cell migration and differentiation in electrospun gelatin. J Biomed Mater Res Part A 94(4):1312–1320
Smith MJ, McClure MJ, Sell SA et al (2008) Suture-reinforced electrospun polydioxanone–elastin small-diameter tubes for use in vascular tissue engineering: a feasibility study. Acta Biomater 4:58–66
Songchotikunpan P, Tattiyakul J, Supaphol P (2008) Extraction and electrospinning of gelatin from fish skin. Int J Biol Macromol 42:247–255
Srivastava CM, Purwar R (2016) Fabrication of robust *Antheraea assama* fibroin nanofibrous mat using ionic liquid for skin tissue engineering. Mater Sci Eng, C 68:276–290
Srivastava CM, Purwar R (2018) Fabrication of 3D self-assembled nonmulberry *Antheraea mylitta* (tasar) fibroin nonwoven mats for wound dressing applications. Macromol Res 26:872–881
Srivastava CM, Purwar R, Gupta AP (2019) Enhanced potential of biomimetic, silver nanoparticles functionalized *Antheraea mylitta* (tasar) silk fibroin nanofibrous mats for skin tissue engineering. Int J Biol Macromol 130:437–453
Stack M, Parikh D, Wang H et al (2019) Electrospun nanofibers for drug delivery. In: Ding B, Wang X, Yu J (eds) Electrospinning: nanofabrication and application, pp 735–764
Stephansen K, Chronakis IS, Jessen F (2014) Bioactive electrospun fish sarcoplasmic proteins as a drug delivery system. Colloids Surf B Biointerfaces 122:158–165
Stitzel J, Liu J, Lee SJ et al (2006) Controlled fabrication of a biological vascular substitute. Biomaterials 27:1088–1094
Subbiah T, Bhat GS, Tock RW, Parameswaran S, Ramkumar SS (2005) Electrospinning of nanofibers. J Appl Polym Sci 96(2):557–569
Sun B, Long YZ, Zhang HD et al (2014) Advances in three-dimensional nanofibrous macrostructures via electrospinning. Prog Polym Sci 39(5):862–890

Suzuki A, Tanizawa K (2009) Poly(ethylene terephthalate) nanofibers prepared by CO_2 laser supersonic drawing. Polymer 50(3):913–921
Tabe S (2014) Electrospun nanofiber membranes and their applications in water and wastewater treatment. In: Hu A, Apblett A (ed) Nanotechnology for water treatment and purification, pp 111–43
Tao SL, Desai TA (2007) Aligned arrays of biodegradable poly(ε-caprolactone) nanowires and nanofibers by template synthesis. Nano Lett 7(6):1463–1468
Taylor GI (1969) Electrically driven jets. The Royal Society 313(1515):453–475
Taylor CA, Braza D, Rice JB, Dillingham T (2008) The incidence of peripheral nerve injury in extremity trauma. Am J Phys Med Rehabil 87:381–385
Thirugnanaselvam M, Gobi N, Karthick SA (2013) SPI/PEO blended electrospun martrix for wound healing. Fibers Polym 14(6):965–969
Thompson ZS, Rijal NP, Jarvis D, Edwards A, Bhattarai N (2016) Synthesis of keratin-based nanofiber for biomedical engineering. J vis Exp 108:1–8
Tonsomboon K, Oyen ML (2013) Composite electrospun gelatin fiber-alginate gel scaffolds for mechanically robust tissue engineered cornea. J Mech Behav Biomed Mater 21:185–194
Tucker N, Stanger JJ, Staiger MP, Razzaq H, Hofman K (2012) The history of the science and technology of electrospinning from 1600 to 1995. J Eng Fibers Fabr 7(2):63–73
Unnithan AR, Gnanasekaran G, Sathishkumar Y, Lee YS, Kim CS (2013) Electrospun antibacterial polyurethane–cellulose acetate–zeincomposite mats for wound dressing. Carbohyd Polym 102:884–892
Uttayarat P, Jetawattana S, Suwanmala P, Eamsiri J, Tangthong T, Pongpat S (2012) Antimicrobial electrospun silk fibroin mats with silver nanoparticles for wound dressing application. Fibers Polym 13(8):999–1006
Vega-Lugo AC, Lim LT (2009) Controlled release of allyl isothiocyanate using soy protein and poly(lactic acid) electrospun fibers. Food Res Int 42:933–940
Venugopal J, Ma LL, Yong T, Ramakrishna S (2005) In vitro study of smooth muscle cells on polycaprolactone and collagen nanofibrous matrices. Cell Biol Int 29:861–867
Venugopal J, Low S, Choon AW, Sampath-Kumar TS, Ramakrishna S (2008) Mineralization of osteoblasts with electrospun collagen/hydroxyapatite nanofibers. J Mater Sci Mater Med 19:2039–2046
Vig K, Chaudhari A, Tripathi S et al (2017) Advances in skin regeneration using tissue engineering. Int J Mol Sci 18(4):1–19
Vogt L, Liverani L, Roether JA, Boccacini AR (2018) Electrospun zein fibers incorporating poly(glycerol sebacate) for soft tissue engineering. Nanomaterials 8(3):150
Wang H, Zhang Y, Shao H, Hu X (2005) Electrospun ultra-fine silk fibroin fibers from aqueous solutions. J Mater Sci 40:5359–5363
Wang G, Hu X, Lin W, Dong C, Wu H (2011) Electrospun PLGA–silk fibroin–collagen nanofibrousscaffolds for nerve tissue engineering. In Vitro Cell Dev Biol Anim 47:234–240
Wang Y, Li P, Xiang P, Lu J, Yuan J, Shen J (2015) Electrospun polyurethane/keratin/AgNP biocomposite mats for biocompatible and antibacterial wound dressings. J Mater Chem B 4(4):635–648
Wei K, Li Y, Lei X et al (2011) Emulsion electrospinning of a collagen-like protein/PLGA fibrous scaffold: empirical modeling and preliminary release assessment of encapsulated protein. Macromol Biosci 11(11):1526–1536
Wen P, Zong MH, Linhardt RJ, Feng K, Wu H (2017) Electrospinning: a novel nano-encapsulation approach for bioactive compounds. Trends Food Sci Technol 70:56–68
Wnek GE, Carr ME, Simpson DG, Bowlin GL (2003) Electrospinning of nanofiber fibrinogen structures. Nano Lett 3(2):213–216
Wray LS, Orwin EJ (2009) Recreating the microenvironment of the native corneafor tissue engineering applications. Tissue Eng Part A15(7):1463–1472
Xie J, Hsieh YL (2003) Ultra-high surface fibrous membranes from electrospinning of natural proteins: casein and lipase enzyme. J Mater Sci 38:2125–2133
Xie J, MacEwan MR, Schwartz AG, Xia Y (2009) Electrospun nanofibers for neural tissue engineering. Nanoscale 2(1):35–44
Xu X, Yang L, Xu X et al (2005) Ultrafine medicated fibers electrospun from W/O emulsions. J Controll Release 108(1):33–42
Xu H, Cai S, Sellers A, Yang Y (2014) Intrinsically water-stable electrospun three-dimensional ultrafine fibrous soy protein scaffolds for soft tissue engineering using adipose derived mesenchymal stem cells. RSC Adv 4(30):15451–15457
Xu L, Wang S, Sui X et al (2017) Mesenchymal stem cells seeded regenerated silk fibroin complex matrices for liver regeneration in an animal model of acute liver failure. Appl Mater Interfaces 9(17):14716–14723
Yan J, Qiang L, Gao Y et al (2011) Effect of fiber alignment in electrospun scaffolds on keratocytes and corneal epithelial cells behavior. J Biomed Mater Res, Part A 100(2):527–535
Yang Q, Li Z, Hong Y, Zhao Y, Qiu S, Wang C, Wei Y (2004) Influence of solvents on the formation of ultrathin uniform poly(vinyl pyrrolidone) nanofibers with electrospinning. J Polym Sci, Part b: Polym Phys 42(20):3721–3726
Yang J, Yamato M, Kohno C et al (2005) Cell sheet engineering: recreating tissues without biodegradable scaffolds. Biomaterials 26:6415–6422
Yang D, Li Y, Nie J (2007a) Preparation of gelatin/PVA nanofibers and their potential application in controlled release of drugs. Carbohydr Polym 69:538–543
Yang D, Lu B, Zhao Y, Jiang X (2007b) Fabrication of aligned fibrous arrays by magnetic electrospinning. Adv Mater 19(21):3702–3706
Yang DJ, Kamienchick I, Youn DY, Rothschild A, Kim ID (2010) Ultrasensitive and highly selective gas sensors based on electrospun SnO2 nanofibers modified by Pd loading. Adv Func Mater 20:4258–4264
Yao C, Li X, Song T (2007) Electrospinning and crosslinking of zein nanofiber mats. J Appl Polym Sci 103(1):380–385. https://doi.org/10.1002/app.24619
Yao CH, Lee CY, Huang CH, Chen YS, Chen KY (2017) Novel bilayer wound dressing based on electrospun gelatin/keratin nanofibrous mats for skin wound repair. Mater Sci Eng, C 79:533–540
Yao Q, Hu Y, Yu F, Zhang W, Fu Y (2018) A novel application of electrospun silk fibroin/poly(L-lactic acid-*co*-ε-caprolactone) scaffolds for conjunctiva reconstruction. RSC Adv 8(33):18372–18380
Yen KC, Chen CY, Huang JY, Kuo WT, Lin FH (2016) Fabrication of keratin/fibroin membranes by electrospinning for vascular tissue engineering. J Mater Chem B 4(2):237–244
Yoon K, Kim K, Wang X, Fang D, Hsiao BS, Chu B (2006) High flux ultrafiltration membranes based on electrospun nanofibrous PAN scaffolds and chitosan coating. Polymer 47:2434–2441
Yuan X, Zhang Y, Dong C, Sheng J (2004) Morphology of ultrafine polysulfone fibers prepared by electrospinning. Polym Int 53(11):1704–1710
Yuan J, Shen J, Kang IK (2008) Fabrication of protein-doped PLA composite nanofibrous scaffolds for tissue engineering. Polym Int 57:1188–1193
Yuan B, Jin Y, Sun Y et al (2012a) A strategy for depositing different types of cells in three dimensions to mimic tubular structures in tissues. Adv Mater 24(7):890–906
Yuan J, Geng J, Xing Z et al (2012b) Novel wound dressing based on nanofibrous PHBV–keratin mats. J Tissue Eng Regen Med 9(9):1027–1035

Yördem OS, Papila M, Menceloğlu YZ (2008) Effects of electrospinning parameters on polyacrylonitrile nanofiber diameter: an investigation by response surface methodology. Mater Des 29(1):34–44

Zafar M, Najeeb S, Khurshid Z et al (2016) Potential of electrospun nanofibers for biomedical and dental applications. Materials (basel) 9(2):1–7

Zhang H, Liu J (2013) Electrospun poly(lactic-co-glycolic acid)/wool keratin fibrous composite scaffolds potential for bone tissue engineering applications. J Bioact Compat Polym 28(2):141–153

Zhang X, Lu Y (2014) Centrifugal spinning: an alternative approach to fabricate nanofibers at high speed and low cost. Polym Rev 54(4):677–701

Zhang H, Qian XM (2011) The applications of electrospun nanofibers in the medical materials. Adv Mater Res 148–149:1138–1143

Zhang K, Wang H, Huang C, Su Y, Mo X, Ikada Y (2009) Fabrication of silk fibroin blended P(LLA-CL) nanofibrous scaffolds for tissue engineering. J Biomed Mater Res, Part A 93(3):984–993

Zhang F, Zuo B, Fan Z et al (2012) Mechanisms and control of silk-based electrospinning. Biomacromol 13(3):798–804

Zhang C, Wen J, Yan J et al (2015) In situ growth induction of the corneal stroma cells using uniaxially aligned composite fibrous scaffolds. RSC Adv 5:12123–12130

Zhang W, Ronca S, Mele E (2017a) Electrospun nanofibres containing antimicrobial plant extracts. Nanomaterials 7(2):1–17

Zhang Z, Tu W, Peijs T, Bastiaancen CWM (2017b) Fabrication and properties of poly(tetrafluoroethylene) nanofibres via sea-island spinning. Polymer 109:321–331

Zhong J, Zhang H, Yan J, Gong X (2015) Effect of nanofiber orientation of electrospun nanofibrous scaffolds on cell growth and elastin expression of muscle cells. Colloids Surf, B 136:772–778

Zhou J, Cao C, Ma X, Lin J (2010) Electrospinning of silk fibroin and collagen for vascular tissue engineering. Int J Biol Macromol 47:514–519

Zhou W, Feng Y, Yang J, Fan J (2015) Electrospun scaffolds of silk fibroin and poly(lactide-co-glycolide) for endothelial cell growth. J Mater Sci Mater Med 26(1):1–14

Zong X, Kim K, Fang D, Ran S, Hsiao BS, Chu B (2002) Structure and process relationship of electrospun bioabsorbable nanofiber membranes. Polymer 43(16):4403–4412

Zuo W, Zhu M, Yang W, Yu H, Chen Y, Zhang Y (2005) Experimental study on relationship between jet instability and formation of beaded fibers during electrospinning. Polym Eng Sci 45(5):704–709

Zuo F, Tan DH, Wang Z, Jeung S, Macosko CW, Bates FS (2013) Nanofibers from melt blown fiber-in-fiber polymer blends. ACS Macro Lett 2(4):301–305

COMADEM
International Journal of Condition Monitoring and Diagnostic Engineering Management

Numerical modelling of Flow characteristics over sharp crested triangular and W-shaped weirs for establishing downstream scour behavior.

Stuti Mishra[a] *, Dr. Munendra Kumar[b], and Vijay Kaushik[c]

[a]*M.Tech. Student, Hydraulics and water resources Department , Delhi Technological University, New Delhi 110042, India*
[b]*Professor, Hydraulics and water resources Department, Delhi Technological University, New Delhi 110042, India*
[c]*Ph.D. student, Hydraulics and water resources Department, Delhi Technological University, New Delhi 110042, India*
∗ *Corresponding author. Tel*. +91 8085969253; *email*: *stuti.mishra44@gmail.com*

ABSTRACT

Weirs are flow control devices that are often used in waterways. As the weirs by nature change the flow patterns in the water ways, significant changes in the flow structure in the vicinity of the weir structure can be easily seen. These significant changes in the flow patterns create significant changes in the surrounding flow geometries. Scour in the downstream of the weir is fairly well-known side effect in the channels. The objective of this study is to evaluate the flow characteristics of a triangular sharp crested weir and its comparison with W-shaped sharp crested weir using CFD (Computational Fluid Dynamics) with Ansys Fluent as the main tool. Important flow characteristics, before and after two weir shapes have been evaluated and their potential to cause scour in the downstream evaluated with a view establish relationship between weir shape and the likely scour area has been investigated. It has been found that weir shape has significant influence on scour behaviour and this study may be used a reference for future evaluation of scour potential of weirs of different shapes.

Keywords: *Weir; CFD;Discharge coefficient.*

.

1. Introduction

Weirs are one of the most commonly used hydraulic structures in the waterways. These are used to serve a variety of purposes, including control and quantification of the flowrates. The weirs are designed to provide obstruction in flowing water to enable raised stream level upstream of the weir. The presence of obstruction in the flow results in complex flow patterns to develop in the vicinity of the weir. In particular, in the downstream section of the weir considerable scouring can occur because of high flow velocity jet impacting on the soft surface. The foundation stability of hydraulic structures can be seriously affected by scour hence it is very essential to establish scour behavior of weirs before these are actually built.

Sharp crested weirs are quite possibly the most well-known hydraulic structures used in open channels. In these structures it is fairly straight forward to link geometric variables with the flow variables from the first principle and thus estimate the global flow features. Sharp peaked weirs have regular geometry and the flow through these weirs has distinct flow patterns associated with it. Types of weirs dependent on the crest form are: Sharp- weir crested, broad crested weir, narrow crested weir and weir which are Ogee-shaped.

For the analysis we have chosen sharp crested weirs since the sharp crested weir are the most common weirs used for estimating the release of the water from the little streams and waterways. Once the relationship between the flow characteristics and scour behavior is established it will be attempted to establish same behavior for complex weir shapes.

From simple analysis it can be shown that, for the triangular weir:

$$Q = \frac{8}{15} Cd \times tan\frac{\theta}{2} \times H^{\frac{5}{2}} \times \sqrt{2g} \qquad (1)$$

Q: flow rate;

H: height above the weir base;

Θ: angle of triangular weir

Cd: discharge coefficient

Ideally Θ ranges from 45° to 90° for this type of weir to get the optimum discharge quantity.

For the W shaped weir, separate calculations are needed for both the triangles considering it as a double V notch.

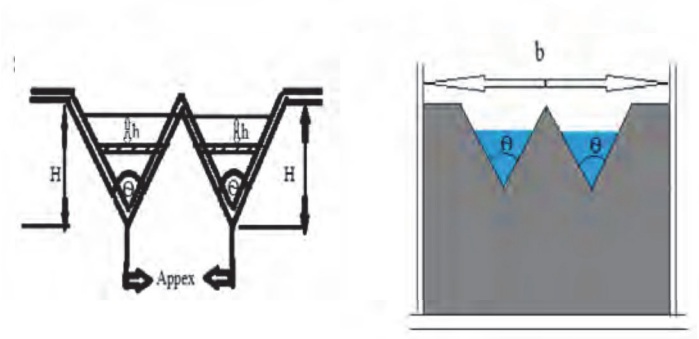


Figure 1. Triangular weir cross-section

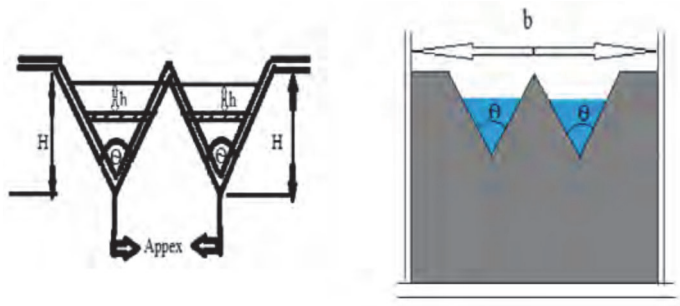


Figure 2. W weir cross-section

Figure 1 represents the cross sectional view of triangular weir placed in a rectangular channel and figure 2 represents the cross sectional view of W weir placed in a rectangular channel.

2. Literature Review

There are several references where the performances of sharp crested weirs of different shapes have been evaluated both for normal and oblique orientations [Borghei et al. (2003), Prakash and Shivapur (2004)]. The discharge equations for different weir shapes for the normal orientation have been modified to take into account the oblique angle.

Tuyen (2006), carried out detailed experiments on rectangular weir and provided details of the flow structures as well as established the effect of oblique angle on the discharge coefficient. Typically the weir shape affects the flow pattern near the weir considerably. This change in flow pattern may be responsible for considerable changes in the flow geometries. Scour process is caused by interaction of water flow with the boundary. Typically scour holes are created because of imbalance in incoming sediment load and outgoing sediment load. In the downstream of weir the flow velocities are typically higher and are associated with vortices and eddies. It is therefore important to understand the interrelation between the flow behavior and the resulting interaction with the downstream boundary surfaces.

Ashour et al. (2008), analyzed scour behavior between the straight and curved weirs. They found that a certain geometric arrangement may result in minimum scour profile. Hamed et al. (2009) carried out extensive investigation on oblique and normal V notch scour profile, however the interrelation between the flow profile and scour profile was not attempted.

Issam A. (2013), conducted various laboratory experiments to investigate the coefficients of momentum and kinetic energy change in asymmetric, rectangular and compound cross-section networks. The Reynolds number was varied from 14,348.9 to 54,868.2 as the relative depth changed from 0.155 to 0.825. Kinetic energy and momentum correction coefficients, α and β, were determined for 9 separate cross sections. The values of α and β for 9 different cross-sections was found to be 1.1525 and 1.1261 respectively

Bautista-Capetillo et al. (2014), found that the three-sided weirs could be effectively used for controlling the flow and offer an effective option for controlling the dispersion of water. In this investigation, the stream profiles were observed using a low-speed photographic strategy. In the experiments various parameters were varied over a wide range for example, weir angles were varied from 30° to 90°), peak heights were varied from 8 to10 cm, and the flow rates were varied from 0.01 l/s to 7.82 l/s. The equations developed in this study would be useful in calculating discharge..

Detailed experimental investigations are limited by the range of parameters and the geometric configuration and hence getting universal correlations are very difficult from these. To overcome this problem recently a large number of investigations have been carried out using computational fluid dynamics.

Khassaf et al. (2016), carried out detailed numerical and experimental investigations on the two compound side weirs (Rectangular and Semi-Circle). For numerical work a Computational Fluid Dynamics (CFD) software named Flow 3D was used to depict the flow structure in subcritical flow conditions. Although the outcomes were of great importance limited information was available on linking geometric variables with flow variables.

Ibrahim (2018), carried out numerical and experimental investigation on two different weir shapes (rectangular and triangular) using multiple nonlinear regression (MNLR) and computational fluid dynamics (CFD) approaches. A good matching was obtained between the experimental and computational results, however the results were not explored in greater detail. Especially the flow structures could have been analyzed to link downstream scour behavior with the flow field.

Ansari and Patil (2020), carried out an interesting study on labyrinth side weirs that are used for diversion and head regulation. Using CFD they carried out parametric investigation for developing empirical equation for discharge coefficient. The computational results were validated against experimental results and were found to be reasonable agreement. It however did not present information on local flow characteristics resulting in limited information about interrelation between local and global parameters.

Rezazadeh et al. (2020), carried out detailed CFD analysis of the flow using Open foam, over sharp crested weirs and compared the discharged coefficient obtained numerically with experimental data. They obtained a very good matching between the two. They also analyzed the velocity and pressure fields in the vicinity of the weir, however this analysis is not in depth. Furthermore, no attempt was made to interrelate the flow field downstream of the weir with the likely scour conditions in the downstream.

The above review of literature has clearly indicated that although considerable amount of work has been done on understanding the relation between the head and flow conditions there is limited work on the effect of weir shape on the flow characteristics. In particular, development of secondary flow before and after the weir has not been investigated. The downstream secondary flow play an important role in scouring phenomena which needs to be investigated. The present work shows the detailed flow features before and after the weir with two distinct shapes and correlates shapes with the flow features both upstream and downstream of the weir. It also interrelates weir shapes and resulting flow features with likely scour profile in the downstream. This work will enable designer to develop mitigating strategies during the design stage for effective operation of weirs.

3. Methodology

To achieve the objectives outlined in the previous section, a numerical scheme has been developed. For this purpose, Ansys Fluent software has been used. Fluent software is a state of the art software that enables a number of functionalities to enable modeling of real life systems reasonably well. This CFD code works based on finite volume formulation. Fluent has the capacity to take care of 2D and 3D modelling issues of open channel stream. In a typical CFD-based numerical modelling there are three stages, namely the pre-processing, solver setup and post-processing stages. The first stage (pre-processing stage) involves geometry

creation, creating mesh for the flow domain, providing parameters for solver setup including materials, boundary conditions, turbulence parameters, residuals, convergence criteria and initially selected solution. All these parameters require careful selection for a physically realistic flow to be modelled reasonably well. This is especially true for multiphase like flow as simulated in volume of fluid method. Once the solver has been used to solve the equations, the post-processing stage is used to analyze the results obtained in order to display flow features within the flow domain. Furthermore, appropriate two-dimensional planes can be created at different sections to quantify variations in the flow variables. In modelling flows using the Reynolds-Averaged Navier-Stokes (RANS) approach, the use of an appropriate turbulence model for simulating complex-geometry flows requires special consideration. This is more so in the case of multiphase flows as in the open channel flow simulations, volume of fluid method is used that tracks the interface between the water and air.

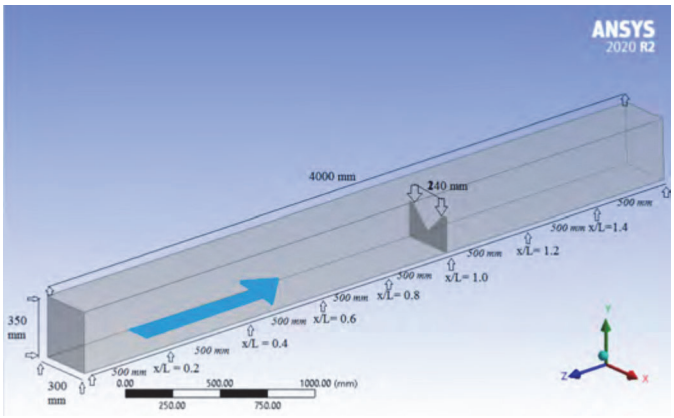


Figure 3: Geometric details of a triangular weir

Simulations were performed over a 4 m long and 30 cm wide tilting flume, using a triangular weir pattern (figure 3) and then for a W-shaped weir (figure 4). For triangular weir simulation Θ was taken as 90° i.e. right angled triangular weir. For the W-shaped weir the crest height was same as that of triangular weir, but Θ for each triangle is 53.12°. The weirs were placed at 2.5 m distance from the upstream side. The upstream length is enough to capture effect of weir on upstream flow and the down-stream length is enough for capturing scour related flow structures. The thickness of weirs was so that it fulfils the criteria of being a sharp crested type. The thickness of both the triangular as well as W-shaped weir was 2 mm that satisfies the sharp crest requirements. For this research work, Volume of Fluid multiphase model was used. It has been well established that volume of fluid method can be used for simulating open channel.

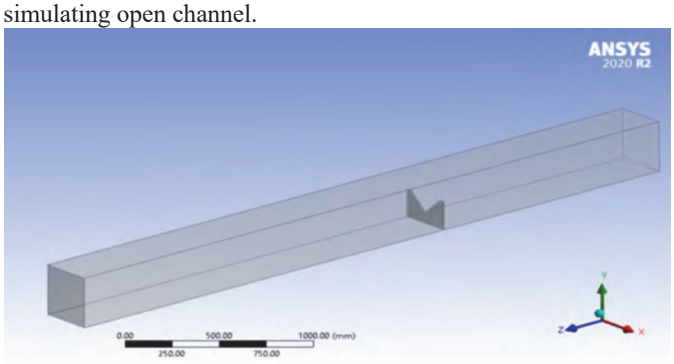


Figure 4: Geometric details of a W weir

flow structures reasonably well. In the volume of fluid method water has been used as the primary phase and air as the secondary phase.

The volume of fluid involves solving mass conservation equation along with momentum equations. Additional equation to take care of multiphase effects are also included in form of volume fraction transfer equation. The volume fraction equation enables tracking of liquid-air interface with volume fraction ranging from 0 to 1. If it is 1 then it indicates 100% water and at zero it is completely air.

The equations used in a typical volume of fluid method area as shown below.

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_i}(\rho u_i) = 0 \tag{2}$$

$$\frac{\partial}{\partial t}(\rho u_j) + \frac{\partial}{\partial x_i}(\rho u_i u_j) = -\frac{1}{\rho}\frac{\partial P}{\partial x_j} + \frac{\partial}{\partial x_i}(\mu + \mu_t)\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)_x + \rho g_i \tag{3}$$

$$\rho = \alpha_a \rho_a + \alpha_w \rho_w \tag{4}$$

Where ui are flow velocity components, αa and αw are volume fractions corresponding to air and water, ρ, ρa and ρw are densities corresponding to mixture, air and water, P is pressure and g is acceleration dur to gravity.

$$\mu = \alpha_a \mu_a + \alpha_w \mu_w \tag{5}$$

In the above equation μ is viscosity of mixture and subscript a, w and t represent air, water and turbulence respectively. The volume of fluid method uses a function for free surface identification. This is obtained from the equation 6, where X is the free surface function.

$$\frac{\partial X}{\partial t} + u_i \frac{\partial X}{\partial x_i} = 0 \tag{6}$$

In this investigation the k epsilon turbulence model has been used as it has been shown to be adequate for this type of investigation.

Meshing is an important part of the simulation process. With the element order (triangular) size 1.45e-0.002 m, and skewness being at 0.90 numbers of elements came out to be 496031 with a good inflation across the boundaries. For the W-shaped weir, element order size was kept at 1.45e-0.002 m only with number of elements at 493682. The mesh were tested for convergence and it was found that this mesh size gives repeatable converged results. Figures 5, 6 and 7 show meshed weir pictures. It can be seen from figure 7 that near the boundaries the mesh density is higher enabling it to capture near boundary phenomena reasonably well

.

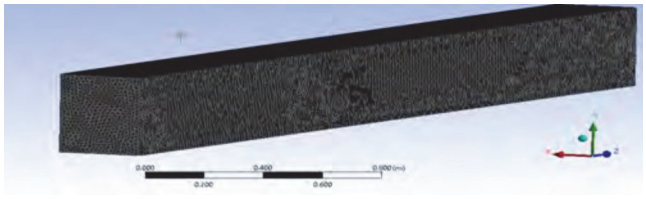


Figure 5: Mesh details for Triangular weir

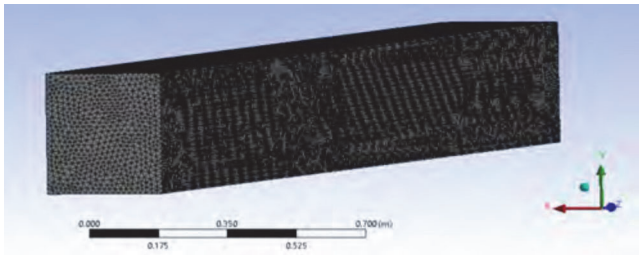


Figure 6: Mesh details for W weir.

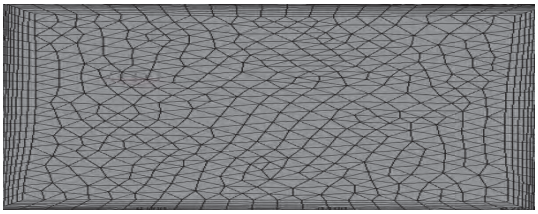


Figure 7: Mesh details for triangular weir at the inlet

The boundary conditions used in the present work are taken from a parallel experimental work conducted in the laboratory. The following table shows various boundary conditions used in the analysis.

Table 1: Boundary conditions used in analysis

Boundary zone	Type	Velocity (m/s)	Free surface level (m)
Inlet	Velocity inlet	0.478	0.155
Ambient	Pressure outlet	--	--
outlet	Pressure outlet	--	0.07

The geometric setup shown in figure 1 was already used in the experimental work for triangular weir and it gave us an idea to place the weir at 2.5 m distance from the inlet to create the replica of the experimental model in CFD. Furthermore, we adopted the same weir location for W weir as well. Velocity of 0.478 m/s was obtained using Acoustic Doppler Velocimeter and head was also calculated in the experimental work. We used the given results as our initial conditions to simulate the flow and Coefficient of discharge was calculated, hence validating the experimental results (Reference 12). Details about the numerical setup is provided in table 2.

Table 2 Details about the numerical setup

Turbulence model	K- epsilon
Time step chosen	Steady state
Number of time step	Steady state
Convergence	Triangular weir- 9000+ iterations W weir- 8500+ iterations

4. Results and discussion

The results obtained from numerical simulations have been evaluated to obtain flow characteristics near the weir for both the weir shapes. The figure 8 shows the pictorial representation of water contour before and after the weir. It can be seen that the air-water interface is correctly modelled, and the wave form is similar to the published literature. Similar observations can be made about W weir (Figure 9). To investigate the flow patterns first it was decided to track the interface along the length of the open channel. The figure 8 shows the water profile before and after the weir, for a triangular weir. It can be seen that the water height at the inlet is 0.155 m and it reaches to 0.160m at the weir after which it changes to 0.09m with its value at exit is 0.07 m. The figure 9 shows the similar characteristics for the W weir. The water height at the inlet is also 0.155 m and it reaches to 0.160 m at the weir after which it changes to 0.09 m with its value at exit is 0.07 m

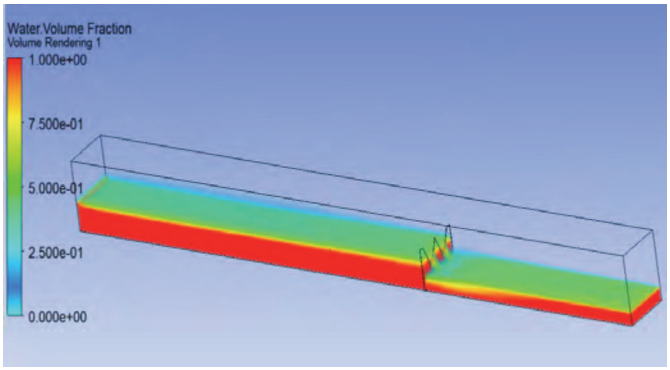


Figure 8: Water profile in the channel for triangular weir

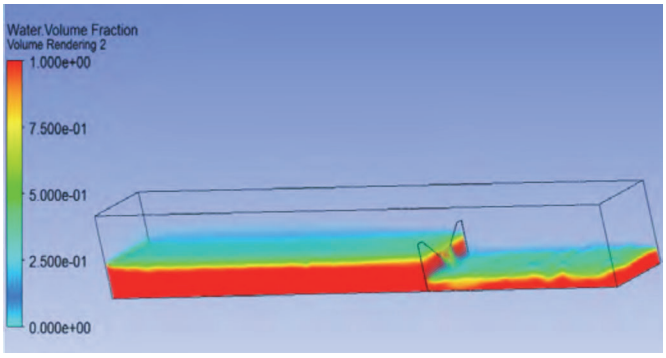


Figure 9: Water profile in the channel for triangular weir

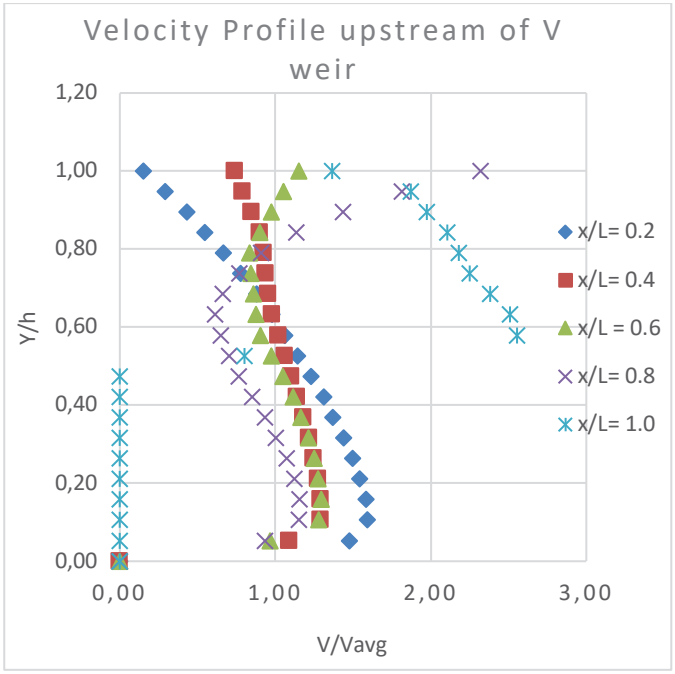


Figure 10: Velocity profile in the channel for triangular weir from inlet to the weir location

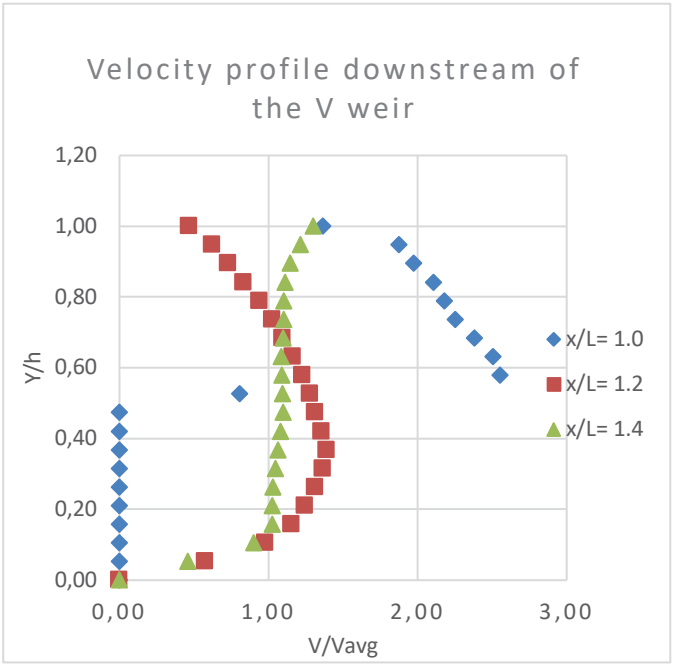


Figure 11: Velocity profile in the channel for triangular weir from the weir location to the outlet

To understand flow characteristics in the upstream region for the triangular weir non dimensional velocity (V/Vav) has been shown against non-dimensional distance at five stations from near the inlet (x/L=0.2) up to the weir location (x/L=1.0) in figure 10. It can be seen that as the flow approaches the weir considerable change in the velocity profile can be seen. At x/L=0.2, it can be seen that at the bed velocity is zero after which it reaches a maximum value of 1.75 times the average velocity at this section at y/H value of 0.11. From this value the non-dimensional velocity decreases continuously to the top of the water surface and attains a value of 16.8% of the average velocity. As the flow moves forward at x/L=0.4, it can be seen that presence of weir is starting to show its effect. Although the trend is similar, more uniformity in the velocity values can be seen. At y/H=0.11, the velocity values is now 1.37 times the average velocity and at the surface the velocity values is 0.78 times the average velocity. Similar trend is seen at x/L=0.6, with velocities at the bottom of channel decreasing and at the top increasing showing the increased effect of presence of weir. There is a clear point of inflexion present at which velocity is least (except that at the bed). The same effect can be seen at x/L=0.8, however these effects are more pronounced with very high surface velocity (2.32 times the average velocity). At the weir the velocities at the nearer the bed have become zero whereas the velocities at the notch section is higher. However, the maximum velocity point is not at the surface but some distance below the surface. The above discussion has clearly indicated evolution of velocity field along the central plane.

To understand flow characteristics downstream of the wear, the non-dimensional velocity (V/Vav) has been plotted against the non- dimensional height from the bottom of the channel. This information will enable us to identify velocity induced scour behavior in the downstream of the channel.

Figure 11 shows the velocity magnitudes from the weir to the outlet locations. It has been extensively reported that downstream of the weir locations are very susceptible for scour and hence it is important to investigate velocity field in these locations. As the weir cause obstruction to the flow, the secondary flows start to become important nearer the weir. These flow features are then transported in the downstream of the weir that may contribute to scour phenomena seen. It can be seen that at x/L =1, velocity near the channel bottom is zero, however at x/L=1.2, the bottom velocities start to rise significantly. The local velocities at x/L=1.2 are significantly higher at the bottom of the channel than at the top of the channel. At x/L=1.4, this velocity has reduced at the channel bottom and increases near the surface. From the above discussion it can clearly be highlighted that the scour potential in the downstream side from x/L=1.0 to x/L=1.2 is very high.

Replacing a V notch with a W notch may be an option especially if the channel is fairly wide and also it is also expected to reduce non-uniformity in the flow profiles. This will also reduce the changes in the flow structure at the downstream sections.

In the W weir (figure 12), at x/L=0.2, it can be seen that at the bed velocity is zero as seen for V weir, after which it reaches a maximum value of 1.62 times the average velocity at y/H value of 0.11. From this value the non-dimensional velocity decreases up to the surface to a V/Vav value of 0.65. We can thus see that although the trend in both the V weir and W weir are similar the actual velocity values are different. As the flow moves forward to a location at x/L=0.4, the point of inflexion can be seen that was seen in V weir at only x/L=0.6. This effect of weir is felt far upstream in W weir as compared to a V weir. The main reason behind this is the presence of two V notches in a single W notch. Similar trend is seen at x/L=0.6, with velocities almost similar that seen at x/L=0.4. At x/L=0.8, however, there are considerable changes in the velocity profile. The velocity from the bottom to the top is continuously increasing, which is quite different to the one seen for V notch. At x/L=1.0, the velocity profile is similar to the one seen for V notch. The above discussion has clearly indicated evolution of velocity field for W notch is quite different as compared to a V notch.

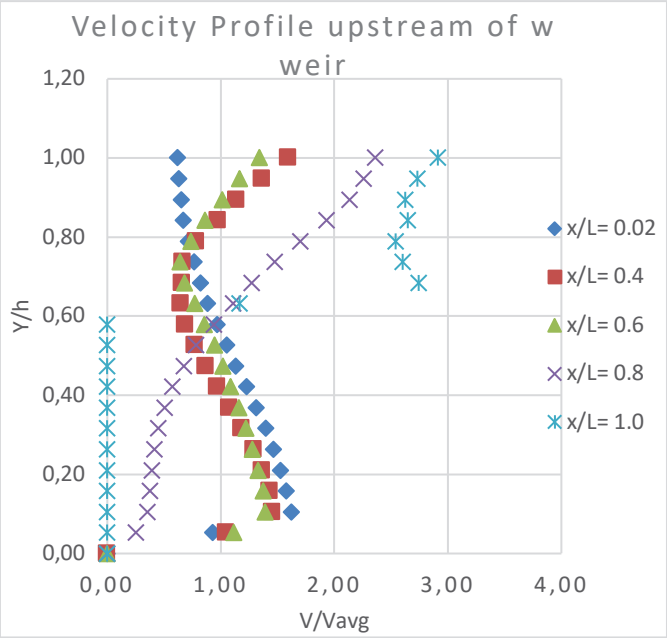


Figure 12: Velocity profile in the channel for W weir from the inlet to the weir location

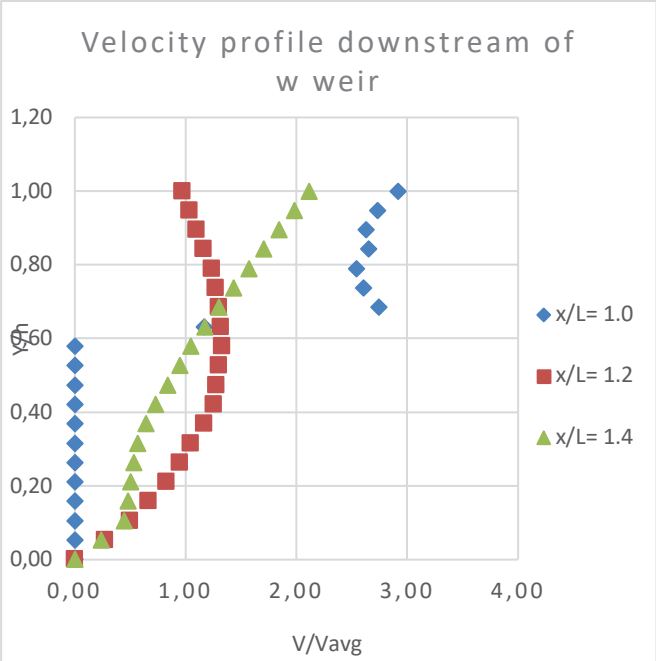


Figure 13: Velocity profile in the channel for W weir from the weir location to the outlet

Figure 13 shows the velocity magnitudes for the W- weir from the weir to the outlet locations. It can be seen that at x/L=1, velocity near the channel bottom is zero, however at x/L=1.2, the bottom velocities start to rise significantly as seen for the V weir. However, the point of inflexion has shifted upwards for W weir, indicating slightly smaller velocities at the channel as compared to the V weir. At x/L=1.4, this velocity has reduced at the channel bottom and increases near the surface, which is similar to the V weir. From the above discussion it can clearly be highlighted that the scour potential in the downstream side from x/L=1.0 to x/L=1.2 is very high for V weir and although it is smaller for W weir it can still be significant.

It is well established now that the flow pattern changes significantly nearer the weir which subsequently causes the scour in the downstream of the wear. To quantify this change a new parameter has been developed which can be expressed as velocity ratio= V_y/V_z. This parameter indicates the ratio of vertical velocity to the axial velocity. This velocity ration will be good indicator of scouring effect in the downstream of the weir. Figure 14 shows the variation of this parameter at different sections along the channel length.

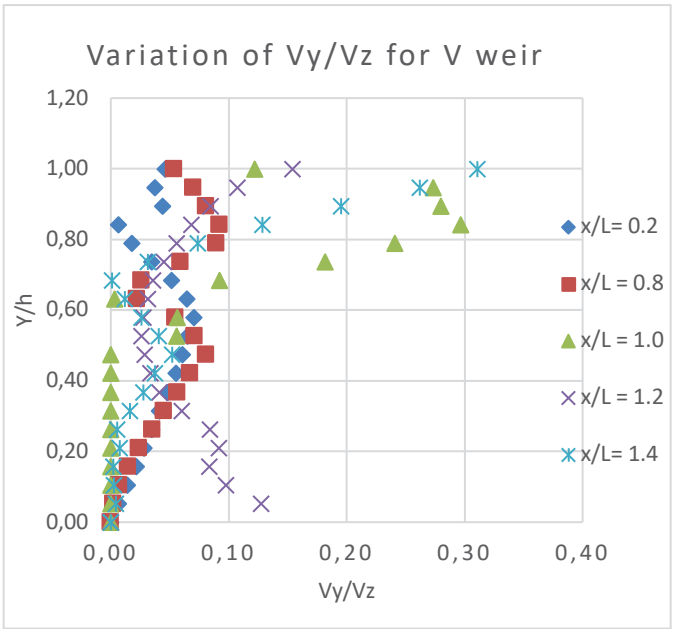


Figure 14: Variation of velocity ratio in the channel for V weir from the inlet to the outlet

It can be seen from the figure 14 that nearer the inlet most of the velocity is along the axial direction. The vertical velocity components is mostly less than 10% of the axial velocity along the vertical plane. The values nearer the bottom of the channel are higher than at the top. At very close to the weir (x/L=0.8), at the bottom of the channel no considerable change has been noticed, however at the top the vertical velocity components is significantly higher (approximately 4 times the value seen at x/L=0.2). At the weir, this value increases to about 30% of the axial velocity. In the downstream of the weir the maximum value of this parameters noted is about 15% and 30% at x/L of 1.2 and 1.4 respectively. From the earlier discussion it can be concluded that velocity ratio becomes quite higher in the downstream of the weir and this may be a good indicator of the scour potential of the weir.

Figure 15 shows the variation of velocity ratio with the non-dimensional distance from the bottom of the channel bed for various central sections along the length of the channel for W weir. The focus has been on the stations at the inlet and near the location of the weir. As scour is more pronounced in the near weir region. It can be seen that near the inlet the vertical velocity is less than 10% of the z-component velocity for 80% of the channel height. In the top 20% the vertical height has reached a value of about 30% of the z- component velocity. Thus near the inlet in W weir the y component velocities are higher as compared to V weir. This may be primarily because of presence of higher area variation along the width of the channel (two V notches in-place of one V notch).

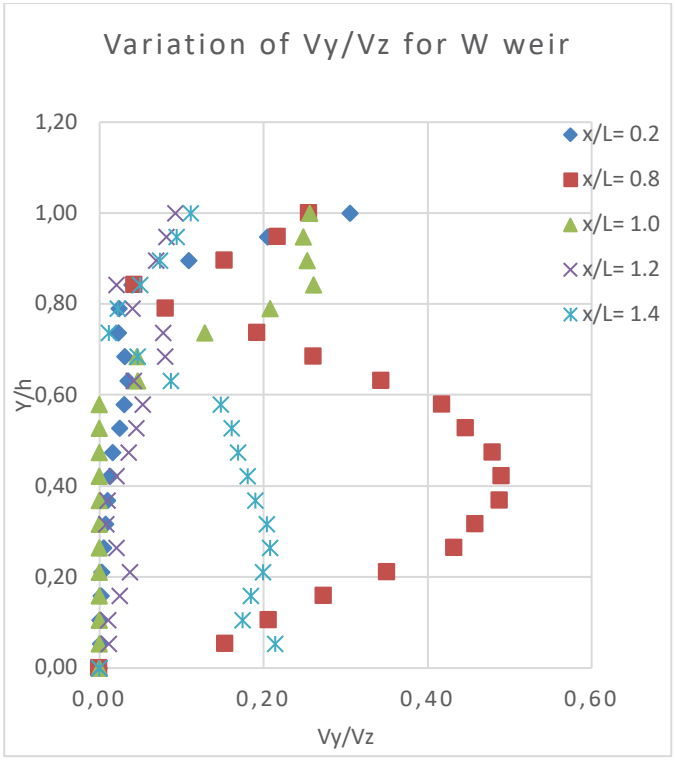


Figure 15: Variation of velocity ratio in the channel for W weir from the inlet to the outlet

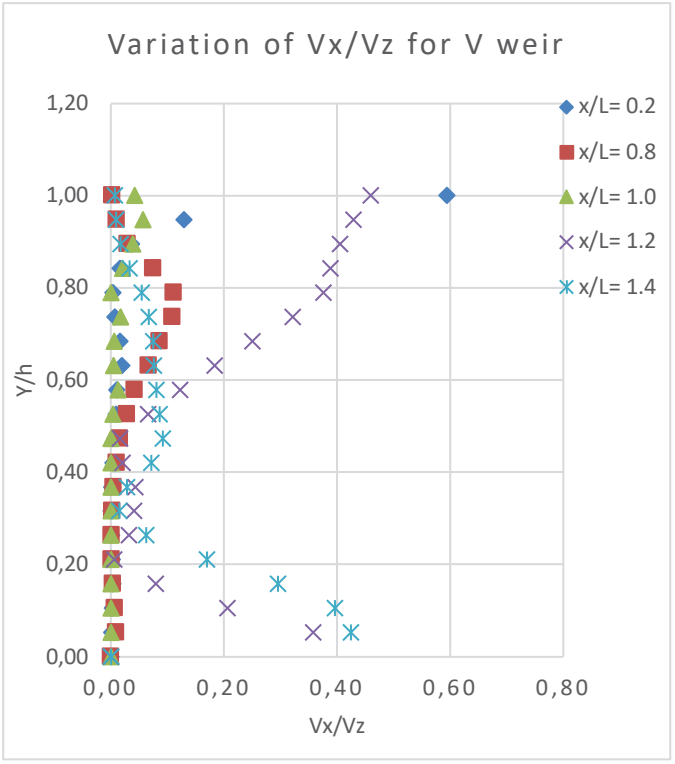


Figure 16: Variation of X velocity ratio in the channel for V weir from the inlet to the outlet

At x/L=0.8, the Y-component flow velocities are changing in the Y directions significantly. Near the mid-channel the vertical component velocity is reaching a value of about 50% of axial flow velocity indicating very strong secondary flow, unlike the V channel. At the weir location x/L=1.0, the vertical velocity components observed in W weir is almost similar to that seen in V weir. In the downstream of the weir however the vertical component velocities seen are smaller than that observed for V weir. It can thus be seen that although the secondary flows will be stronger in W channel in the upstream of the weir as compared to V weir, in the downstream W weir will have smaller secondary flows as compared to V weir. Thus, W weir will have smaller scour potential as compared to V weir in the downstream of the weir.

The velocity components in X direction also contributes in development of secondary flows in the horizontal plane. To analyze this non-dimensional X component velocity has been plotted against non-dimensional distance from the bed of the channel at various sections along the length of the channel.

Figure 16 shows variation of Vx /Vz for V weir at different sections along the length. Near the inlet, x/L=0.2, it can be seen that the X components velocity is almost negligible indicating that most of the flow is directed along z axis, except at the top. At x/L =0.8, i.e. near the V weir, it can be seen that only from y/H=0.8 to Y/H=1.0, significant motion along X direction is seen. At the weir X components velocity is almost zero as expected because the two dominating velocity components at the weir will be y and z components.

At x/L=1.2, we can see that the X components velocity is about 40% of the Z components velocity. This indicates significant circulation taking place in the downstream which may be a severe cause of scour. At x/L=1.4, there is significant X components velocity nearer the bed as compared to that at the top.

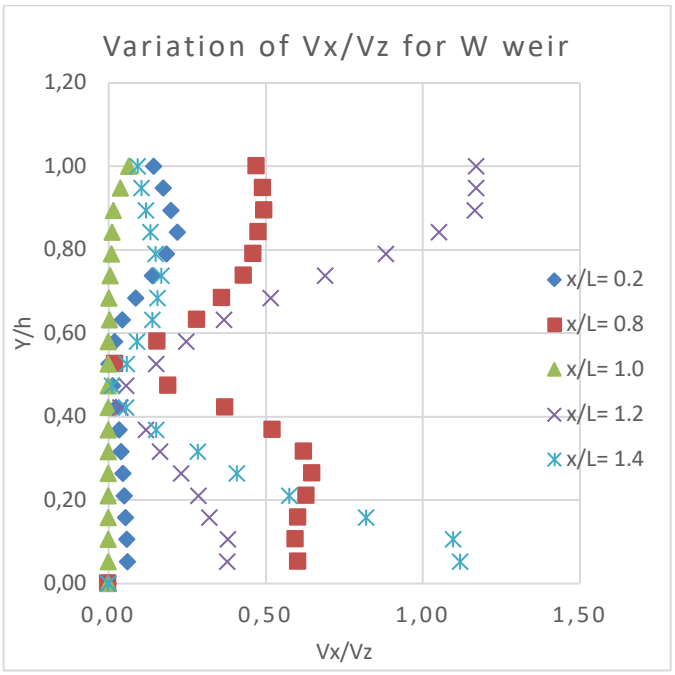


Figure 17: Variation of X velocity ratio in the channel for W weir from the inlet to the outlet

To investigate X components velocity variations figure 17 has been prepared that shows strength of secondary flows for W channels. It can be seen that at X/L=0.2, x components velocities are negligible, and are of same order of magnitude as seen in V weir. At x/L=0.8, similar trend as seen for V weir can be seen and the order of magnitudes are also similar. At the weir as expected the x component velocity is much smaller. At x/L=1.2, significant x components velocities can be seen that are of higher order of magnitude as seen for V weir and at x/L=1.4 also similar trends

can be seen. The above discussion has indicated that X components of velocities are higher of W weir as compared to V weir, which may be expected because a W weir consist of two V notches and the flow combination from the two notches results in significant flow redistribution in x direction. However because y component velocity is primarily the main reason behind the scour phenomena and hence from the earlier discussion it has been proven that the W weir will offer a better solution in alluvial channel and the possibility of scouring can be minimized in such channels.

Besides the findings of velocity field at the different sections of weir, the measurement of water level at the upstream side and at the crest level is also very important to calculate the coefficient of discharge of the weir. For the triangular weir, coefficient of discharge can be calculated by equating the discharge through 2 different formulas. For discharge to be calculated by the product of area and velocity, cross section area is taken into consideration with width being one parameter and water level being another (water level in the inlet is already simulated from the boundary conditions),which was then equated to the standard formula for the discharge calculation in the notch as discussed in the introduction.

For the W-shaped weir, the coefficient of discharge was measured separately for the triangles and discharge was also taken as half. Since they were identical and symmetric in measurements. The water level in both the V notches were found to be same.

Table 3: The discharge coefficient of the weirs

Type of weir	Angle of weir (Θ) (Degrees)	Coefficient of discharge
Triangular weir	90	0.61
W-shaped weir	53.12	0.632 0.632

It can be seen from the table 3 that the V weir has a discharge coefficient of 0.61 and the triangular weir has a discharge coefficient of 0.63.

5. Conclusions

The investigations carried out in this study has clearly indicated the flow features associated with a V notch and W notch. It has been clearly seen that velocity profiles are significantly affected by the presence of the weir. The resulting flow fields are complex. The region close to the weirs shows very complex behavior and flow becomes truly three dimensional at x/L=0.8. In the downstream of the weirs again the flow becomes three dimensional and x and y component velocities are significant. It has been seen that Y components velocities are higher in the downstream of V notch as compared to W notch that may be the reason for significant scour. However X components velocities are higher in the downstream of the W notch. There is no significant difference in the discharge coefficient of the weirs.

References

1. S.M. Borghei, Z Vatannia, Masoud Ghodsian, Masoud Ghodsian, Mohammad Reza Jalili Ghazizadeh, (2003), Oblique rectangular sharp-crested weir, ICE Proceedings Water and Maritime Engineering, Volume 156 Issue 2, June 2003, pp. 185-191.
2. Mysore N. Shesha Prakash, AV Shivapur, (2004)., Generalized Head-discharge equation for flow over sharp crested inclined inverted V notch weir, Journal of Irrigation and Drainage Engineering, 130(4).
3. Tuyen, N.B., (2006), Flow Over Oblique Weirs, M.Sc. Thesis. Delft University of Technology, Delft, the Netherlands.
4. Ashour, M.A., Bestawy, A.M., and Ali, K.A., (2008), Optimization of row numbers of baffles downstream heading-up structures for reducing scour hole, 1st International Conference of Civil Engineering Science, ICCESI, Vol. 2, pp. 82-88.
5. Y. A. Hamed, M. H. El-Kiki and A. M. Mirdan, (2009), Scour downstream oblique v-notch weir, Thirteenth International Water Technology Conference, IWTC 13, Hurghada, Egypt.
6. Issam A. Al-Khatib, (2013), Investigation of momentum and kinetic energy correction coefficients in asymmetric compound cross-section flumes, Turkish Journal of Engineering & Environmental Sciences, (2013) 37: 69 – 78.
7. C. Bautista-Capetillo, A.M.Asce; O. Robles; H. Júnez-Ferreira; and E. Playán, (2014), Discharge Coefficient Analysis for Triangular Sharp-Crested Weirs Using Low-Speed Photographic Technique, Journal of Irrigation and Drainage Engineering Vol. 140, Issue 3.
8. Saleh I. Khassaf, Ali N. Attiyah and Hayder A. Al-Yousify, (2016), Experimental investigation of compound side weir with modeling using computational fluid dynamic, International Journal of Energy and Environment (Vol. 7, Issue 2).
9. Hasan Ibrahim Al Shaikhli, Kadhim al-taeeKadhim al-taee, (2018), Development an equations for flow over weirs using mnlr and cfd simulation approaches, International Journal of Civil Engineering and Technology, Volume 9, Issue 3, March 2018, pp. 70–79.
10. Ansari U.S., Patil L.G., (2020), Numerical analysis of triangular labyrinth side weir in triangular channel, ISH Journal of Hydraulic Engineering.
11. Shiva Rezazadeh, Mohammad Manafpour, Hamzeh Ebrahimnezhadian, (2020), Three-Dimensional Simulation of Flow Over Sharp-Crested Weirs Using Volume of Fluid Method, Journal of Applied Engineering Sciences 10(1):75-82.
12. Vijay Kaushik, (2020), M.tech. dissertation on ‘Study of sediment transport on sharp crested triangular weir’, Hydraulics and water resources department, Delhi Technological University, New Delhi.

Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

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

Numerical study of microscale passive mixing in a 3-Dimensional spiral mixer design

Akash Sinha [a,*], M. Zunaid [a], Sulekh Tokas [a], Mubashshir Ahmad Ansari [b]

[a] Department of Mechanical Engineering, Delhi Technological University, Delhi 110042, India
[b] Department of Mechanical Engineering, ZHCET, Aligarh Muslim University, Uttar Pradesh 202002, India

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Micromixer
CFD
Spiral channel
Water
Passive mixing

ABSTRACT

Mixing of fluids becomes difficult if the length of the mixing channel is very small with size in microns. This is due to the lack of inertia and pre-eminence of viscous forces at low Reynolds numbers. The present study proposes a passive 3-dimensional spiral micromixer (SMM) and its Computational Fluid dynamics (CFD) study to scrutinize the mixing performance of water over a broad range of Reynolds numbers ($2 \leq Re \leq 320$). The obtained results were compared with a simple straight T-micromixer (STM) keeping the axial length same for both micromixers, which is 3000 μm. SMM appears to be much better and shows promising results with complete mixing (97.5%) of fluids species achieved at Re = 320 and minimum efficiency of 74.3 % achieved at Re = 66. Contrarily, STM shows comparatively poor results with a maximum mixing efficiency of 32 % for the highest Reynolds number (Re = 320). Thus, it is concluded that the proposed SMM design is an effective and significant choice for various micromixing applications.
Copyright © 2021 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Materials, Mechanics & Modelling.

1. Introduction

There is a growing interest in micro-total analysis system (μ-TAS) as it governs the overall efficiency of the μ-TAS and thus, enhancement of micromixing efficiency has been a focus of research for several years because micromixers are an essential part of it. Depending on the mechanism used to mix fluids in the microchannels, Micromixers can be broadly classified as Active and Passive. Active micromixers use external sources of energy like the electric field, magnetic field, vibrational, etc. to obtain enhanced mixing at a faster rate [1,2]. Whereas, passive mixing involves geometrical modifications of flow channels of the micromixer to disturb the flow and introduce the chaotic advection which enhances the mixing performance [3,4]. Since passive micromixers don't utilize external power, they are more cost-effective and easier to integrate with the complete μTAS. Although microscale synthesis and analysis carry various advantages like the low inventory requirement, higher magnitude of surface-to-volume ratios, better control over flow rates, and reduction in the analysis time [5]. The major bottleneck for employing the micro-total analysis system devices is the inefficient mixing at the microscale. The flow is laminar and dominated by the viscous forces leaving mass diffusion to be the sole mechanism of mixing which is a time-consuming process with the requirements of prolonged lengths of the microchannel to attain the desired results. The most basic micromixer has a simple T-shape straight channel structure with two inlets and one outlet. Many studies including both experimental and numerical have been to understand the mixing patterns in this type and can be referred to in the literature [6–10].

Several Modifications were introduced in the straight channel to mix two or more fluids at a faster rate. Andreussi et al. [10] examined different flow regimes in a simple T-micromixer using a numerical model for three different aspect ratios (width-to-height ratios of the channel) that are 0.75, 1, and 2. At low Reynolds number, using large aspect ratios of mixer appeared detrimental for the mixing. Ansari et al.[11] numerically studied the effect of the location of the fluid stream interface on the mixing performance and found that the location of contact region doesn't have a significant effect on the mixing for $Re \leq 10$ but for higher Reynolds number, a notable difference in mixing index was observed. Several researchers suggested modifications in the straight channel T-micromixer to increase the mixing efficiency.

* Corresponding author.
E-mail address: sinha76akash@gmail.com (A. Sinha).

https://doi.org/10.1016/j.matpr.2022.02.508
2214-7853/Copyright © 2021 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Materials, Mechanics & Modelling.

Please cite this article as: A. Sinha, M. Zunaid, S. Tokas et al., Numerical study of microscale passive mixing in a 3-Dimensional spiral mixer design, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2022.02.508

Sarkar et al. [12] numerically investigated different designs and arrangements of symmetric and asymmetric inlet junctions. It was observed that asymmetric type performed better than the conventional symmetric inlet junctions whereas W-type junction presented the best results amongst the 1–1 type of junctions evaluated in the study. Sarkar et al. [13] used CFD to study the effect of inward wall protrusions of basic shapes namely rectangular, semi-circular, and triangular on the mixing performance of fluid streams. The semi-circular shape provided the best mixing out of all the conventional shapes and several new designs of protrusions apart from basic shapes were also examined. Moreover, the pressure drop that occurred inside each micromixer was calculated for all the designs and a comparative study was presented. More work on the similar simple modifications of the straight channel micromixer can be referred to in the available wide literature [14–17].

One of the effective ways to advance the strength of the secondary flows in a passive micromixer is to design a curved flow channel that will impart the centrifugal forces on the inner walls of channels leading to mixing enhancement. Sudarsan and Ugaz [18] performed experiments to analyze mixing performance in five in-plane spiral microchannels for a wide range of Reynolds Numbers (0.02 to 18.6). At high mass flow rates, secondary dean vortices were visible in the flow path which improved the mixing performance within the system. Afzal and Kim [19] used CFD simulations to study and propose a sinusoidal channel walls passive micromixer. It was inferred from the results that sinusoidal micromixers were more efficient than the previously available square-wave and zig-zag micromixers at all the values of flow rates. Vatankhah and shamloo [20] numerically studied the effect of cross-sectional shape and area on the mixing performance of a planar spiral passive micromixer. A parametric study was conducted it was inferred that the square cross-sectional (aspect ratio = 1) area performed better than the other cross-sections of the mixing channels. Similarly, various other spiral and helical micromixers were proposed by several researchers [21–23].

Most of the research work focuses on mixing analysis with an in-plane spiral channel having circular or/and square cross-section. Numerical analysis is carried out to study the flow characteristics of the fluids and the mixing performance of water with the suggested 3d-spiral micromixer (SMM) and the simple T-micromixer (STM). The study analyses the performance of micromixers with working fluid water with the help of CFD simulations. Furthermore, a detailed study on pressure drop is presented which is a strong indicator of the energy requirements of a micromixer.

2. Micromixer model and specifications

The extracted flow domains of the two micromixers used in the current investigation are shown in Fig. 1. To perform the analogy, the axial length of both the micromixers are fixed (3 mm). The inlet cross-sections are square with a side of 100 μm, whereas the cross-section of the mixing channel is rectangular (200 × 100 μm). The top-view of conventional straight channel micromixer (SCM) is illustrated in Fig. 1(a). Fig. 1 (b) depicts the proposed spiral channel design of a micromixer with geometrical parameters like pitch (P_s), number of spiral turns (n), mean spiral diameter (d_s), etc. summarized in Table 1.

Table 1
Spiral channel Micromixer (SCM) Geometrical definition.

Parameter Name of Spiral Micromixer	Magnitude
Pitch of the spiral curve (P_s)	370 μm
Mean diameter of the spiral curve(d_s)	380 μm
Number of spiral turns (N)	7
Flow path length (L_f)	8.74 mm
Axial length (L_{ax})	3 mm
Inlet channel length (L_i)	300 μm

3. Numerical methodology

The governing equations were discretized and numerically solved using ANSYS FLUENT to calculate the mixing performance and flow characteristics of micromixers. The basic governing equations are given as

Continuity:

$$\nabla.\vec{V} = 0 \tag{1}$$

Navier-stokes Equation:

$$\rho\left(\vec{u}.\nabla\right)\vec{u} = -\nabla p + \nabla.\vec{\tau} \tag{2}$$

Where $\vec{\tau}$ is the stress tensor.

$$\vec{\tau} = \eta\left(\nabla\vec{u} + \nabla\vec{u}^{t}\right) \tag{3}$$

Species conservation equation:

$$\left(\vec{V}.\nabla\right)C_x = D_B\left[\frac{\partial^2 C_x}{\partial x^2} + \frac{\partial^2 C_x}{\partial y^2} + \frac{\partial^2 C_x}{\partial z^2}\right] \tag{4}$$

Where $\vec{V}$ is the velocity vector, C_x is the species concentration × and D_B is the diffusion coefficient. Normal velocity component boundary condition was specified at both the inlets whereas pressure outlet boundary is specified at the outlet of the micromixer. Water is considered as the working fluid for the current analysis and the respective properties of water at 20 °C were considered. The dynamic viscosity, density, and diffusivity are constant at 0.001 Pa-s, 998.2 Kg/m3, and 2 × 10–9 m2/s, respectively.

Qualitative analysis of mixing at various planes and interfaces was performed with the help of a broadly used parameter known as the mixing index. The mixing index can be obtained by

$$Mi = \left(1 - \sqrt{\frac{\sigma^2}{\sigma^2_{max}}}\right) \times 100 \tag{5}$$

$$\sigma^2 = \frac{1}{n}\sum_{j=1}^{n}\left(c_j - c_{avg}\right)^2 \tag{6}$$

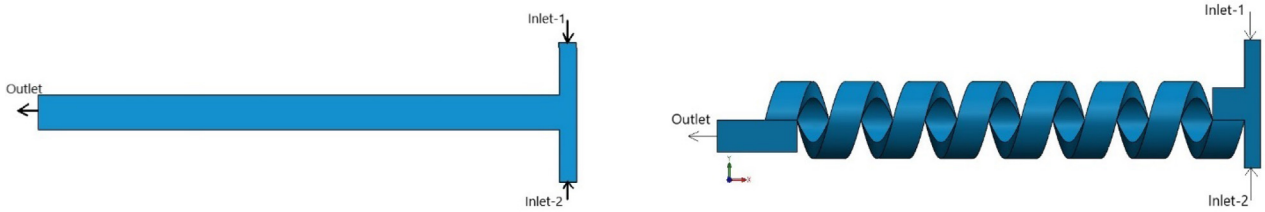


Fig. 1. Model description of (a) Straight channel micromixer (STM); (b) Spiral design Micromixer (SMM).

Where σ^2 and σ^2_{max} are the actual and maximum variance of the mass fraction of the fluid respectively, considered to be on a plane. c_j and c_{avg} are the value of actual species concentration and average species concentration respectively. Here, n is the total number of considered points over the plane which is 800 for the current study. The numerical solutions were considered to converge when the root mean square value of the residual reached 10–9. The accuracy and reliability of the numerical setup were verified with a validation test and results were compared with the existing literature of Dundi et. al [24]. The mixing index was calculated at various axial distances along the simple T- micromixer at Re = 266 for this particular case and it was found that all the results were in good agreement with that of the existing literature. Fig. 2 demonstrates the comparison of mixing index results between the current numerical method and that of the results of Dundi et. al. A grid independence test was also performed to make the numerical solutions more reliable and approximately 3.2 million mesh elements were picked to proceed with the analysis as good accuracy was achieved at that number of elements. Fig. 3 depicts a grid sensitivity analysis and shows the change in the pattern of mixing efficiency with an increase in the number of elements used in the simulation for the conventional T-micromixer at Re = 235. The mixing tends to appear a constant value after a certain number of elements telling about the grid independence of the numerical results.

4. Results and discussion

As discussed previously, to present a comparative picture, SMM axial length of the mixer is fixed and is equal to the length of the conventional TM with a straight channel, i.e., 3000 μm. Although the axial length is assumed to be constant for the micromixers, the mixing length or the peripheral length of SMM is higher (~8810 μm) than STM due to the helical path of the channel introduced in the proposed design.

First, a comparison of mixing performance is presented between the TM and the SMM based on the CFD visualization. Fig. 4 (a) and (b) depicts the streamlines defined by the species concentration of water for STM and SMM, respectively at Re = 266.

It is inferred from the figure that higher mixing efficiency (87.8 %) is attained in the SMM as compared to the STM with a mixing index of 30.7 % at the same Reynolds number of 266. Closer observation reveals that the complex 3-dimensional design of the mixing channel of SMM produces significantly better mixing. The twisting channel of the spiral micromixer provides rotation to the fluid, this curvature then introduces the centrifugal forces at

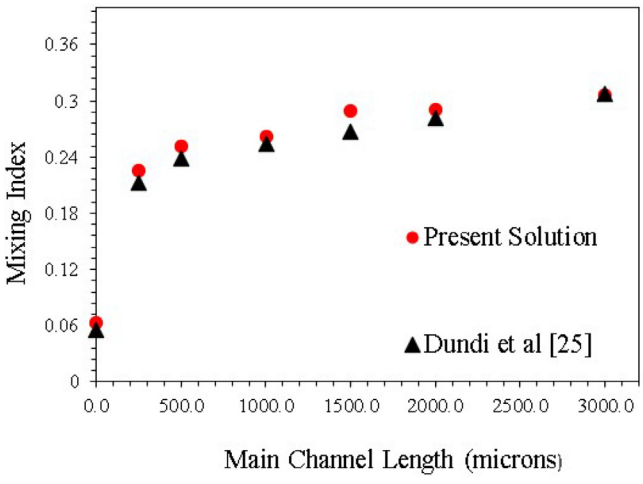


Fig. 2. Validation test comparison for the reliability of the numerical methodology.

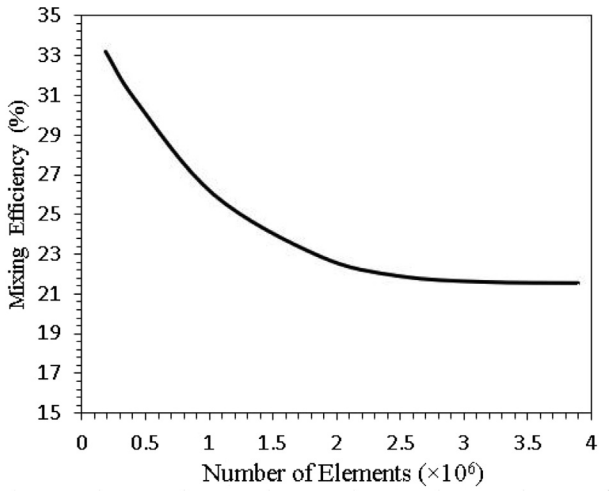


Fig. 3. Grid sensitivity study.

the walls of the channel. The direction of this force is changing due to its spiral shape as the fluid travels in the axial direction towards the outlet causing the enhancement in the mixing performance. The secondary flow enhances the mass diffusion and thus the mixing performance due to the creation of chaotic advection. Secondary flow usually develops around curved three dimensional surfaces and are perpendicular to the main direction of fluid flow. Fig. 5 represents the variation of the mixing index with Reynolds for both STM and SMM. It is obvious that at lower values of Reynolds number (Re $\leq$ 10), the fluids have enough time to be in contact with mass diffusion to take place even though the flow falls in the laminar diffusion regime. Thus, a mixing index of 11.2 % for STM and 80.7 % for SMM is obtained at Re = 2. So, the basic reason for the higher mixing efficiency in SMM is due to, firstly, higher mixing length of the SMM compared to STM and more stretching and secondly due to the folding of the flowing fluid due to the 3-dimensional twisting of the channel. Because this twisting of the channel enhances the area of contact between the two streams promoting the diffusion process.

With an increase in Reynolds number, weak vortices are visible however a decreasing trend in the mixing index is noticed owing to the decrease in the residence time of fluids inside the mixing channels. This results in a lesser time of contact between the two species and the weak strength of vortices don't seem to have a significant impact on the mixing efficiency. The lowest value of the mixing index is 3.9 % for the conventional STM and 74.3 % for the proposed SMM in this study, both at Re = 66.

Here, the Vortex diffusion regime is observed until Re = 160 for STM and Re = 100 for SMM, subsequently, the flow enters into the engulfment flow, and a drastic rise in the mixing performance is noted. Together with the engulfment flow, stronger secondary flow is visible in the SMM due to the continuous rotation, twisting, and stretching of the fluid layer. This amplifies the mixing performance and complete mixing of fluid streams (97.5 %) is achieved at the maximum Reynolds number of 320 considered in the current investigation, whereas the maximum efficiency is 31.9 % for the straight channel STM. Fig. 6 depicts the graph between the mixing index of SMM with the axial distance (X) from the T-junction.

Axial distance (see Fig. 6,Fig. 7) is the distance between the T-junction and the centre of the plane taken for the calculation of the mixing performance, where X = 0 μm represents the first plane as the fluid enters the helical channel after the T-junction. The sharpest rise in the mixing performance from the inlet (5.89%) of the helical channel to the outlet (80.7 %) is seen for Re = 2 due to the larger time available for the molecular diffusion to take place

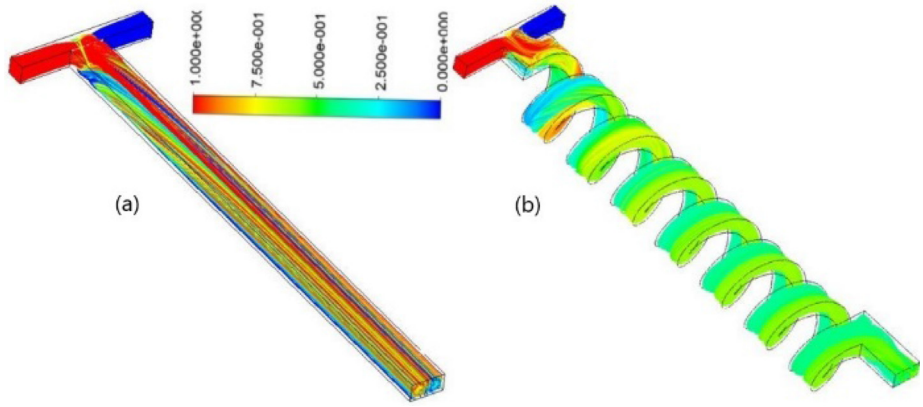


Fig. 4. Mixing performance with water for (a) STM and (b) SMM at Re = 266.

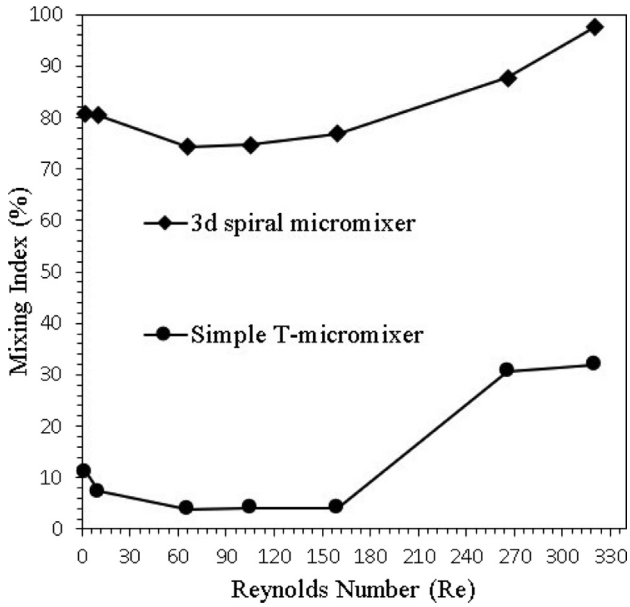


Fig. 5. Variation of mixing index with Reynolds number.

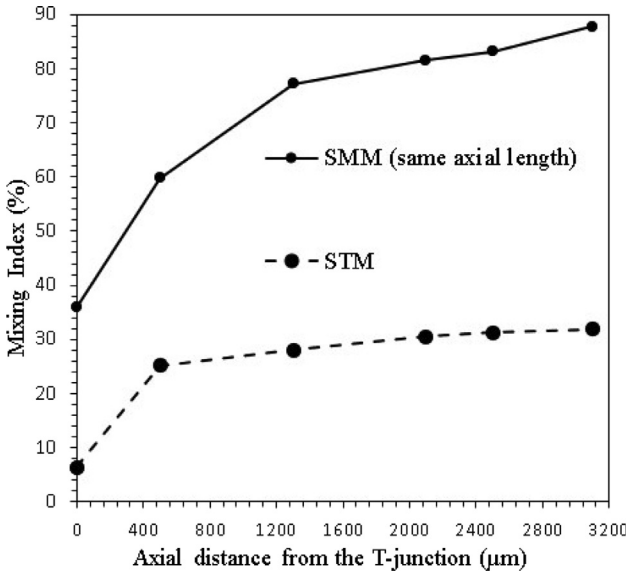


Fig. 7. Comparison of the mixing index of STM and SMM at various cross-sections along with different axial distances for Re = 266.

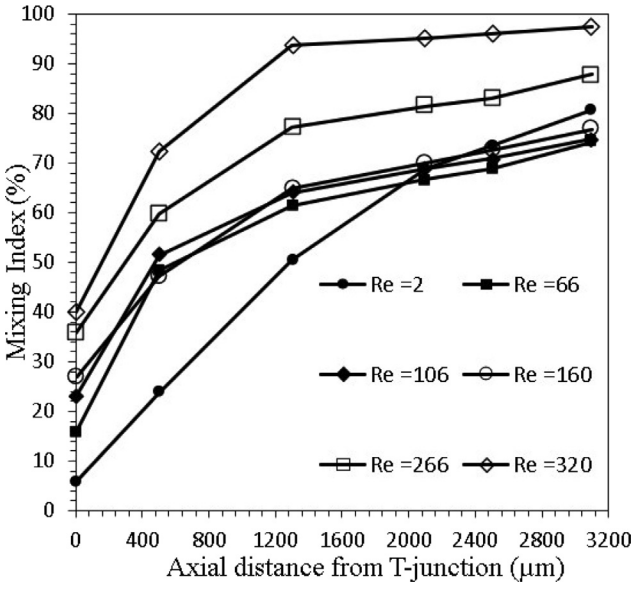


Fig. 6. Change in the value mixing index with the axial distance for different values of Reynold number of water.

as discussed earlier (called the residence time). The slope of mixing efficiency at different axial distances for Re = 66, 106 & 160 appears to be almost the same. The mixing index for Re = 320 rises from 40% to 97.5 % resulting in the complete mixing of the two streams. The variation of mixing index with axial distance from the inlet plane of the SMM is shown in Fig. 7 for Re = 266. The slope of the curve of STM appears to flatten after the axial length of 500 μm with a small increase in efficiency from 25.1 % at 500 μm to 31.9 % at the outlet section of the conventional STM. At high Reynolds number as in this case, Re = 266 the initial collision of the fluid streams at T-junction tend to provide a swirling motion to the fluids and penetrate each other but owing to the straight channel in STM and lack of inertial forces in the flow field, there is an only mild increase in the mixing efficiency after 500 μm. Whereas, a continuous rise is observed from 35.9 % on the inlet section of SMM to 87.8% at the outlet section because of the stronger strength of secondary flow. Therefore, the SMM is a better choice over STM to achieve superior mixing within a short axial distance for both low and high Reynolds number flow.

To visualize the mixing of fluids achieved at different sections considered at different axial distances from the inlet, contours of species concentration at different planes for five different Reynolds numbers that are 2,66, 106, 160 & 320 is shown in Fig. 8. At Re = 2,

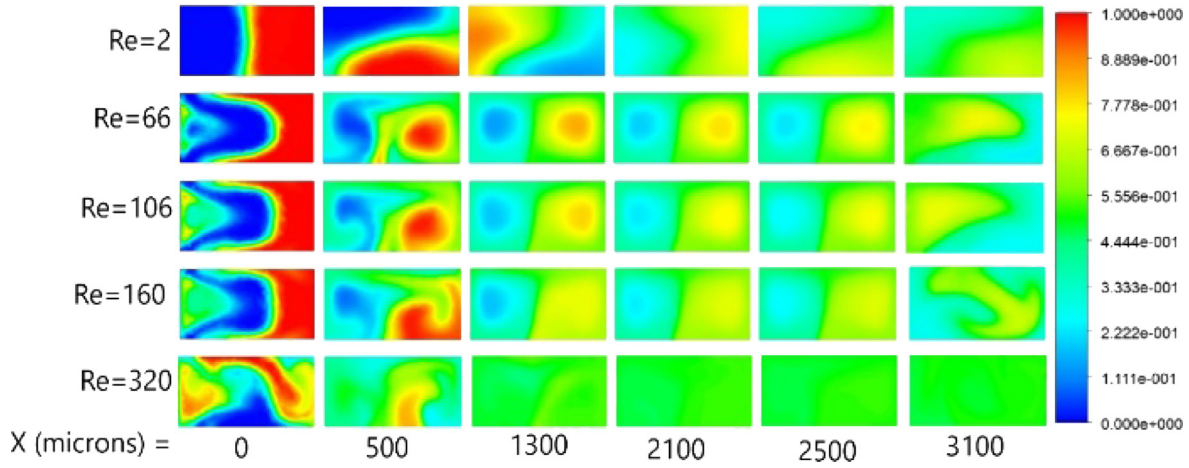


Fig. 8. Species concentration contour of water in SMM at different sections.

the interface between the mixing fluid emerges narrow and thin on initial planes informing about the laminar diffusion regime with no signs of vortices or secondary flow. Still, good mixing is achieved at the outlet because of the stretching of the fluid species and the high time of contact (high residence time) of the fluids inside the microchannels as seen in the successive sections of Re = 2 in Fig. 8. For Re = 66, 106 & 160, the interfacial area between the fluid species increases as the flow progress to further planes. This is due to the weak vortices which are introduced with an increase of the Reynolds number. At Re = 320, the inertia of the fluid supports the rotation of the fluid streams, and the strong secondary motion results in the complete mixing of the streams at the outlet section.

Pressure drop is a crucial parameter to understand the power requirements to drive the fluids inside the micromixer. Fig. 9 represents how pressure drop changes with the Reynolds number across the inlets and the outlet of the SMM. It can be seen that with the increase in Reynolds number the pressure drop increases rapidly.

At the lowest value of Reynolds number considered in the study, i.e., 2 the pressure drop is 0.278 KPa whereas it is 1.40 KPa at Re = 10. This increasing trend is observed throughout the wide range of Reynolds numbers with a maximum pressure drop of 95.183 KPa calculated at the maximum value of Reynolds number, i.e., 320. Since the mixing lengths of the STM and SMM are different in this case, it is not feasible to make a comparison between the pressure drops of the two micromixers.

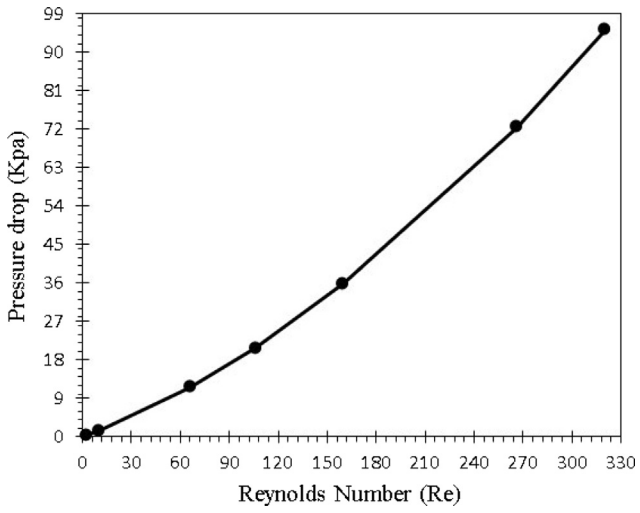


Fig. 9. Pressure drop variation in SMM with Reynolds number.

5. Conclusions

A CFD study has been carried out to study and compare the mixing efficiencies and performances of a simple T-micromixer (STM) & a novel 3-dimensional spiral micromixer (SMM) where water is considered as the fluid to be mixed. It is concluded that the discussed micromixer SMM provided better mixing efficiency and performance for all the Reynold number values taken in the study with almost complete mixing achieved fluids mixed completely (97.5%) at a high Reynolds number of 266 making it a promising design for implementation at microscale mixing.

CRediT authorship contribution statement

Akash Sinha: Data curation, Writing- Original draft preparation, Visualization, Investigation, Writing- Reviewing and Editing. **M. Zunaid:** Conceptualization, Methodology, Software, Supervision. **Sulekh Tokas:** Data curation, Writing- Original draft preparation, Visualization, Investigation, Writing- Reviewing and Editing. **Mubashshir Ahmad Ansari:** Conceptualization, Methodology, Software , Supervision.

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

We gratefully thank Delhi Technological University for their support.

References

[1] W. Raza et al., Active and passive micromixers: A comprehensive review, Micromachines 11 (9) (2020).
[2] V. Hessel, H. Löwe, F. Schönfeld, Micromixers - A review on passive and active mixing principles, Chem. Eng. Sci. 60 (8-9) (2005) 2479–2501.
[3] G. Cai, L.i. Xue, H. Zhang, J. Lin, A review on micromixers, Micromachines 8 (9) (2017) 274, https://doi.org/10.3390/mi8090274.
[4] W. Raza, S. Hossain, K.-Y. Kim, A Review of Passive Micromixers with a Comparative Analysis, Micromachines 11 (5) (2020) 455, https://doi.org/10.3390/mi11050455.

[5] A. Gu, Multiphase microfluidics: from flow characteristics to chemical and materials synthesis, 2006, pp. 1487–1503.
[6] D. Bothe, C. Stemich, H.-J. Warnecke, Computation of scales and quality of mixing in a T-shaped microreactor, Comput. Chem. Eng. 32 (1-2) (2008) 108–114.
[7] D. Bothe, C. Stemich, H.-J. Warnecke, Fluid mixing in a T-shaped micro-mixer, Chem. Eng. Sci. 61 (9) (2006) 2950–2958.
[8] A.S. Lobasov, A.V. Minakov, V.V. Kuznetsov, V.Y. Rudyak, A.A. Shebeleva, Investigation of mixing efficiency and pressure drop in T-shaped micromixers, Chem. Eng. Process. - Process Intensif. 134 (2018) 105–114.
[9] A. Soleymani, E. Kolehmainen, I. Turunen, Numerical and experimental investigations of liquid mixing in T-type micromixers, Chem. Eng. J. 135 (2008) S219–S228.
[10] T. Andreussi, C. Galletti, R. Mauri, S. Camarri, M.V. Salvetti, Flow regimes in T-shaped micro- mixers, Comput. Chem. Eng. 76 (2015) 150–159.
[11] M.A. Ansari, K.-Y. Kim, S.M. Kim, Numerical study of the effect on mixing of the position of fluid stream interfaces in a rectangular microchannel, Microsyst. Technol. 16 (10) (2010) 1757–1763.
[12] S. Sarkar, K.K. Singh, V. Shankar, K.T. Shenoy, Numerical simulation of mixing at 1–1 and 1–2 microfluidic junctions, Chem. Eng. Process. Process Intensif. 85 (2014) 227–240.
[13] S. Sarkar, K.K. Singh, V. Shankar, K.T. Shenoy, CFD simulations to study the effects of wall protrusions on microfluidic mixing, J. Micromech. Microeng. 25 (8) (2015) 084008, https://doi.org/10.1088/0960-1317/25/8/084008.
[14] S. Hossain, K.-Y. Kim, Numerical Study on Mixing Performance of Straight Groove Micromixers, Int. J. Fluid Mach. Syst. 3 (3) (2010) 227–234.
[15] M. Kashid, A. Renken, L. Kiwi-Minsker, Mixing efficiency and energy consumption for five generic microchannel designs, Chem. Eng. J. 167 (2-3) (2011) 436–443.
[16] A.A.S. Bhagat, E.T.K. Peterson, I. Papautsky, A passive planar micromixer with obstructions for mixing at low Reynolds numbers, J. Micromech. Microeng. 17 (5) (2007) 1017–1024.
[17] N. Ait Mouheb, D. Malsch, A. Montillet, C. Solliec, T. Henkel, Numerical and experimental investigations of mixing in T-shaped and cross-shaped micromixers, Chem. Eng. Sci. 68 (1) (2012) 278–289.
[18] A.P. Sudarsan, V.M. Ugaz, Fluid mixing in planar spiral microchannels, R. Soc. Chem. 6 (1) (2006) 74–82.
[19] A. Afzal, K.-Y. Kim, Mixing performance of passive micromixer with sinusoidal channel walls, J. Chem. Eng. Japan 46 (3) (2013) 230–238.
[20] P. Vatankhah and A. Shamloo, AC SC, Anal. Chim. Acta, 2018.
[21] S. Tokas, M. Zunaid, M.A. Ansari, Numerical investigation of the performance of 3D-helical passive micromixer with Newtonian fluid and non-, Newtonian fluid blood 16 (1) (2021), https://doi.org/10.1002/apj.v16.110.1002/apj.2570.
[22] L. Balasubramaniam, R. Arayanarakool, S.D. Marshall, B. Li, P.S. Lee, P.C.Y. Chen, Impact of cross-sectional geometry on mixing performance of spiral microfluidic channels characterized by swirling strength of Dean-vortices, J. Micromechanics Microengineering 27 (9) (2017) 095016, https://doi.org/10.1088/1361-6439/aa7fc8.
[23] J. Yang, L.i. Qi, Y.i. Chen, H. Ma, Design and fabrication of a three dimensional spiral micromixer, Chinese J. Chem. 31 (2) (2013) 209–214.
[24] T.M. Dundi, V.R.K. Raju, V.P. Chandramohan, Numerical evaluation of swirl effect on liquid mixing in a passive T-micromixer, Aust. J. Mech. Eng. 00 (00) (2019) 1–15.

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/TCDS.2022.3158613, IEEE Transactions on Cognitive and Developmental Systems

Obscenity Detection in Videos through a Sequential ConvNet Pipeline Classifier

Neil Gautam, Dinesh Kumar Vishwakarma, *Senior Member, IEEE*

***Abstract*- The amount of pornographic material available on the web is staggering. This content is available freely on the internet and without any restrictions which pose a threat to minors as they might be introduced to such content which is harmful to their mental health for the long term and also there is huge threats of pornographic content generation to the celebrities, known-figures due to invention of deep-fakes. There are many software that blocks access to the visually disturbing sites that contains obscene, child pornography or material "harmful" to minors but only a few software analyses motion content of the videos and mostly only image features are analyzed. To tackle this problem, we propose a Frame Sequence ConvNet Pipeline using ResNet-18 for features extraction and analyzing N frame feature-map using a proposed ConvNet for the frame sequence classification, therefore implicitly encapsulating the motion information by encoding changes in the ResNet output feature vector which achieves a state-of-the-art accuracy of 98.25% in classifying pornographic videos of Pornography-800 dataset and state-of-the-art accuracy of 97.15% in classifying videos of Pornography-2k Dataset.**

***Index Terms*-**Porn-detection, Convolution Neural Networks, Human Detection Model, End-to-end trainable Network.

I. INTRODUCTION

WITH the growing access to the internet, everyone stumbles upon adult content intentionally or unintentionally. Pornographic scenes sometimes contain physical aggression, including spanking, gaggling, slapping and verbal aggression, including abusing. This type of disturbing content is harmful to minors, especially considering how their mental health might be affected for the long term.

Recently, a major challenge in pornography that has been observed is that fake porn videos of different celebrities were created through deep learning based approaches and these videos seem to be realistic visually while these are just deep fakes [1]. Hence, the detection of pornography is becoming very important over the era of the internet.

As reported by Quadra et al [2], teenagers who watch pornography frequently are more likely to cultivate aggressive behaviour among themselves. This creates a necessary c[1]ondition to find solutions that are accurate, computationally less expensive, and fast both at the time of training and inference.

Pornography is a subjective topic that does not have a fixed definition. There has been no clear definition of something, whether it is pornographic or not. For this work, we consider pornography to be a visual depiction in the form of images or videos having sex organs or activities, which is mainly aimed to simulate sexual excitement in a person.

In this paper, we propose a Frame Sequence Classification pipeline to classify and detect pornographic video's frame sequence. We attempt to achieve high accuracy by using the network which does not use motion information explicitly as used by Perez et al [3] where they trained different streams for spatial and temporal data. Also, our technique requires lesser data pre-processing as compared to Perez et al [3], as they classify pornographic videos using a two-stream network inspired by Simoyan et al [4], processing static and temporal information in separate streams and also calculate optical flow over frames which also increase computational cost. Our technique of feature extraction using ResNet is inspired by Wehrmann et al [5], but the way we use that information is entirely different. The main contribution of the work is as follows:

- We designed a two components single stream pipeline for classifying the sequence of N (8 or 16) continuous frames from video using fine-tuned ResNet-18 on frames to work as feature extractor network and feeding its outputs of $R^{N\times 512}$, where N is sequence length, to the Sequence Classifier ConvNet that we propose in this work.
- The two different Sequence ConvNets Classifiers are built up for N frames feature-map, 8 and 16 respectively, whose input is feature-map of $R^{N\times 512}$ made from the consecutive output of ResNet-18 extracted features. Our method aims to achieve high accuracy while using low parameters of deep learning model.
- We use filtering while feeding data to the network when the data is noisy. We detect that each sequence that we pass should contain humans in the frames in case of pornography. We use TensorFlow Object Detection API 2.1 [6] to filter frame sequence which contain humans

(Corresponding author: Dinesh Kumar Vishwakarma)
Neel Gautam and Dinesh Kumar Vishwakarma are with the Biometric Research Laboratory, Department of Information Technology, Delhi Technological University, (Formerly Delhi College of Engineering) Delhi-110042, India (aineilgautam@gmail.com, dvishwakarma@gmail.com)

2379-8920 (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 March 23,2022 at 07:10:22 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/TCDS.2022.3158613, IEEE Transactions on Cognitive and Developmental Systems

while passing data in the pipeline in case of porn category data.

- For validating the effectiveness of this pipeline, we tested this pipeline on Pornography-800 by Avila et al [7] and Pornography-2k dataset by Moreira et al [8] and achieved high accuracy as compared to earlier state-of-the-arts.

This paper is organized in the following way. In section II, we discuss the previous methods that were used, both handcrafted feature and deep learning-based. In section III, we propose the methodology that we have used in this work, including all the steps of our work intuitively. In section IV, we present the experimental methodology, which includes the evaluation measures that we have used and ConvNets architecture hyperparameters settings. In section V, we present the conclusion of our work and future scope.

II. Related Work

The foundational work of pornographic detection started when researchers started approaching the problem by detecting nudity [5], skin-based features [9]. The first approach using deep learning for pornographic detection was introduced by Moustafa et al [10] where Ensemble of AlexNet [11] and GoogLeNet [12] Architectures was used for getting the majority vote to achieve the best performance for obscenity classification. Moustafa et al [10] used CNN architectures pre-trained on ImageNet, This work by Moustafa et al [10] was further improved by Wehrmann et al [5] where they extracted the features from the various ResNet models and processed the sequence of extracted features using the RNN Model. Our work is somewhat inspired by the work of Simoyan et al [4] that they have done, but our approach to process sequence is different. Also, we have used frame filtering using Human Detection Faster-R-CNN model when the data was present with redundant information. Along with that, we have tried to achieve maximum accuracy using models with minimum parameters to make the inference of the model as fast as possible.

The most recent work on pornographic detection, which achieves the highest accuracy is by Perez et al [3], they have applied a two-stream approach i.e. one stream to classify the static frames and the other for motion detection. They encapsulated motion information using MPEG vectors and dense optical flow. They trained separate stream for static and motion features and merged them using various fusion techniques to get a majority vote to classify the videos. Deep learning technique proposed by Perez et al [3] achieved the highest adult video classification accuracy and was the first work to classify videos considering motion information. A comparatively less explored deep learning approach for pornography detection was proposed by Silva et al [13], where the author explored the Convolution3D based CNN architectures to encapsulate Spatio-temporal information in a single stream network using Conv3D layer which learns Spatio-temporal embeddings on training from the videos.

To review how research in adult content detection problem started, refer to work by Short et al [14] for the detailed analysis of the earlier approaches that did not use deep learning techniques.

III. Proposed Methodology

This section discusses the proposed approach that we have used in this work, as shown in Fig.1 and Fig.2. The proposed work consists of human detection in a frame, selecting key frames, feature extraction, and frame-sequence classifier. The proposed approach has been devised to develop a precise pornographic detection method. The details of the proposed work are described in the following subsections.

A. Feature Extraction Network

While preparing the data for the feature extraction network, we extract the frames from videos at 4 FPS (frame per second) rate for this work and this rate is selected based on observing the nature of data and experimental observation with the aim to encapsulate small motion changes in the sequence of frames and process these images at the size of 224×224. This choice was more based on the experimental side on two choices i.e faster prediction while using at the test time on videos and the other is to encapture the small motion changes that happens over time while processing the sequence with the minimum number of frames. If we were to process the video at a normal frame rate i.e 30 FPS, then it would make training a computationally expensive operation and also for capturing small motion changes would require us to use more frames. Selection of key frames from the extracted frames for both datasets is done manually or randomly, by picking up the equal number of random frames for each video so that the data remains balanced throughout fine-tuning of the ResNet-18 [15] for this specific task. Before feeding the images to the ResNet-18, we use the data augmentation techniques such as normalization with Image-Net statistics across each RGB channel, color-jitter, random rotation at 45 °, random horizontal-flip or vertical-flip, so that network learns the generalized features and does not overfit.

Let us consider a set of k frames v extracted from the videos at 4 frame-per-second-rate,

$$v \in \{ v_1, v_2, v_3, \dots \dots \dots \dots, v_k \} \quad (1)$$

Let us consider the ResNet-18 to be a function represented by $f(v)$, $f(v) \in R^{512}$. After training, the classification layers are removed and weights are freezed to function as feature extractor, that leaves the output of $R^{7\times7\times512}$. We average pool that layer using $Average\ Pool\ 2d$ layer which gives the output as $f(v(i))$ as R^{512}, which is the required output to further pass on to the sequence classifier network. In Fig. 1 in the green box, it has been shown how this model structure is incorporated in the overall architecture of the pornographic video classification framework where its function is to extract the features of images.

We could have used other common ResNet Architectures such as ResNet-51, 101, et cetera. But we wanted to make sure two things, first is that we should have the model with lesser parameters and second is that we have lesser dimensional feature vector output, which is R^{512} vector, whereas in the case

2379-8920 (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 March 23,2022 at 07:10:22 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/TCDS.2022.3158613, IEEE Transactions on Cognitive and Developmental Systems

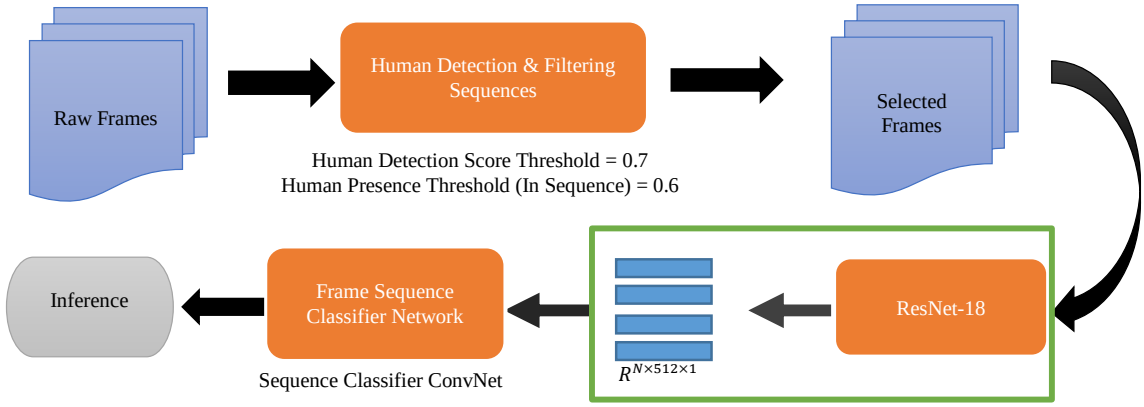


Fig 1. Framework for Pornography Classification. Green Box here represents the feature extraction network.

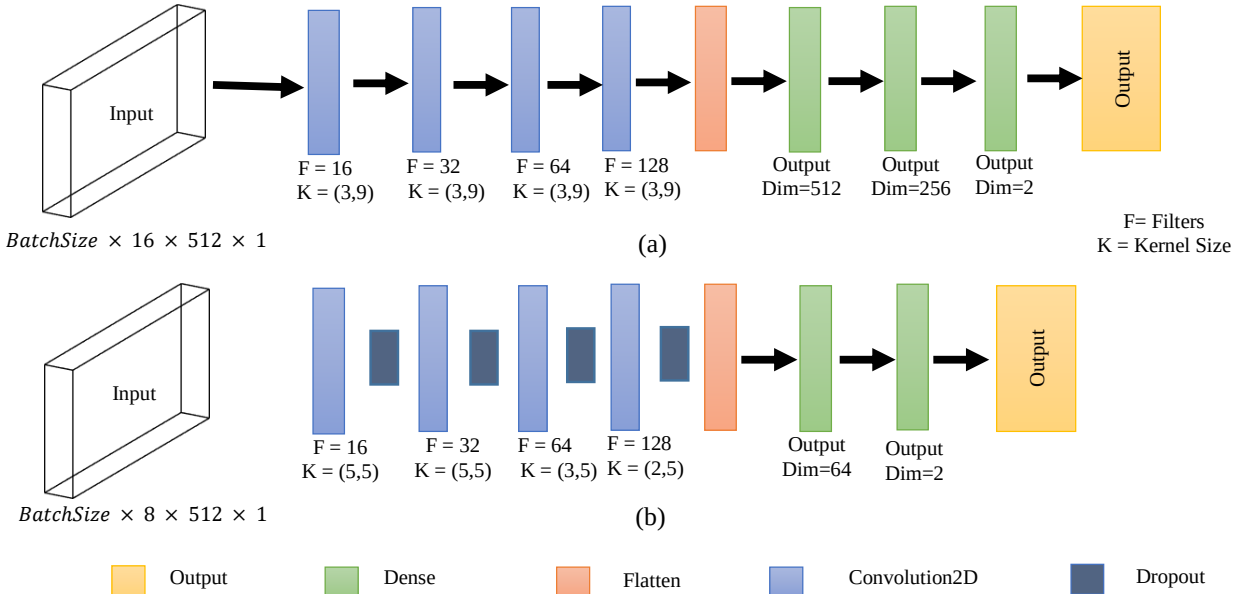


Fig 2. (a) Sequence Classifier ConvNet for frame sequence 16, (b) Sequence Classifier ConvNet for frame sequence 8

of ResNet-34 the output is R^{1024} , and in the case of ResNet-51 and 101 output vectors is R^{2048} on removing the classifier layer and average pooling the layer.

B. Sequence Classifier Convolutional Network

After the ResNet-18 outputs the R^{512} vector from a single frame, we combine a sequence of N consecutive frames feature vector to create a feature map of $R^{N\times512\times1}$, which is then fed to the Sequence Classifier ConvNets as shown in Fig. 2.

Let us consider l be the set of k output of $f(v)$, where the input is v as given in Eq. 1,

$$l = \{f(v(1)), f(v(2)), f(v(3)), \dots\dots\dots, f(v(k))\} \quad (2)$$

$$l(i) = f(v(i)) \quad (3)$$

$l(i)$ can be represented as in Eq. (4), we can represent l as,

$$l = \{l_1, l_2, l_3 \dots\dots\dots, l_k\} \quad (4)$$

where $l(i) \in R^{512}$, from the extracted feature set l, we make two sets of feature-map f_1 and f_2, for two different values of N, i.e 8 and 16 respectively, for $N = 8$,

$$f_1(i) = l(i) \dots l(i+N) \quad (5)$$

where $f_1(i) \in R^{N\times512\times1}$.

presented as in Eq. (4), we can represent

$$f_1 = \{f_1(1), f_1(2), \dots\dots\dots,\} \quad (6)$$

$$f_1 = \{f_1^1, f_1^2, f_1^3, \dots\dots\dots,\} \quad (7)$$

Similarly, for $N = 16$, $f_2 \in R^{N\times512\times1}$

$$f_2 = \{f_2^1, f_2^2, f_2^3, \dots\dots\dots,.\} \quad (8)$$

As we can observe in Eq. 7 and Eq. 8, f_1 and f_2 represents the sequence feature map inputs that will be fed in the Sequence Classifier ConvNet, where we choose K number of random

2379-8920 (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 March 23,2022 at 07:10:22 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/TCDS.2022.3158613, IEEE Transactions on Cognitive and Developmental Systems

feature-map from the set f_1 and f_2 from a single video for feeding to the network given in Fig. 2(a) and 2(b) respectively.

The Sequence Classifier ConvNets shown in Fig. 2 shows the network to classify the feature-map of sequence length 16 and 8 respectively. It aims to encapsulate sequence dependencies and detect the type of motion that is present in the feature map of the sequence. We are training a sequence of a fixed length of a short-range because this way this architecture can also be put to use to detect obscenity in between the videos also as it will be trained on classifying the moments of the video feed. This feature will be beneficial as we observe that even the pornographic video does not contain obscenity in every frame. Especially the starting frames of the videos are shown to have the logo of the company, no human, other irrelevant backgrounds with respect to that category et cetera.

C. Frame Filtering by detecting human in sequence

While training the model, it was observed that Network was also training on the sequence of frames which does not even contain a human in the frames. We assume that, in most cases, the maximum percentage of obscenity is there only when humans are present in the frames. To get better results and deal with this problem, we deployed a human detection model before passing frames sequence to feature-extraction network to detect the presence of humans.

As given in Eq. 1, we have set v of k frames extracted from a video. We pass this set of frames through TensorFlow human detection API, the passed set of frames is h,

$$h = \{h_1, h_2, h_3, \dots \dots \dots, h_k\} \quad (9)$$

where $h(i) \in \{0,1\}$, the condition to pass the sequence of N frames through the network is given below,

$$Average(h_{i\dots i+N}) = (h_i + h_2 + \cdots + h_{i+N})/N \quad (10)$$

When $Average(h_{i\dots i+N}) > \beta$, where $\beta = 0.6$ as used in this work, then we pass that sequence of frames through the ResNet-18. β is one of the hyperparameters that is required to change the threshold of the ratio of human consisting frames to the frame sequence length. We have kept $\beta = 0.6$ in the stringent setting while training the network so that network learns the exact cases of pornography where human is present.

We exploited the TensorFlow Objection Detection API [16] for detecting humans. In particular, we used the pre-trained weights of the Faster RCNN-Inception ResNet V2 model for human detection as architecture achieves fairly high accuracy and Top-1 accuracy of 80.4%. The only drawback of using this architecture was that it has a comparatively lower speed in frames processing as compared to other latest architectures. We filtered the frames only during training while passing the data for the pornographic category as it is only natural to assume that for the non-pornographic category human presence or absence does not matter. Fig. 1 shows the working of complete Frame Sequence Classifier ConvNet, including every important element intuitively.

IV. Experimental Methodology

In order to evaluate the performance of our proposed method, we have conducted an experiment on two standard baseline dataset i.e Pornography-800 and Pornography-2k.

All our experiments were performed on the system provided in Biometric Research Laboratory, Delhi Technological University, which contains Intel Xeon Silver 4114 processor, 128 GB RAM and Nvidia Titan RTX graphics card having 24 GB Memory.

The hyperparameter settings for the feature extraction network are as follows: we fine-tuned the ResNet-18 for both Dataset on the frames using $learning\ rate = 5 \times 10^{-3}$ with $Adam$ optimizer and $binary\ cross\ entropy$ loss.

For Sequence Classifier ConvNet we kept the $learning\ rate = 10^{-3}$ with $Adam$ optimizer and $binary\ cross\ entropy$ loss.



Fig. 3 Sample frames of Pornography -2k Dataset, shows the diversity and difficulty in videos of dataset.

2379-8920 (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 March 23,2022 at 07:10:22 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/TCDS.2022.3158613, IEEE Transactions on Cognitive and Developmental Systems

A. *Pornography-800 Dataset*

This dataset was developed by Avila et al. [7] on which all the previous algorithms have been tested. Pornography-800 dataset contains videos of two categories i.e pornographic and non-pornographic, where non-pornographic is bifurcated into two categories, i.e. easy, difficult. It consists of a video of around 77 hours total, out of which 57 hours video consists of pornography video and other time consists of the non-pornographic video. This dataset proposes a challenging problem of classifying videos because the videos contain people of a wide range of skin color i.e all sorts of skin color ranging from white to brown to dark skin complexion and people of different ethnicity i.e Asian, white, multi-ethnic. Other challenges provided by this dataset lies in the difficult non-pornographic category. In that category, certain videos containing activities such as sumo-wrestling with high skin exposure and contains two people that provides a challenging case to classify as the model may consider it to be an adult clip if we do not classify this as a sequence or consider to include motion stream directly as done by Perez et al [3].

B. *Pornography -2k Dataset*

This dataset was developed by Avila et al. [7], which is an extension of the Pornography-800 dataset and has a larger size as compared to the Pornography-800 Dataset. It consists of 140 hours of 1000 Pornographic and 1000 Non-Pornographic videos. The length of videos varies from 6 seconds to 33 minutes. This dataset is more challenging since it represents the variety of photo styles such as number of genres, cartoon from live actions, diverse behavior and ethnicity. The performance measure used to evaluate this dataset was the 5×2-fold cross-validation method in the previous works. The sample images of this dataset are as shown in Fig. 3. These images present in Fig.3 shows the diversity that is present in the dataset. The figure contains three rows where first represents the images from pornographic videos whereas the other two rows belong to the non-pornographic category. The middle row consists of frames of the difficult cases of non-pornographic videos.

C. *Evaluation Measures*

The output of the sequence classifier network is a probability for each class as we have used the Softmax activation function at the output layer. Hence, the final prediction is calculated using a threshold of 50%, which is arbitrary. We choose to evaluate the algorithm using the ROC (Receiver Operating Characteristic) curve for comparing the results of various approaches that we have used. The second evaluation parameter that we have used is accuracy i.e fraction of the sum of true positive and true negative divided by total videos.

For evaluating the Pornography-2000 Dataset, we use the $5 \times 2\ fold$ cross-validation technique to check the performance of the proposed model on the Dataset.

D. *Feature Extractor Network*

We utilize the pre-trained ResNet-18 Network in PyTorch based on the work of Zhang et al [15] for feature extraction. The reason to choose ResNet-18 over other ResNet architectures is that deeper networks are more difficult to train and uses more space in RAM but we were looking for a simpler network to gain the maximum performance. Therefore, we chose ResNet-18 which has $11.8\ Million$ weight parameters. The second reason for choosing ResNet-18 was that its output vector encoding the frame's information in R^{512} vector as discussed in Section. III (A).

We fine-tuned the ResNet-18 Network on randomly selected frames from the extracted frames. The selected frames were key-frames that were selected randomly from videos, selecting mostly the frames present in the middle of the video as frames in the starting contains logo like information for the videos. This choice was based on the analysis and observation of the data that is usually present in pornography class.

E. *Sequence Classifier Conv. Network*

After the feature extraction from the ResNet network, we process the feature map of the $R^{N\times 512\times 1}$ formed by the N consecutive feature vector from extracted frames. We experimented with two values of N i.e 8 and 16 continuous frames where we achieved the best results whereas when we chose $N = 24$, the performance started degrading. The reason for the performance to start decreasing is because of longer sequence dependency in pornographic videos can lead the model to misconstrue the motion information because of frequent changes in the angle of the camera in the videos which is the characteristic feature present in the pornographic videos. We use different CNN architecture for both sequence lengths. Both the networks are shown in Fig. 2 along with their hyperparameters detail. For making a prediction on the whole video, we make the sequences of N consecutive frames and pass their feature map through Sequence Classifier ConvNet which output the probability of that feature-map belonging to each class i.e pornography and non-pornography and the average is taken over all feature maps of the video. This average calculated for each class represents the probabilities of video belonging to each class.

ConvNet to classify the sequence of 8 frames is given in Fig. 2(b). Parameters in this network are $\approx 4.2\ Million$, which amount to a total of 16 $Million$ parameters for the end-to-end network having $\approx$11.8 $Million$ parameters for ResNet-18 and 4.2 for Sequence Classifier ConvNet. For regularization in this ConvNet, we use $Dropout$ after each $Conv2D$ layer. Conv2D learn its filters using propagation [17] after each epoch. ReLU (Rectified Linear Unit) activation function introduced by Nwankpa et al [18] is used as an activation function in this Network.

However, we also experimented with $LeakyReLU$ activation function having $a = 0.2$, we observed that Network was able to converge in almost half epochs as it took while training for $ReLU$ activation.

2379-8920 (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 March 23,2022 at 07:10:22 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/TCDS.2022.3158613, IEEE Transactions on Cognitive and Developmental Systems

While processing the sequence of 16 frames, we designed the ConvNet given in Fig. 2(a). The parameters in this ConvNet is approximately $2\,Million$, therefore, bringing the total parameters of end-to-end Sequence Classifier ConvNet $\approx 13.8\,Million$.

The output of the sequence classifier ConvNet is the probability representing the confidence with which the sequence belongs to each category. While we make a prediction for the whole video, we divide the video into sequence clips on which predictions are made, in the end, the majority vote of the predictions for clips is taken to predict the category that a video belongs to according to the model. As the model is trained on the filtered data, we can use that model to predict obscenity in between the video's small sections as we discussed before the whole video does not contain obscenity and that may lead to false prediction i.e. false positive.

F. Performance Assessment on Pornography-800 Dataset

The performance of the network is assessed on this dataset in terms of the confusion matrix, training and validation accuracy, and Receiver Operating Characteristics (ROCs). The classification result is obtained on 8 and 16 frames sequence with human and without human, as shown in Fig. 4 and Fig. 5 with the help of confusion matrix.

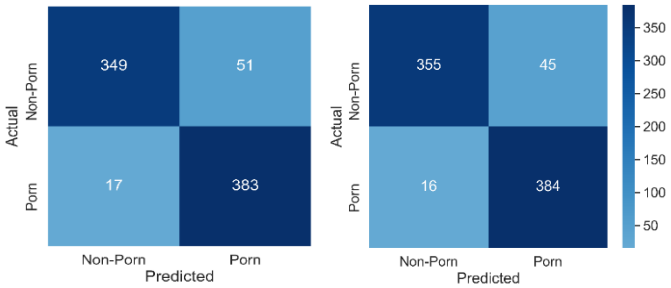


Fig.4 Classification Result of Pornography-800 dataset of sequence length a) 8 and b) 16 without human detection

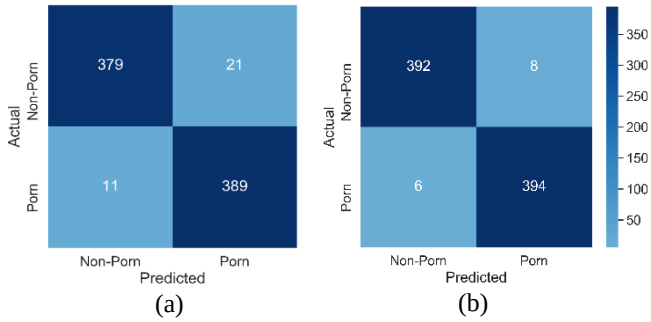


Fig. 5 Classification Result of Pornography-800 dataset of sequence length a) 8 and b) 16 with human detection

In Table.1, we observe that our proposed Pornographic Classification Architecture achieves state-of-the-art results on pornographic-800 dataset, surpassing the accuracy achieved by Perez et al [3] by 0.35% which was the state-of-the-art method.

We can observe that our proposed Pornography Classification Architecture with sequence length 16 and having human filtering performs better than every other method such as a bag of words-based methods [19][20][7][21] and deep learning-based approaches [10][13][5]. We are able to surpass their accuracy even though we have lesser weight parameters in our model consisting of total 16 $Million$ weight parameters. In the work done by Silva et al [13] they have used VGG-C3D that has around 179.1 $Million$ weight parameters. Therefore, it puts our approach to an advantage that smaller architectures can also be used to achieve high accuracy with techniques employed such as human filtering in pornography detection. While we see that our Sequence Classifier ConvNet for sequence length of 8 (with Human Detection Filtering) does not achieve as state-of-the-art accuracy but this approach is also able to achieve better accuracy than other deep learning-based approach used in the previous works.

TABLE 1. COMPARISON OF RESULTS ON PORNOGRAPHY-800 DATASET

Method	Accuracy (%)	AUC
BOSSA [20]	87.1±2	---
BOSSANova(HueSIFT) [7]	89.5±1	---
BOSSANova VD (Bin Boost 16) [21]	90.9±1	---
BNVD(Bin Boost d= 16) [19]	92±1	0.973
BOW-VD(Bin Boost d = 16) [19]	92.4±1	0.976
AGNet [10]	93.8±2	---
AGbNet [10]	94.1±2	---
Spatio-Temporal C3D [13]	95.1±1.7	---
ACORDE-101 [5]	95.6±1	0.99
Perez et al [3]	97.9±0.7	---
E2E VCN (8 Frame Sequence without human filtering)	**91.5±1**	**0.975**
E2E VCN (8 Frame sequence with human filtering)	**96±1**	**0.985**
E2E VCN (16 Frame sequence without human filtering)	**92.37±1**	**0.973**
E2E VCN (16 Frame sequence with human filtering)	**98.25±1**	**0.985**

The training-validation accuracy plots are shown in Fig. 6-Fig. 9 that show an increasing trend that infers that the model is training properly. We stop training the model when the validation accuracy starts decreasing for a certain number of epochs. The training accuracy given in these plots is different from the accuracy given in Table 1. While training, we pass randomly selected K number of feature-map from every video and do not pass the whole video through the network (i.e training on random fragments of the video) and we also calculate the accuracy and loss on these. At the test time, while we provide the accuracy for video classification, we take prediction on every feature map of each video and take the average of the probabilities coming from the $softmax$ function over each class for every feature map of the video. It gives the probability of video belonging to each class and we select the class having higher probability.

The performance of all proposed models can also be seen from Fig. 10 where AUC-ROC curve is given which gives $AUC = 0.985$ for the best performing model. Also, the performance of model in detecting specific class of the video

2379-8920 (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 March 23,2022 at 07:10:22 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/TCDS.2022.3158613, IEEE Transactions on Cognitive and Developmental Systems

can also be seen in Fig. 11.

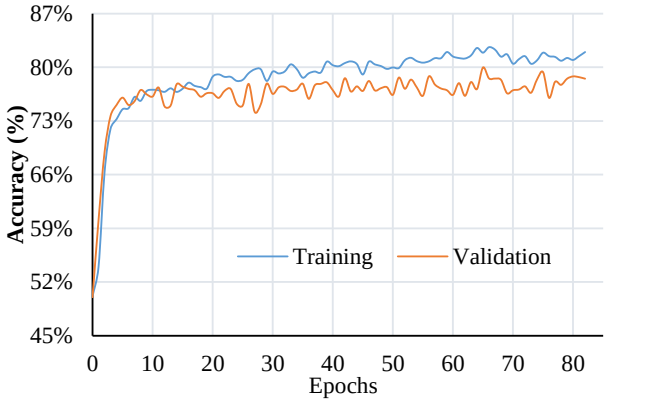


Fig. 6. Training Plot Sequence Length 8 (WHD) for Pornography 800 Dataset

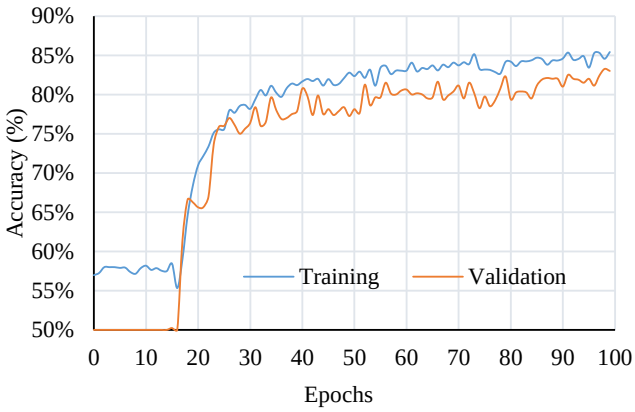


Fig. 7. Training Plot for Sequence Length 8 (HD) for Pornography 800 Dataset

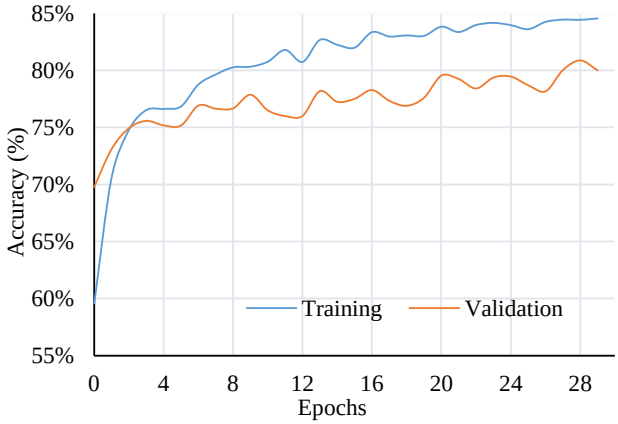


Fig. 8. Training Plot For Sequence Length 16 (WHD) for Pornography-800 Dataset

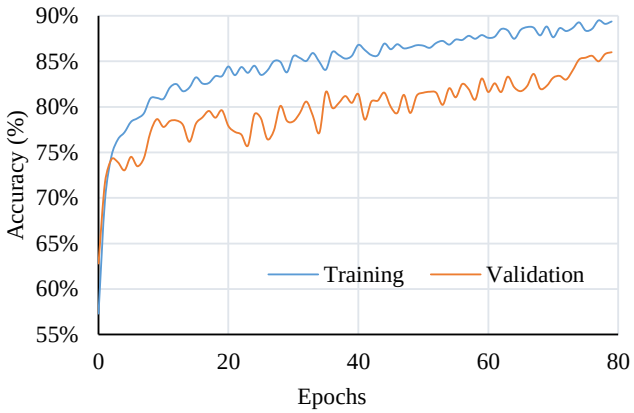


Fig. 9. Training Plot Sequence Length 16 (HD) for Pornography-800 Dataset

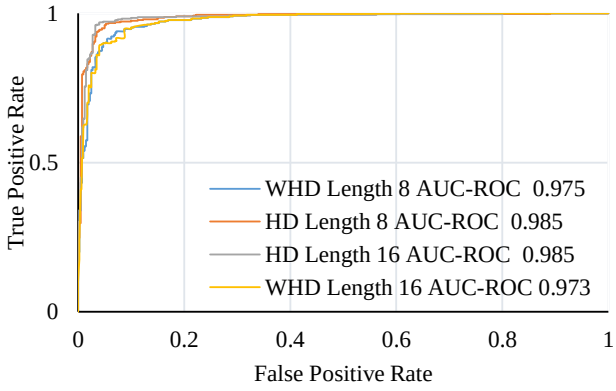


Fig.10. AUC-ROC curve for all proposed model HD-Human Detection, WHD-Without Human Detection on Pornography-800 Dataset

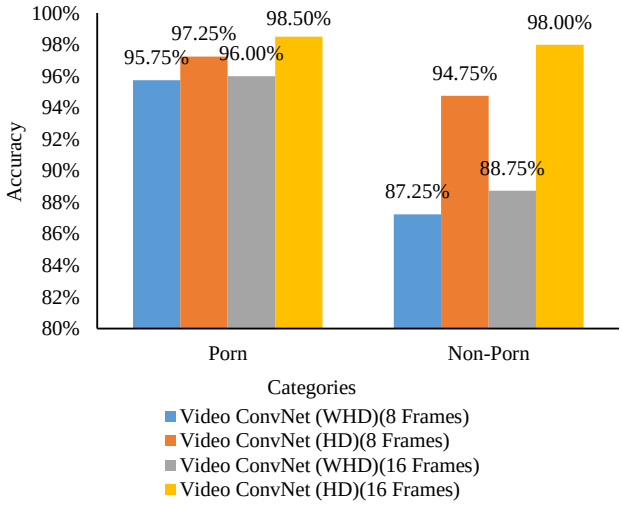


Fig. 11. Comparison Plot of Sequence Classifier ConvNet (HD-Human Detection) (WHD-Without Human Detection)-Pornography 800 Dataset

2379-8920 (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 March 23,2022 at 07:10:22 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/TCDS.2022.3158613, IEEE Transactions on Cognitive and Developmental Systems

G. *Performance Assessment on Pornography-2K Dataset*

The results on pornographic detection of the network on Pornography-2k dataset is as shown in Fig. 12 with different sequence lengths i.e 8 and 16 without human filtering in the form of the confusion matrix. Upon experimentation on the Pornography-2K dataset, we observe that model can achieve better accuracy than other works even without human filtering because of the presence of sufficient and diverse data as juxtaposed with the performance on Pornography-800. The performance of the proposed methods can be observed using confusion matrix, ROC Curves, training plots and accuracy achieved as shown in Fig 12, 13, 14 and 15. The overall highest accuracy reported on this dataset is 97.15% with 8 sequence length. The good performance of the network is also observed from the ROC curve shown in Fig. 12, where AUC is 0.994.

The training and validation accuracy of the network can be seen in Fig. 13 and Fig.14 in the case of 8 and 16 sequence lengths respectively, where it can be clearly visualized that the network is training properly and an upward trend can be seen in the training and validation accuracy altogether.

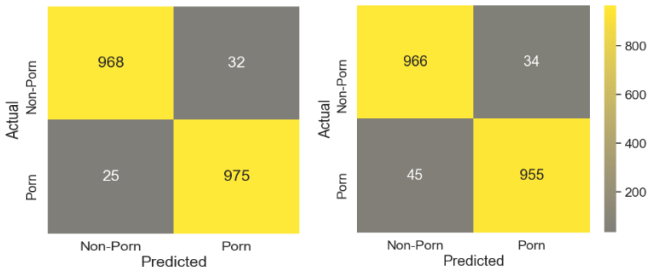


Fig. 12. Classification Result of Pornography-2K dataset of sequence length a) 8 and b) 16 without human detection

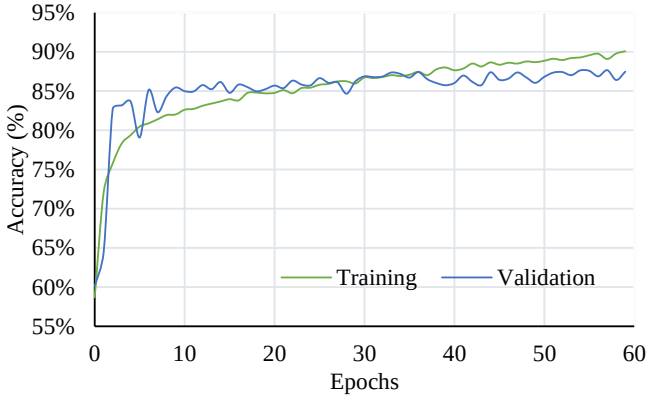


Fig. 13 Training Plot of Pornography-2K Dataset (8 sequence length) without human detection

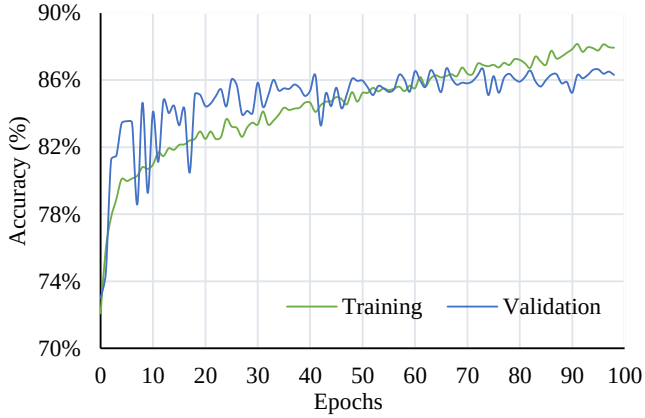


Fig. 14. Training Plot of Pornography-2K Dataset (16 sequence length) without human detection

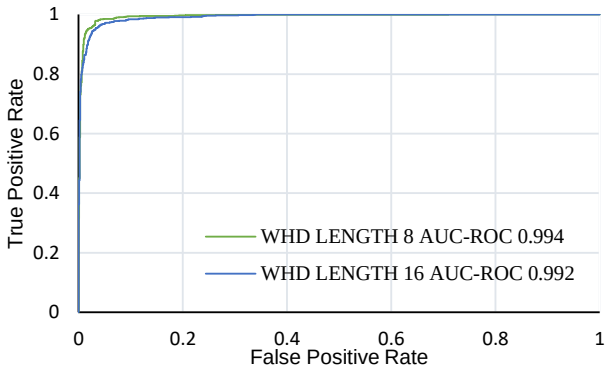


Fig. 15. AUC-ROC curve for all proposed WHD-Without Human Detection Model on Pornography-2k Dataset

In Table 2, we see that our proposed model has achieved state-of-the-art results by achieving 97.15% accuracy using this model on Pornography-2k Dataset without using human detection filtering. Pornography-2k is a dataset of larger size as compared to Pornography-800 dataset and also contain fairly cleaned videos containing lesser redundant frame sequence which also plays an important part in achieving an even higher accuracy using our proposed method and without even using Human Detection Filtering Technique.

While looking at the training curve given in Fig. 13 and 14, we observe that the model achieves the training accuracy of approximately 90% when $N = 8$ and 88% when $N = 16$. This accuracy is different from the accuracy on videos, it is explained as follows, when we train the network, we randomly pick up $K(16\ is\ used\ by\ us)$ sequences from all the videos and calculate the accuracy and loss over that as given in Fig. 13 and Fig. 14. When we check the accuracy over the whole video by predicting and taking the average over all the sequences, we achieve the accuracy as given in Table 2.

The fluctuations in the training plots in Fig. 13 and 14 can be accounted for because of the random pick up of K feature-maps from a video.

2379-8920 (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 March 23,2022 at 07:10:22 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/TCDS.2022.3158613, IEEE Transactions on Cognitive and Developmental Systems

TABLE 2. COMPARISON OF RESULTS ON PORNOGRAPHY -2000 DATASET

Method	Accuracy (%)	AUC
TRoF [8]	95.6	---
Late Fusion (Motion Vector) [3]	96.4	---
Late Fusion (Optical Flow) [3]	96.4	---
E2E VCN (8 Frame Sequence without human filtering)	**97.15±1**	**0.994**
E2E VCN (16 Frame sequence without human filtering)	**96±1**	**0.992**

V. CONCLUSION

Our proposed method has achieved a state-of-the-art accuracy of 98.25% on evaluating our model accuracy on Pornography-800 Dataset and state-of-the-art accuracy of 97.15% on Pornography-2000 Dataset using Pornographic Classification Architecture of 16 sequence length with human detection and 8 sequence length without human detection respectively.

The results of our proposed architectures achieve better accuracy than all the deep learning architectures in the case of both datasets. We are able to achieve better accuracy than the best approach by Perez et al [3], our proposed technique outperformed all other deep learning architectures proposed in previous works by Wehrmann et al [5], Moustafa et al [10] and also the motion-optical flow approach by Perez et al [3]. This approach also has another benefit of classifying sections in between the videos, which contain pornographic scenes because our proposed network is the end-to-end network, so it is easy to reduce computation using that.

Thinking about how future improvements can be made, we emphasize that Capsule Network architectures should be explored further for this classification problem. These types of networks have not been explored enough for this particular application. Capsule Networks are better at capturing the geometrical intricate details better than the CNNs so that could prove to be a promising explorable area to advance the pornography detection field.

In future, one can use this approach in real-time for developing a firewall, which can block pornographic content over the web.

VI. REFERENCES

[1] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: A compact facial video forgery detection network," in *10th IEEE International Workshop on Information Forensics and Security, WIFS 2018*, 2019, pp. 1–7, doi: 10.1109/WIFS.2018.8630761.

[2] A. Quadra, A. El-Murr, and J. Latham, *The effects of pornography on children and young people*, vol. 1, no. 1. 2017.

[3] M. Perez *et al.*, "Video pornography detection through deep learning techniques and motion information," *Neurocomputing*, vol. 230, pp. 279–293, 2017, doi: 10.1016/j.neucom.2016.12.017.

[4] K. Simoyan and A. Zisserman, "Simonyan, Karen & Zisserman, Andrew. (2014). Two-Stream Convolutional Networks for Action Recognition in Videos. Advances in Neural Information Processing Systems. 1.," *Nips*, vol. 32, no. 5, pp. 849–855, 2014, doi: 10.1016/0006-2952(83)90587-7.

[5] J. Wehrmann, G. S. Simões, R. C. Barros, and V. F. Cavalcante, "Adult content detection in videos with convolutional and recurrent neural networks," *Neurocomputing*, vol. 272, pp. 432–438, 2018, doi: 10.1016/j.neucom.2017.07.012.

[6] "{TensorFlow}: Large-Scale Machine Learning on Heterogeneous Systems," 2015. https://www.tensorflow.org/.

[7] S. Avila, N. Thome, M. Cord, E. Valle, and A. De A. Araújo, "Pooling in image representation: The visual codeword point of view," *Comput. Vis. Image Underst.*, vol. 117, no. 5, pp. 453–465, 2013, doi: 10.1016/j.cviu.2012.09.007.

[8] D. Moreira *et al.*, "Pornography classification: The hidden clues in video space–time," *Forensic Sci. Int.*, vol. 268, pp. 46–61, 2016, doi: 10.1016/j.forsciint.2016.09.010.

[9] M. J. Jones, M. J. Jones, and J. M. Rehg, "Statistical Color Models with Application to Skin Detection Statistical Color Models with Application to Skin Detection," *Comput. Vis. Pattern Recognit.*, vol. 46, no. January, pp. 1–23, 2016, doi: 10.1023/A.

[10] M. Moustafa, "Applying deep learning to classify pornographic images and videos," 2015.

[11] A. Krizhevsky, I. Sutskever, and G. Hinton, "Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.," *Nips*, 2012, doi: 10.1201/9781420010749.

[12] G. Zeng, Y. He, Z. Yu, X. Yang, R. Yang, and L. Zhang, "Szegedy, Christian & Liu, Wei & Jia, Yangqing & Sermanet, Pierre & Reed, Scott & Anguelov, Dragomir & Erhan, Dumitru & Vanhoucke, Vincent & Rabinovich, Andrew. (2014). Going Deeper with Convolutions.," *arxiv*, 2014, doi: 10.1002/jctb.4820.

[13] M. V. da Silva and A. N. Marana, "Spatiotemporal CNNs for pornography detection in videos," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11401 LNCS, pp. 547–555, 2019, doi: 10.1007/978-3-030-13469-3_64.

[14] M. B. Short, L. Black, A. H. Smith, C. T. Wetterneck, and D. E. Wells, "A review of internet pornography use research: Methodology and content from the past 10 years," *Cyberpsychology, Behav. Soc. Netw.*, vol. 15, no. 1, pp. 13–23, 2012, doi: 10.1089/cyber.2010.0477.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 2015.

[16] J. Huang *et al.*, "Speed/accuracy trade-offs for modern convolutional object detectors," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 3296–3305, 2017, doi: 10.1109/CVPR.2017.351.

[17] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation. In: Rumelhart D E, McClelland J L et al. (eds.) Parallel Distributed Processing: Explorations in the Microstructure of Cognition.," *MIT Press. Cambridge, MA*, vol. 1, no. V, pp. 318–362, 1986.

[18] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation Functions: Comparison of trends in Practice and Research for Deep Learning," pp. 1–20, 2018.

[19] C. Caetano, S. Avila, W. R. Schwartz, S. J. F. Guimarães, and A. de A. Araújo, "A mid-level video representation based on binary descriptors: A case study for pornography detection," *Neurocomputing*, vol. 213, pp. 102–114, 2016, doi: 10.1016/j.neucom.2016.03.099.

[20] S. Avila, N. Thome, M. Cord, E. Valle, and A. De, "BOSSA: Extended bow formalism for image classification," *Proc. - Int. Conf. Image Process. ICIP*, no. 1, pp. 2909–2912, 2011, doi: 10.1109/ICIP.2011.6116268.

[21] C. Caetano, S. Avila, S. Guimarães, and A. A. De Araújo, "Pornography detection using BossaNova video descriptor," *Eur. Signal Process. Conf.*, pp. 1681–1685, 2014.



Neil Gautam is the under graduate student of the Department of Information Technology, Delhi Technological University, Delhi, India. He is also a Research Associate at Biometric Research Laboratory. His research Interest Includes Computer Vision, Machine Learning, and Deep Learning.

2379-8920 (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 March 23,2022 at 07:10:22 UTC from IEEE Xplore. Restrictions apply.

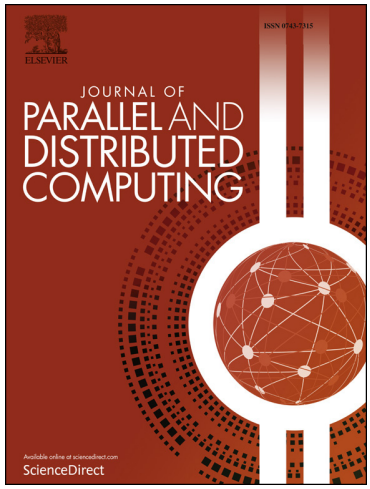
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCDS.2022.3158613, IEEE Transactions on Cognitive and Developmental Systems



Dinesh Kumar Vishwakarma (M'16, SM'19) received a Doctor of Philosophy (Ph.D.) degree in the field of Computer Vision and Machine Learning from Delhi Technological University, New Delhi, India, in 2016. He is currently a Professor with the Department of Information Technology, Delhi Technological University. His current research interests include Computer Vision, Deep/Machine Learning, Sentiment Analysis, Fake News Detection, Multimedia Data Analytics, and Crowd Behaviour Analysis. He received research excellence awards from the Delhi Technological University in the years 2017, 2018, 2019, 2020, and 2021. He is Associate Editor of IEEE Transactions on Circuits Systems for Video Technology. He is a reviewer of various journals/transactions of the ACM, IEEE, Elsevier, and Springer. He is a senior member of IEEE, Member of ACM and a lifetime member of ISTE

2379-8920 (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 March 23,2022 at 07:10:22 UTC from IEEE Xplore. Restrictions apply.

Journal Pre-proof



Oppositional chaos game optimization based clustering with trust based data transmission protocol for intelligent IoT edge systems

M. Padmaa, T. Jayasankar, S. Venkatraman, Ashit Kumar Dutta, Deepak Gupta et al.

PII: S0743-7315(22)00065-X
DOI: https://doi.org/10.1016/j.jpdc.2022.03.008
Reference: YJPDC 4531

To appear in: *Journal of Parallel and Distributed Computing*

Received date: 7 June 2021
Revised date: 27 January 2022
Accepted date: 7 March 2022

Please cite this article as: M. Padmaa, T. Jayasankar, S. Venkatraman et al., Oppositional chaos game optimization based clustering with trust based data transmission protocol for intelligent IoT edge systems, *Journal of Parallel and Distributed Computing*, doi: https://doi.org/10.1016/j.jpdc.2022.03.008.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier.

Journal Pre-proof

Highlights

- Propose an energy efficient EEC-SDTP technique for Intelligent IoT Edge systems.
- The EEC-SDTP technique aims to select CHs and secure routes for data transmission.
- Designs a novel OCGOC technique for proper CH selection and cluster construction.
- Derive a SRP-QSPO technique to achieve secure data transmission in IoT edge systems.
- Validate the performance of the EEC-SDTP technique interms of different measures.

Journal Pre-proof

Journal Pre-proof

Oppositional Chaos Game Optimization Based Clustering with Trust Based Data Transmission Protocol for Intelligent IoT Edge Systems

M.Padmaa[1], T. Jayasankar[2], S. Venkatraman[3], Ashit Kumar Dutta[4], Deepak Gupta[5,*], Shahab Shamshirband[6,*], Joel J. P. C. Rodrigues[7,8]

[1]Department of Electronics and Communication Engineering, Saranathan College of Engineering, Tiruchirapalli, India.

[2]Department of Electronics and Communication Engineering, University College of Engineering, BIT Campus, Anna University, Tiruchirappalli, India.

[3]School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India

[4]Department of Computer Science and Information System, College of Applied Sciences, AlMaarefa University, Riyadh, 11597, Kingdom of Saudi Arabia

[5]Maharaja Agarsen Institute of Technology, New Delhi, India

[6]Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Douliou 64002, Taiwan

[7] College of Computer Science and Technology, China University of Petroleum (East China), Qingdao 266555, China

[8]Instituto de Telecomunicações, Covilhã, Portugal

Email: padmaa-ece@saranathan.ac.in, jayasankar27681@gmail.com, venkats23@gmail.com, adotta@mcst.edu.sa, deepakgupta@mait.ac.in, shamshirbands@yuntech.edu.tw, joeljr@ieee.org

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Abstract

In the past decade, the Internet of Things (IoT) becomes essential in consumer and industrial applications. The accessibility of high bandwidth Internet connection particularly with the arrival of robust 5G networks rises to innovative IoT solutions such as smart city, automobiles, industry 4.0, etc. IoT analytics represents edge computing as a term commonly employed for defining intelligent computational resources placed closer to the source of data generation. Despite the benefits of the IoT edge systems, security and energy efficiency remains major challenging issues. With this motivation, this paper presents an energy-efficient clustering based secure data transmission protocol (EEC-SDTP) for Intelligent IoT Edge systems. The goal of the EEC-SDTP technique is to select an appropriate set of cluster heads (CHs) and optimally secure routes for data transmission in the network. The proposed model involves oppositional chaos game optimization-based clustering (OCGOC) technique for proper CH selection and cluster construction. Besides, a trust-based model is designed to determine the trustworthiness of the node in the IoT edge systems. The proposed OCGOC technique derives

Journal Pre-proof

a fitness function utilizing 3 input parameters like trust level, distance to neighbors, and energy. Finally, a trust-based secure routing protocol using the quantum sand piper optimization (SRP-QSPO) technique is employed to derive routes for secure data transmission. For examining the better efficiency of the proposed EEC-SDTP algorithm, an extensive group of experimentations were performed and the outcomes are investigated under several performance measures. The experimental outcomes highlighted the improved efficiency of the proposed method over the other related techniques.

Keywords: Internet of Things, Edge Computing, Artificial intelligence, Security, Trust sensing model, Metaheuristics

1. Introduction

Currently, the Internet of Things (IoT) and mobile edge computing (MEC) become popular to fulfil future technological needs. The technology growth focuses on the need for data transmission, minimize the network load and throughput. The increasing need focused on managing minimum delay, data storage, and processing in major application regions like smart grid, smart city, sustainable platform, and transportation. A large number of data could be deal with MEC devices that have a particular framework for attaining network analytics, network optimization using IoT [1]. With the rapidly increasing amount of IoT devices, scientists on autonomous network creation for IoT devices have attained more interest. In the perception of network implementation, the primary objective is to create an energy efficient multi hop network which could link source to destination nodes through mobile relay nodes using energy constraint [2]. Therefore, the network topology has been defined according to wireless connections among relay node that are IoT device.

The feasible application includes military network [3] and disaster network whereas constrained networking framework is presented, and mobile agents like smart IoT devices are spread sparsely over a huge area should perform as relay node for delivering data in an ad hoc way. In this condition, mobility of the agent could not be managed centrally due to the constraints in networking framework that implies the network could not be implemented with the determination of the position of mobile relay nodes. There was a large surge in utilization of IoT in various application fields like weather predicting, agriculture, industrial automation, energy system, healthcare, etc [4, 5]. IoT hasn’t still understood its large significance because of various bottlenecks in its execution.

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof

Few on-going problems in the execution of IoT network contain energy utilization, security, privacy, data integration, and storing of the data produced using IoT devices. Since the IoT device works on battery (or constrained power supply), and it always gather and allocate the data there is large need for energy efficiency. The energy utilization in IoT device should be enhanced as battery in the device drains rapidly [6]. A common method is the grouping of IoT device to cluster and selects an optimum node as cluster head (CH) for this cluster [7]. Fig. 1 displays an IoT framework with CH. Primarily, the sensor nodes are placed in network. In every cluster, an optimum CH is selected depending upon various efficiency metrics, and each node in the cluster should transmit their packet to the resultant CH. The CH transmit data packet to IoT BS. Data gathered from the BS would be distributed on cloud server.

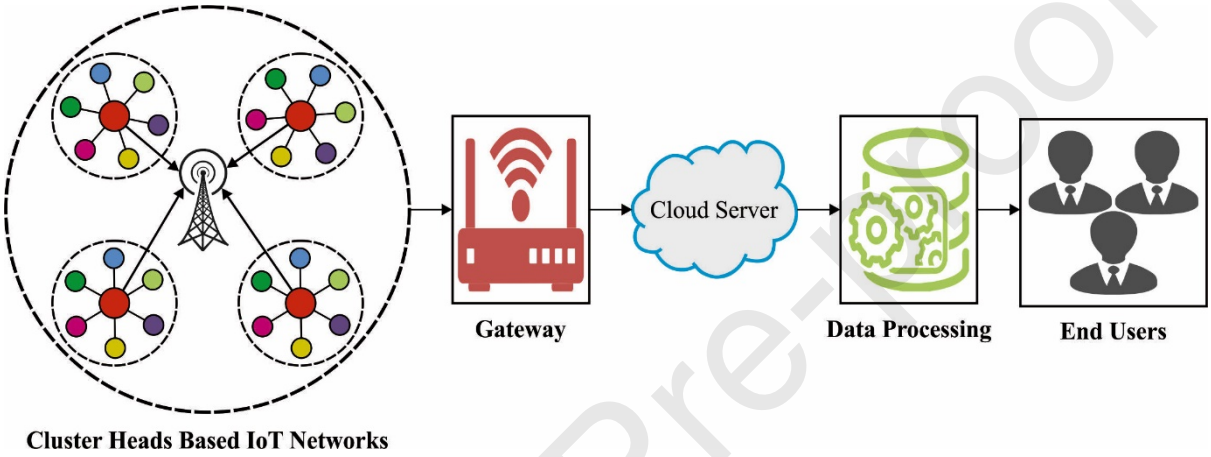


Fig. 1. Clustering architecture of IoT Nodes

In the present modules, sensor node determines function and transmission in distributed clusters. When the cluster size is distinct, e.g., lower/higher node cluster, the efficiency of the IoT network would decrease for few nodes the distance amongst BS & CH. The dynamic clustering approach suffers from higher computation time that might result in increased delay & load [7]. Because of this, various evolution metaheuristic techniques are presented for optimizing energy consumption. A few common metaheuristics and evolution techniques contain genetic algorithm (GA), cuckoo search algorithm, firefly algorithm, bat algorithm, and so on. This method focuses on nodes that are continuous and few nodes for sleeping to minimize utilization of energy however IoT networks result in utilizing higher energy because of various aspects such as temperature, load, and data traffic.

The routing protocol is vital in the transmission stack. Routing protocol creates the network to allow the set of data to pass by manifold hops from starting to targeted nodes. The features of

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof

WSN distinguish them from wireless & wireline networks. The haphazard feature of wireless connection sternly affects transmission feasibility and limited computing abilities and based on constrained battery. Therefore, routing is deliberated as a difficult obstacle [8]. Routing is deliberated significant as nodes that exist in the IoT network rely on hosts and routers whereas data distributing to gateways. Various routing protocols are developed in sensors taking into the account IoT network. The transmission of data between nodes affects power consumption when transmitting data.

Because of randomized behavior, stochastic methods are utilized for investigating the power consumption of every node and finish the network [9]. This approach utilizes historical events to predict the behavior of the future. Moreover, routing contains nodes that define route to destination which might suffer overhead. All over beaconing, source node provides a route request to its neighbors that retransmit the packet attained at the end. The destination responses to requests and a route are generated. Additionally, features like beacon interval affect the rate when beacons are transmitted. Thus, there is a requirement to enumerate the performance and energy effect of routing approach and examine similar overhead and controls.13 Scalable routing protocols are utilized in several conditions and network size variants can define optimum routes. Trust based systems are utilized in IoT to provide safety in routing. The IoT is utilized in several applications, such as business, smart homes, and battlefield. Henceforth, the security of routing is required that is enhanced with trust based routing.

In this view, this paper presents an energy-efficient clustering based secure data transmission protocol (EEC-SDTP) for Intelligent IoT Edge systems. The EEC-SDTP technique aims is to elect a proper collection of CHs and secure routes for data transmission in the network. The proposed model involves oppositional chaos game optimization-based clustering (OCGOC) technique and a trust-based secure routing protocol using the quantum sand piper optimization (SRP-QSPO) technique. Besides, a trust-based model is designed to determine the trustworthiness of the node in the IoT edge systems. The proposed OCGOC technique derives a fitness function utilizing 3 input parameters like trust level, distance to neighbors, and energy. For examining the improved efficiency of the proposed EEC-SDTP technique, a detailed set of simulation analyses is made and the results are inspected in terms of distinct evaluation parameters.

In short, the contribution of the paper is summarized as follows.

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof

- Propose a new EEC-SDTP technique to achieve energy efficiency and security in Intelligent IoT Edge systems.
- Designs a new OCGOC technique by incorporating oppositional based learning (OBL) concept to the CGOC technique for proper CH selection and cluster construction.
- Derives a fitness function utilizing three input parameters such as trust level, distance to neighbors, and energy.
- Present a trust-based model to compute the trustworthiness of the node in the IoT edge systems.
- Develop a new SRP-QSPO algorithm for optimal selection of routes for secure data transmission.
- Validate the performance of the proposed EEC-SDTP technique under varying number of nodes and execution rounds.

The rest of the paper is arranged as follows. Section 2 offers a brief review of existing security and energy efficient solutions for intelligent IoT edge systems. Section 3 introduces the proposed model and section 4 validates the performance of the proposed model. At last, section 5 concludes the work.

2. Related works

Ilyas et al. [10] intended to extend network lifespan, enhance throughput, reduce packet loss or latency, and in addition improvise in facing malicious nodes. For tuning the network lifespan regarding wireless harvesting energy is recommended projected 3 layer cluster based WSN routing protocols. The projected technique is a 3 tier clustering method with embedded security technique for encountering malicious activity of sensor node and to slant them to blacklist. It has been centred based clustering protocol, whereas grid head and CH selection are executed using sink node depending upon value of thier cost function. Furthermore, hardware based link quality estimator is utilized for checking the connection efficiency and enhance routing performance. Santos et al. [11] present detection of the sinkhole and selective forward to Support secure routing for IoT, a novel IDS toward sinkhole, and selective forward attacks are target routing techniques in huge and mobile IoT network. For handling mobility and density problems in the recognition of hackers and ensure consistency, THATACHI utilizes trust, watchdog, and reputation approaches.

In Fang et al. [12], a trust management based and energy efficient hierarchical routing protocol, i.e., LEACH-TM, is projected. In LEACH-TM, with density of neighbor nodes, amount of

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof

dynamic decision CH nodes and residual energy, the size of the cluster could be constrained better to enhance energy efficiency, and evade unnecessary energy consumption of a node. Markiewicz and Sgandurra [13] examine the application of unsupervised learning methods, especially clustering, to improve a new IDS for network consists of lower power devices. They define resolution, named Clust-IT (Clustering of IoT), for managing heterogeneous data gathered from distributed and cooperative networks of connected devices and search this information for compromising indicators when residual protocol agnostic. In Awan et al. [14], a trust based clustering technique permits clusters to define trustworthy CH. A new feature combined in the presented technique contains trust based CH selection which includes experience, reputation, and knowledge of a node. The great importance of utilizing trust in clustering is the detection of compromised and malicious nodes. The detection of this node aids to remove the threat of invalid data.

Alshehri and Hussain [15] detects on-off attacks by the presented FL based method, and identify contradictory performance attacks and other malicious node. Next, they improve FL based method for detecting malicious nodes included in bad service provision. Lastly, to preserve the security of IoT network, they developed a safe messaging scheme that allows safe interaction among nodes. In Wang et al. [16], a novel module is presented for collecting trustworthy data based on edge computing in the IoT. Also, through mapping the trust value of node on force to mobile data collectors, the optimum mobility path is created with higher trust. Kalkan [17] proposed a SUTSEC that gives benign CH for the groups with the consideration of trust, power, mobility, and priority. SDN inherited features are leveraged to provide a dynamic secure selection. In addition, secure key distribution is deliberated in trust clustering. In this security operation, they deliberated various problems interrelated to QoE and QoS like reliable communication, energy efficiency, user preferences awareness, and lower latency.

In Sadrishojaei et al. [18], a CLRP MMS technique for Mobile IoT is proposed. In recent times, mobile sinks are utilized in routing additional energy saving and durability in WSN. In this study, the whole nodes are separated as to clusters, and later every cluster chose a CH by estimating the CHCF. If clustering operates on network with movable node, the likelihood of disconnecting the nodes from CH nodes would affect several data loss. It would alter the number of energy and rate of data attained, however, the number of unused energy is decreased by forecasting the position and reduce the sink & CH node distances.

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof

Malchi et al. [19] presented an ESF TNN method that allows three techniques. Initially, EPS module tackles the coverage fraction of every IoT sensor. Next, Sugeno Fuzzy Processing module controls the energy utilization by respectively allocating data to nodes with no defuzzification procedure. At last, Trust based Neural Data Storage technique improves a sufficient data storage capability with the consideration of average classification ratio when processing recreated data packets for concerning every interaction information's through Trust method. In Alqahtani et al. [20], a TBM system was implemented to enhance security features in cloud supported IoT platform. This security system utilizes intelligent agents and middleware to manage communication and user level security. TBM runs in 3 security administering stages i.e. trust construction, message authentication, and spoof detection. The intelligent agents are in charge of guaranteeing secure transmission by swapping trust and signal strength observation using middleware.

Earlier researchers have started to move towards mobile communication and edge computing. Billions of devices have been linked to the Internet and generated zettabytes of data items. Despite the benefits of the IoT edge systems, security and energy efficiency remains major challenging issues. The use of Artificial Intelligence, machine learning, neural networks, and data analytic techniques in edge processing resulted in a new interdisciplinary field that enables distributed intelligence with edge devices and is known as distributed edge AI or edge intelligence. Therefore, this work presents a new EEC-SDTP technique to achieve energy efficiency and security in Intelligent IoT Edge systems.

3. The Proposed Model

The overall working principle involved in the proposed model is given in Fig. 2. The proposed EEC-SDTP model involves two stage processes namely OCGOC based Clustering and SRP-QSPO based routing. Initially, the OCGOC technique is derived by incorporating the concept of oppositional based learning (OBL) into the conventional CGO algorithm to improve the convergence rate. Besides, the SRP-QSPO algorithm also involves the concept of quantum computing to the SPO algorithm.

3.1. Energy Efficient Clustering using OCGOC Technique

At this stage, the OCGOC technique is designed to select CHs and construct clusters in an effective way. The principal idea of the CGO technique is depending upon few principles of chaos concept where the configuration of fractals using chaos game theory and fractals self

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof

similarity problems are in perception [21]. In arithmetic, the word chaos game originally refers to a technique of making a fractal, with polygon and an earlier stage is chosen at arbitrary inside it. The fractal is made through iteratively making a series of points, beginning with early arbitrary stage, where every stage in the series is a provided fraction of the distance among the prior stage and other vertices of the polygon; the vertex is selected at arbitrary in every iteration. Repeating this iterative procedure a huge amount of times, choosing the vertex at arbitrary on every iteration, and removing the initial stages in the series, would frequently generate a fractal shape. By a standard triangle and the factors would lead to Sierpinski triangle, when making the appropriate plan with 4 stages and a factor would make a presentation of a "Sierpinski Tetrahedron", the 3D analog of Sierpinski triangle.

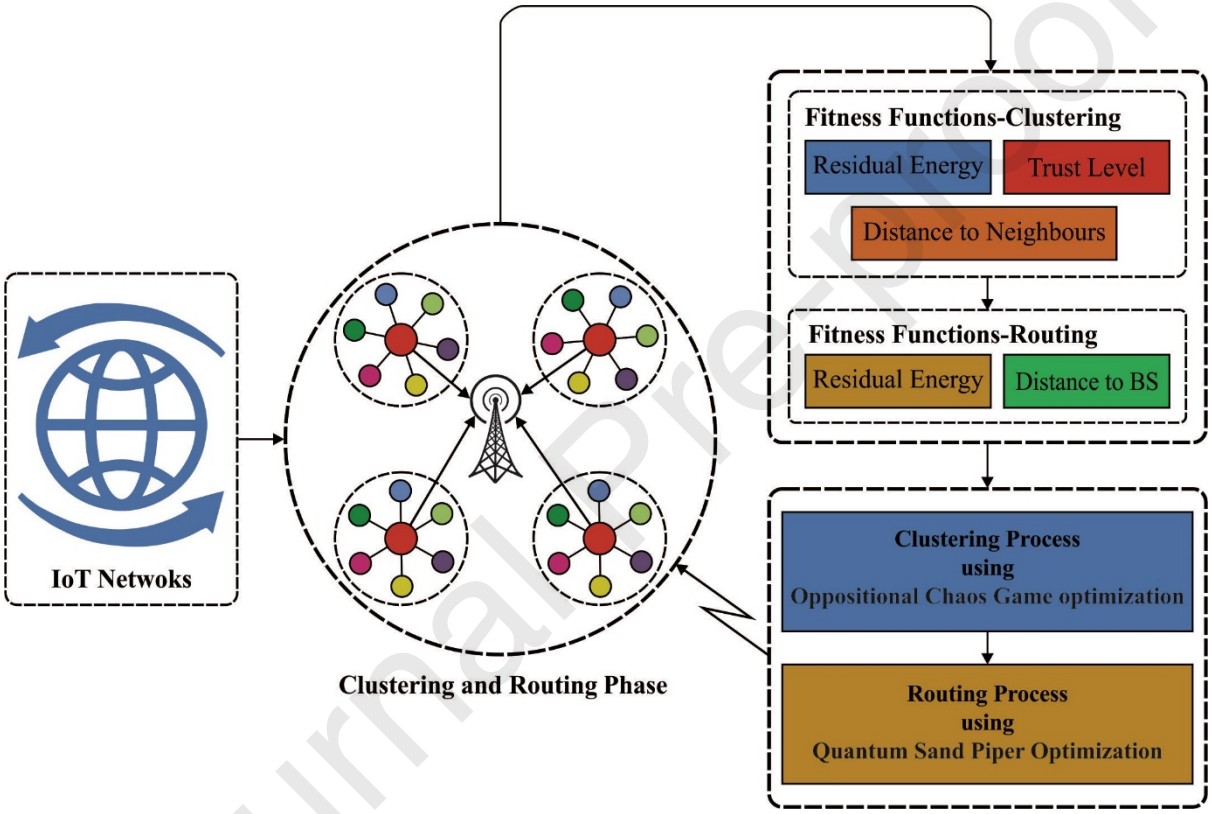


Fig. 2. Working process of EEC-SDTP model

Using "chaos game" a novel fractal could be created and when creating a novel fractal few variables could be attained. This variable is beneficial for the application of fractal theories like identification and classification. The novel fractal is self-similar to the original in few significant features like fractal dimension. When "chaos game" initiates at every vertex and endure every probable path that the game could take, they would attain a similar image as arbitrary path. But, taking several paths is hardly made as the overhead to keep track of all

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof

paths makes it farther slow to estimate [21]. This technique has the benefits of demonstrating how the fractal is made accurately compared to regular approach and being determined.

An optimization technique is projected according to the introduced principles of the chaos concept. The main ideas of the chaos game and fractals are used for formulating scientific modules to CGO technique. Due to several natural evolution methods maintaining a population of solution is developed using arbitrary selection and alterations. The CGO technique considers a number of solution candidates (S) in this objective that denotes few qualified seeds inside a Sierpinski triangle. The scientific demonstration of these factors is given by:

$$S = \begin{matrix} S_1 \\ \vdots \\ S_n \end{matrix} = \begin{bmatrix} S_1^1 & S_1^2 & S_1^j & \cdots & S_1^d \\ S_2^1 & S_2^2 & S_2^j & & \\ & \vdots & & \ddots & \vdots \\ S_i^1 & S_i^2 & S_i^j & \cdots & S_n^d \\ S_n^1 & S_n^2 & S_n^j & & \end{bmatrix} \quad (1)$$

$i = 1,2, \dots, n.\ j = 1,2, \dots, d$ whereas n denotes amount of eligible seed (solution candidate) inside the Sierpinski triangle (search space), and d indicates dimension of this seed. The early position of this eligible seed is defined arbitrarily in the search space by:

$$S_1^j(0) = S_{1,min}^j + R\left(S_{1,min}^j - S_{1,max}^j\right) \quad (2)$$

Where R arbitrary amount in the interval of zero and one. The procedure for the initial seeds are denoted by

$$Seed_i^1 = S_i + x_i * (y_i * Global\ best - z_i * Mean\ value) \quad (3)$$

Where x_i, y_i, z_i represents arbitrary integer of 0/1 to model the likelihood of rolling a dice. Later the schematic demonstration of the defined procedure for the subsequent seed is equated by,

$$Seed_i^2 = Global\ best + x_i * (y_i * S_i - z_i * Mean\ value) \quad (4)$$

A schematic demonstration of 3rd and 4th seeds is defined by,

$$Seed_i^3 = Mean\ value + x_i * (y_i * S_i - z_i * Global\ best) \quad (5)$$

$$Seed_i^4 = S_i\left(S_i^k = S_i^k + Rand\right) \quad (6)$$

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Where k indicates arbitrary integer in the interval of zero and one. The CGO technique, distinct forms are introduced for x_i that controls the motion limits of the seeds.

$$x_i = \begin{cases} 2 * rand \\ (\Psi * rand) + 1 \\ (\Omega * rand) + \sim\Omega \end{cases} \quad (7)$$

Where Rand denotes uniform distribution of arbitrary amount in the interval of zero and one, whereas Ψ and Ω denote arbitrary integer in the interval of zero and one.

The OCGO technique is derived from the OBL with classical CGO technique. It has been initially presented by Tizhoosh [22]. The OBL approach creates corresponding opposite solution for every first candidate solutions. From these 2 types of solutions (matching opposite solution and candidate solutions), with moderately optimum fitness are chosen as members of the first population. This would assist to enhance the convergence rate in the optimized procedure. The interrelated numerical concept of the OBL approach is given by

Where $x \in R$ denotes real number determined on specific interval: $x \in [a, b]$. The opposite number $\tilde{x}$ can be given by:

$$x = a + b - x \quad (8)$$

Where $P(x_1, x_2, \dots, x_n)$ denotes point in n dimensional coordinate scheme using $x_1, x_2, \dots, x_n \in R$ and $x_i \in [a_i, b_i]$. The opposite point P^* is entirely determined as coordinates $x_1^*, x_2^*, \dots, x_n^*$ whereas

$$x_i^* = a_i + b_i - x_i \; i = 1, \dots, n \quad (9)$$

Here $f(x)$ indicates function in focus and $g(\cdot)$ represents appropriate calculation function. When $x \in [a, b]$ is a first (arbitrary) guess and x^* denotes opposite value, then in all iterations, they evaluate $f(x)$ and $f(x^*)$. The learning remains with x if $g\big(f(x)\big) \geq g\big(f(x^*)\big)$, or else with x^*. The first solution of CGO population is given by

$$x_i = x_{low} + rand\big(x_{up} - x_{low}\big) \; j = 1, \dots, S \quad (10)$$

Whereas S denotes size of the population. The matching opposite solution x_i^* of CGO population is depending upon OBL is given by

$$x_i^* = x_{up} + x_{low} - x_i \quad (11)$$

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof

Then, the OCGOC technique derives a fitness function using three input parameters namely trust level, distance to neighbors, and energy for CH selection [23]. Fig. 3 demonstrates the flowchart of CSG.

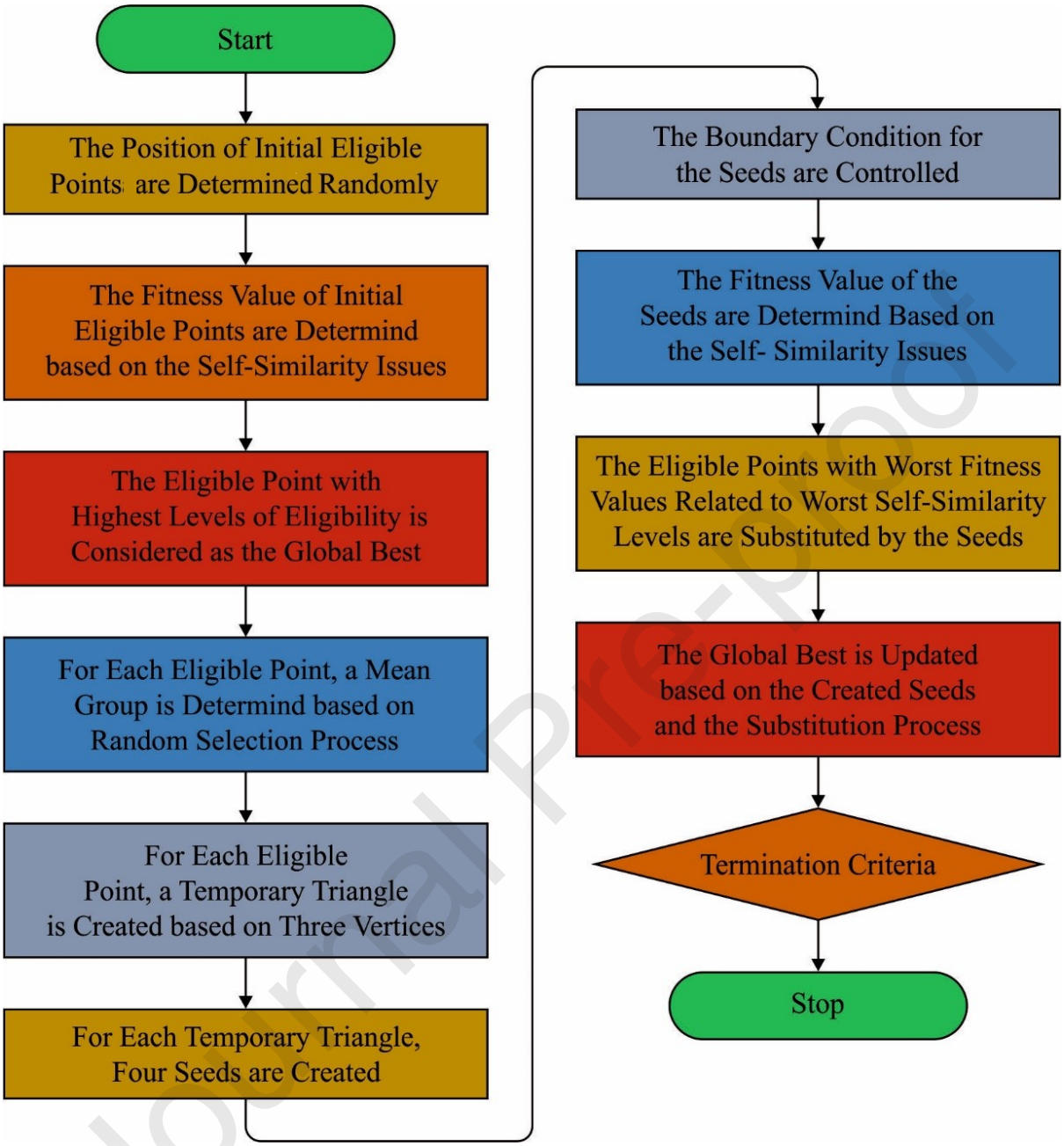


Fig. 3. Flowchart of CSG

Distance to neighbors: It can be appropriate for electing CHs with minimal distance among neighbouring nodes. During the intra-cluster communications process, sensor node consumptions power to CH transmission. If the neighbour node's distance is diminished, afterward the power of intra-cluster transmission is also minimized.

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof

Objective 1: Minimize

$$f_1 = \sum_{j=1}^{m} \frac{1}{l_j} \left(\sum_{i=1}^{l_j} dis(CH_j, s_i) \right) \quad (12)$$

Trust factor (TF): Firstly, all the nodes are stated that TF is one. The value of TF is diminished by anomalous forecast model when the node procedures the abnormal task and node is called a malicious node.

Objective 2: Maximize

$$f_2 = \sum_{j=1}^{m} \frac{1}{m} (TF_j) \quad (13)$$

Energy: It can be quantity of power consumed as CHs to RE of CHs. If a CH intake lesser power use as sensing, processes, and communication procedures also with higher RE is collected of lower energy ratio. Thus low as energy ratio, the CH election develops further feasible.

Objective 3: Minimize

$$f_3 = \sum_{j=1}^{m} \frac{E_c(CH_j)}{E_R(CH_j)} \quad (14)$$

During the projected MOSSA-CR model, it can be essential for reducing the linear combination of on objective function. Therefore, the potential energy function of MOSSA-CR technique is executed as:

$$Minimize\ Potiential\ energy\ function = \alpha_1 \times f_1 + \alpha_2 \times f_2 + \alpha_3 \times f_3 \quad (15)$$

Where $\alpha_1 + \alpha_2 + \alpha_3 = 1, \alpha_2 \geq (\alpha_1 + \alpha_3). Also\ 0 < f_1, f_2, f_3 < 1.$

3.2. Trust based Secure Routing using SRP-QSPO Technique

The primary motivation of SPO is the attacking and migration behaviors of sandpiper. They are seabirds that could be established around the world and they eat earthworms, insects, fish, etc. Generally, they live in colonies [24]. It utilizes its intelligence to attack and find the prey. Migration is determined as seasonal motion of sandpipers from one position to other for locating the food rich and plentiful sources that would give essential energy. They often attack

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof

migrating birds on sea if they migrate from one position to other. This behavior is equated in this manner which could be related to the objective function to be enhanced.

This migration and attacking behaviors are modeled as follows. The set of sandpiper moves from one place to another in migration the novel search agent location to evade the collision avoidance among neighboring sandpipers are given by;

$$\vec{S}_{CA} = \vec{M}_s * \vec{A}_{Cp}(i) \tag{16}$$

Whereas $\vec{S}_{CA}$ denotes location of search agent that doesn't collide with another search agent, $\vec{A}_{Cp}$ represents present location of agent, i denotes present iteration, $\vec{M}_s$ indicates motion of search agent in search space.

$$M_s = M_v - (i * M_v / Max_{iteration}) \tag{17}$$

$$i = 0,1,2 \ldots Max_{iteration} \tag{18}$$

Whereas M_v indicates control frequency for adjusting the parameter M_s that is gradually reduced from M_v to zero, when the variable M_v is fixed to two, the parameter CA is always reduced from two to zero.

Afterward collision avoidance, the search agent moving to the direction of optimum neighbour.

$$\vec{L}_{SA} = C_R * \left(\vec{A}_{Best}(i) - \vec{A}_{cp}(i)\right) \tag{19}$$

Whereas $\vec{L}_{SA}$ denotes location of search agent to an optimum fittest search agent. C_R represents arbitrary parameter that is in charge of optimum exploration. Later the C_R is estimated by:

$$C_R = 0.5 * Rand \tag{20}$$

Whereas $Rand$ indicates arbitrary amounts in the range of zero and one. Lastly, the search agent/sandpiper could upgrade its location resultant to optimum search agent.

$$\vec{G}_{AF} = \vec{S}_{CA} + \vec{L}_{SA} \tag{21}$$

Whereas $\vec{G}_{AF}$ determines the gap among the search agent & optimum fittest search agent.

Sandpiper generates spiral behavior, whereas attacking the prey, in the air. This behavior can be given by.

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof

$$a' = radius * \sin(n) \quad (22)$$

$$b' = radius * \cos(n) \quad (23)$$

$$c' = radius * n \quad (24)$$

$$R = p * b^{zq} \quad (25)$$

Where represents radius of every turns of spiral, n denotes parameter in the range of $[0 \leq k \leq 2\pi]$. P and q indicate constant for defining the spiral shape, and b denotes base of natural logarithm. If the value of this constant exceeds one then the shape of spiral is highly complicated. Thus, the upgraded location of search agent can be estimated by:

$$\vec{A}_{Cp}(i) = \left(\vec{G}_{AF} * (a' + b' + c')\right) * \vec{A}_{Best}(i) \quad (26)$$

Whereas $\vec{A}_{Cp}(i)$ upgrade the location of other search agents and save optimum solutions. To improve the performance of the SPO algorithm, QPSO algorithm is derived.

For determining a better group of routes, the provided function is utilized for defining the succeeding hop to BS and is definite as:

$$f(x) = \{i, for\ which \left|\left(\frac{i}{k} - X_{i_f j}\right)\right| is\ minimum, \forall_i 1 \leq i \leq k \quad (27)$$

The purpose is for determining a better group of routes in CHs to BS utilizing an FF containing 2 parameters like energy and distance. Initially, the RE of next-hop nodes are defined and the node with maximal energy is given as relay nodes. For transmitting data, the source nodes send to relay nodes that more forwarding to BS using inter CHs. So, the nodes with superior RE are given as next-hop nodes. The initial sub-objective $f1$ is providing as:

$$f1 = E_{CH} \quad (28)$$

Also, Euclidean distance is implemented for determining the distance in CHs to BS. The minimization of energy dissipation is mostly dependent upon the transmission distance. With a lower distance, the energy is kept considerable. Therefore, the node with minimal distances is preferred to a relay node. Thus, the next sub-objective by means of distance is $f2$ that is expressed as:

$$f2 = \frac{1}{\sum_{i=1}^{m} d\ is(CH_i, NH) + dis(NH, BS)} \quad (29)$$

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof

The aforementioned sub-objective is reviewed as to FF as provided under in which the α_1 and α_2 refers to the weights allocated for all sub-objectives.

$$Fitness\ = \alpha_1(f1) + \alpha_2(f2), where \sum_{i=1}^{2} \alpha_i = 1 \alpha_i \varepsilon(0,1); \quad (30)$$

Quantum process is a new type of calculation model which implies the techniques significant to quantum theory as quantum estimate, state superposition, and quantum entanglement. The basic module of quantum calculation is qubit. In 2 basic situations are $|0>$ and $|1>$ progress a qubit that is signified as linear integration of two important situations as provided under.

$$|Q>= \alpha|0> +\beta|1>. \quad (31)$$

$|\alpha|^2$ implies the possibilities of detecting state $|0>$, $|\beta|^2$ refers the possibility of detecting state $|1>$, where $|\alpha|^2 + |\beta|^2 = 1$. In quantum is manufactured utilizing n qubits. Due to quantum superposition, all the quanta are collected of 2^n possible values. A n-qubits quantum is signified as.

$$\Psi = \sum_{x=0}^{2^n-1} C_\chi |x>, \sum_{x=0}^{2^n-1} |C_\chi|^2 = 1. \quad (32)$$

Quantum gates are implied for modifying the situation of qubits as rotation gate, NOT gate, Hadamard gate, etc. Primarily, the Rotation gate is determined as mutation function to create an efficiency quanta method and recognizes the global optimum solution. The rotation gate is showcased as given under:

$$\begin{bmatrix} \alpha^d(t+1) \\ \beta^d(t+1) \end{bmatrix} = \begin{bmatrix} \cos(\Delta\theta^d) & -\sin(\Delta\theta^d) \\ \sin(\Delta\theta^d) & \cos(\Delta\theta^d) \end{bmatrix} \begin{bmatrix} \alpha^d(t) \\ \beta^d(t) \end{bmatrix} ford = 1,2,\dots,n. \quad (33)$$

$\Delta\theta^d = \Delta \times S(\alpha^d, \beta^d), \Delta\theta^d$ represents the rotation angle of qubit, where Δ and $S(\alpha^d, \beta^d)$ implies the size and direction of rotation respectively.

4. Experimental Validation

This section examines the performance of the proposed EEC-SDTP technique with other techniques in terms of different measures. The EEC-SDTP technique is simulated using MATLAB. The experimental results are inspected under varying number of nodes and execution rounds. The simulation analysis is performed on network, with the node count of

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof

100-1000, which are placed under random topology. The results are examined under several rounds up to 20000. The searching speed considered is 0.02(m/s). The parameter setting is given as follows: Initial energy: 0.1K, number of nodes: 1000, maximum bandwidth 54Mbps, and MAC Layer protocol IEEE802.11a. The EEC-SDTP technique is executed on a PC Motherboard - MSI Z370 A-Pro, Processor - i5-8600k, Graphics Card - GeForce 1050Ti 4GB, RAM - 16GB, OS Storage - 250GB, and SSD File Storage - 1TB HDD. A detailed comparison study is made with grey wolf optimization algorithm based clustering (GWOA-C), krill herd algorithm based clustering (KHA-C) [25], gravitational search algorithm based clustering (GSA-C), and cuckoo search optimization algorithm based clustering (CSOA-C) techniques.

A brief average residual energy analysis of the EEC-SDTP method with other existing techniques takes place in Table 1 and Fig. 4. The experimental outcomes demonstrated that the presented EEC-SDTP technique has offered better performance with the worse average residual energy values under the distinct number of nodes. For instance, with 1000 nodes, the proposed EEC-SDTP model has gained effective results with an average residual energy of 1.00J whereas the GWOA-C, KHA-C, GSA-C, and CSOA-C techniques have accomplished worse results with the PLR of 0.91J, 0.93J, 0.95J, and 0.97J correspondingly. In line with, with 6000 nodes, the presented EEC-SDTP technique has accomplished better performance with the average residual energy of 0.97J whereas the GWOA-C, KHA-C, GSA-C, and CSOA-C techniques have accomplished worse results with the average residual energy of 0.45J, 0.48J, 0.50J, and 0.41J respectively. Also, with 9000 nodes, the projected EEC-SDTP model has resulted in an effective performance with an average residual energy of 0.84J whereas the GWOA-C, KHA-C, GSA-C, and CSOA-C techniques have accomplished worse outcomes with the average residual energy of 0.15J, 0.19J, 0.22J, and 0.18J correspondingly.

Table 1 Average residual energy analysis of different methods with proposed EEC-SDTP algorithm

No. of Rounds	**Average Residual Energy (J)**				
	GWOA-C	**KHA-C**	**GSA-C**	**CSOA-C**	**EEC-SDTP**
0	1.00	1.00	1.00	1.00	1.00
1000	0.91	0.93	0.95	0.97	1.00
2000	0.90	0.92	0.94	0.95	1.00
3000	0.87	0.90	0.92	0.90	1.00
4000	0.83	0.85	0.91	0.80	0.99
5000	0.76	0.78	0.85	0.54	0.98
6000	0.45	0.48	0.50	0.41	0.97

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

7000	0.36	0.39	0.40	0.30	0.93
8000	0.24	0.28	0.30	0.24	0.87
9000	0.15	0.19	0.22	0.18	0.84
10000	0.00	0.08	0.10	0.12	0.76
11000	0.00	0.00	0.03	0.05	0.68
12000	0.00	0.00	0.00	0.02	0.64
13000	0.00	0.00	0.00	0.00	0.58
14000	0.00	0.00	0.00	0.00	0.45
15000	0.00	0.00	0.00	0.00	0.34
16000	0.00	0.00	0.00	0.00	0.23
17000	0.00	0.00	0.00	0.00	0.14
18000	0.00	0.00	0.00	0.00	0.09
19000	0.00	0.00	0.00	0.00	0.04
20000	0.00	0.00	0.00	0.00	0.00

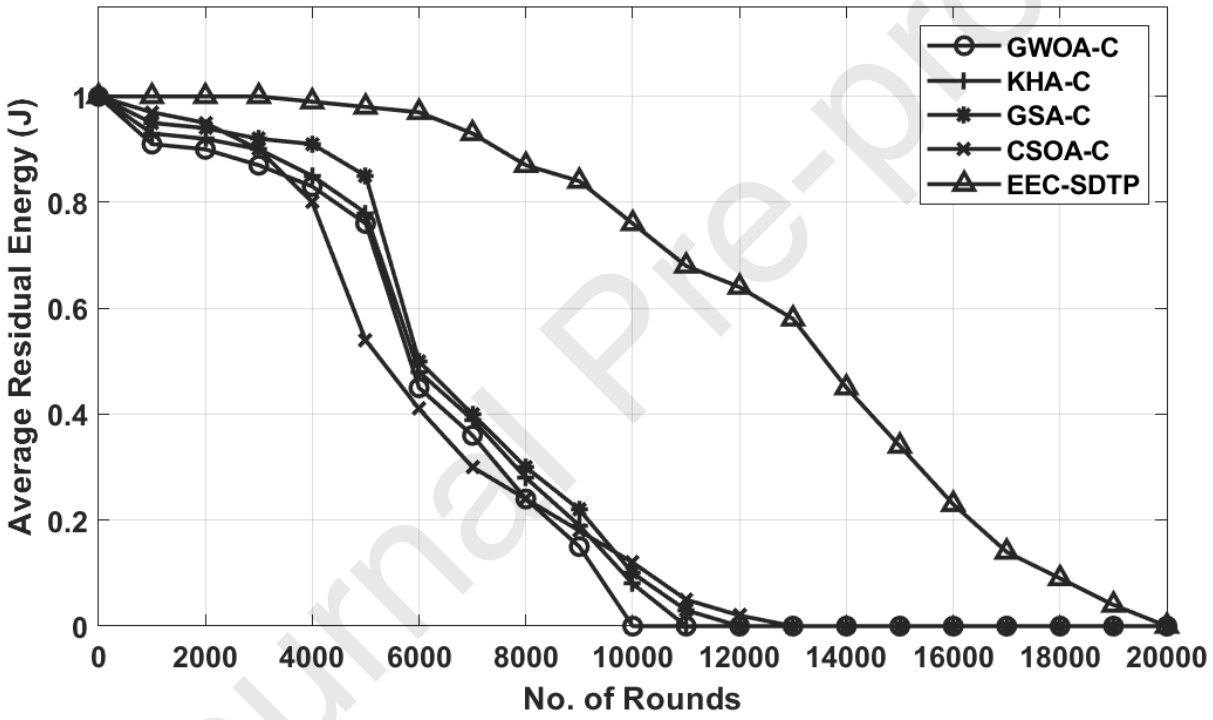


Fig. 4. Average residual energy analysis of EEC-SDTP model

The network lifetime analysis of the EEC-SDTP technique with other techniques takes place in Table 2 and Fig. 5. From the figure, it is evident that the EEC-SDTP technique has obtained maximal network lifetime with the lengthened FND, HND, and LND of 4008, 13247, and 19644 rounds respectively.

Table 2 First Node Die (FND), Half Node Die (HND), and Last Node Die (LND) analysis of different methods

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

	Network Lifetime (in Rounds)				
	GWOA-C	**KHA-C**	**GSA-C**	**CSOA-C**	**EEC-SDTP**
FND	423	733	912	994	4008
HND	5674	6154	7564	8221	13247
LND	9754	10342	11211	12023	19644

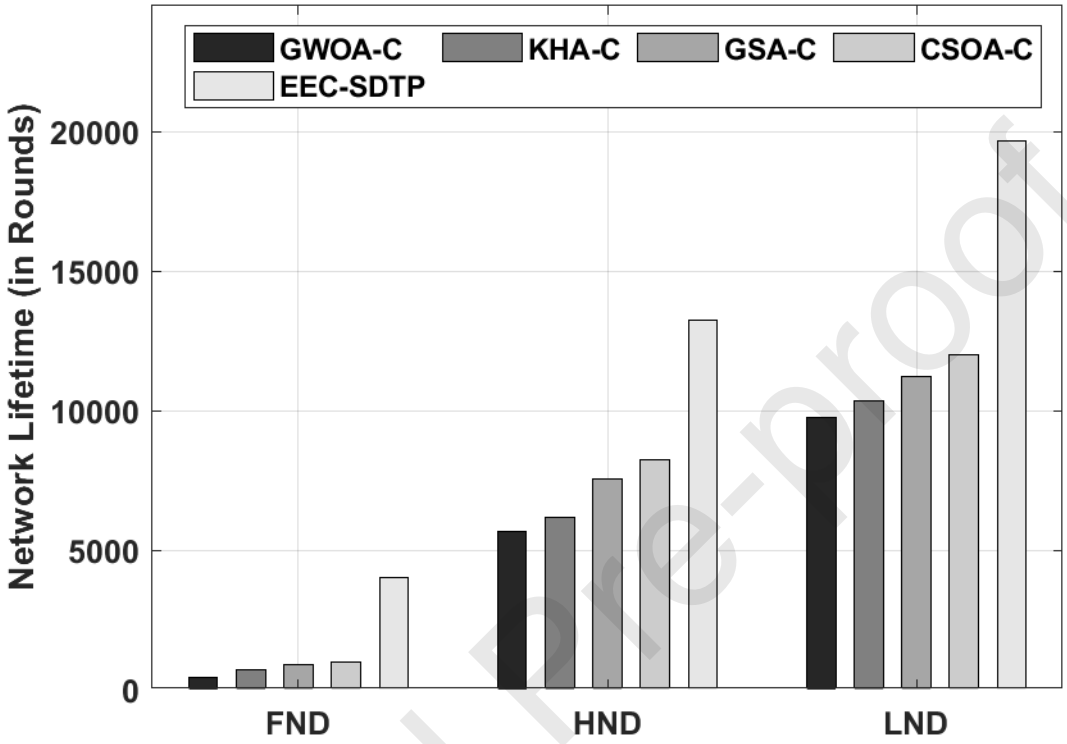


Fig. 5. Network lifetime analysis of EEC-SDTP model

A brief ETE delay analysis of the EEC-SDTP technique with other existing approaches take place in Table 3 and Fig. 6. The experimental outcomes exhibited that the proposed EEC-SDTP model has offered optimum performance with the least ETE delay values under distinct number of nodes. For sample, with 100 nodes, the proposed EEC-SDTP model has gained effective results with the ETE delay of 81ms whereas the GWOA-C, KHA-C, GSA-C, and CSOA-C techniques have accomplished worse results with the ETE delay of 128ms, 122ms, 104ms, and 90ms respectively. Likewise, with 500 nodes, the proposed EEC-SDTP model has accomplished better performance with the ETE delay of 123ms whereas the GWOA-C, KHA-C, GSA-C, and CSOA-C techniques have accomplished worse results with the ETE delay of 168ms, 155ms, 149ms, and 133ms correspondingly. Similarly, with 1000 nodes, the proposed EEC-SDTP model has resulted to an effective performance with the ETE delay of 149ms

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof

whereas the GWOA-C, KHA-C, GSA-C, and CSOA-C methodologies have accomplished poor results with the ETE delay of 236ms, 210ms, 193ms, and 158ms correspondingly.

Table 3 End to End Delay analysis of different methods

No. of Nodes	Average Delay (ms)				
	GWOA-C	KHA-C	GSA-C	CSOA-C	EEC-SDTP
100	128	122	104	90	81
200	135	132	119	103	94
300	147	140	134	118	106
400	153	150	141	123	111
500	168	155	149	133	123
600	178	170	157	138	128
700	183	175	169	143	130
800	198	192	179	148	136
900	218	200	188	153	143
1000	236	210	193	158	149

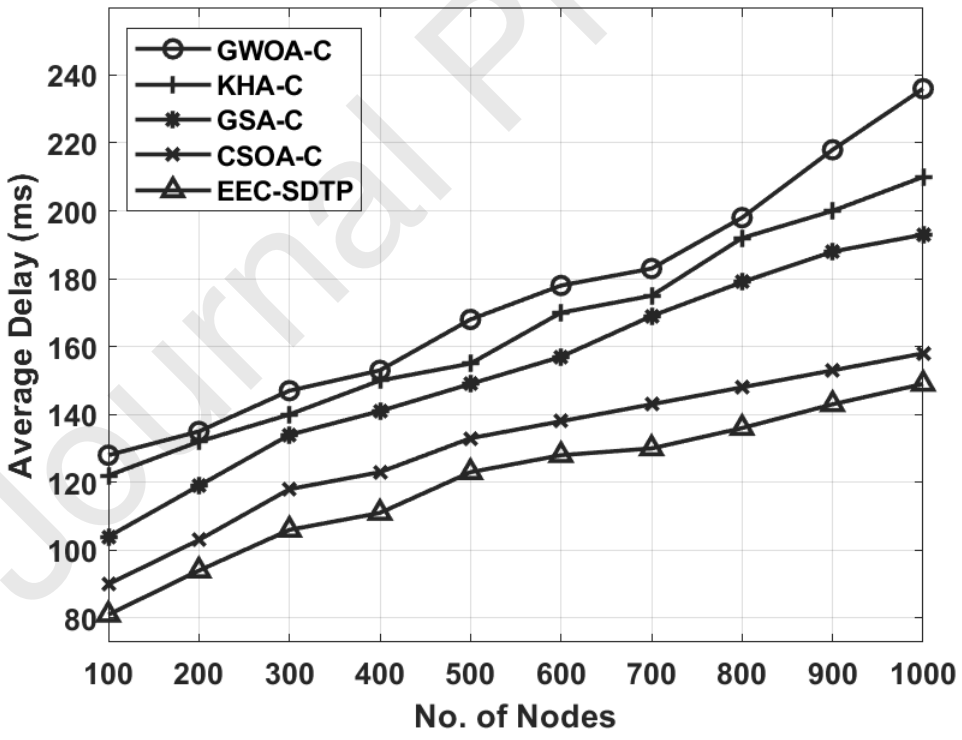


Fig. 6. Average delay analysis of EEC-SDTP model

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof

A comprehensive PDR analysis of the EEC-SDTP technique with other techniques takes place in Table 4 and Fig. 7. The experimental results demonstrated the supremacy of the EEC-SDTP technique over the other methods with the maximal PDR. For instance, under 100 nodes, the EEC-SDTP technique has reached a maximal PDR of 0.950 whereas the GWOA-C, KHA-C, GSA-C, and CSOA-C techniques have accomplished minimum PDR of 0.770, 0.834, 0.850, and 0.890 respectively. Moreover, under 500 nodes, the EEC-SDTP model has gained a higher PDR of 0.915 whereas the GWOA-C, KHA-C, GSA-C, and CSOA-C techniques have attained a lower PDR of 0.680, 0.760, 0.780, and 0.840 correspondingly. Furthermore, under 1000 nodes, the EEC-SDTP methodology has gained an increased PDR of 0.830 whereas the GWOA-C, KHA-C, GSA-C, and CSOA-C techniques have offered a decreased PDR of 0.605, 0.630, 0.670, and 0.770 correspondingly.

Table 4 PDR analysis of different methods

No. of Nodes	**Packet Delivery Ratio**				
	GWOA-C	**KHA-C**	**GSA-C**	**CSOA-C**	**EEC-SDTP**
100	0.770	0.834	0.850	0.890	0.950
200	0.750	0.820	0.840	0.860	0.945
300	0.720	0.810	0.830	0.850	0.936
400	0.700	0.780	0.780	0.840	0.924
500	0.680	0.760	0.770	0.830	0.915
600	0.650	0.720	0.750	0.820	0.905
700	0.630	0.710	0.730	0.810	0.894
800	0.620	0.680	0.710	0.800	0.875
900	0.610	0.660	0.690	0.790	0.855
1000	0.605	0.630	0.670	0.770	0.830

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

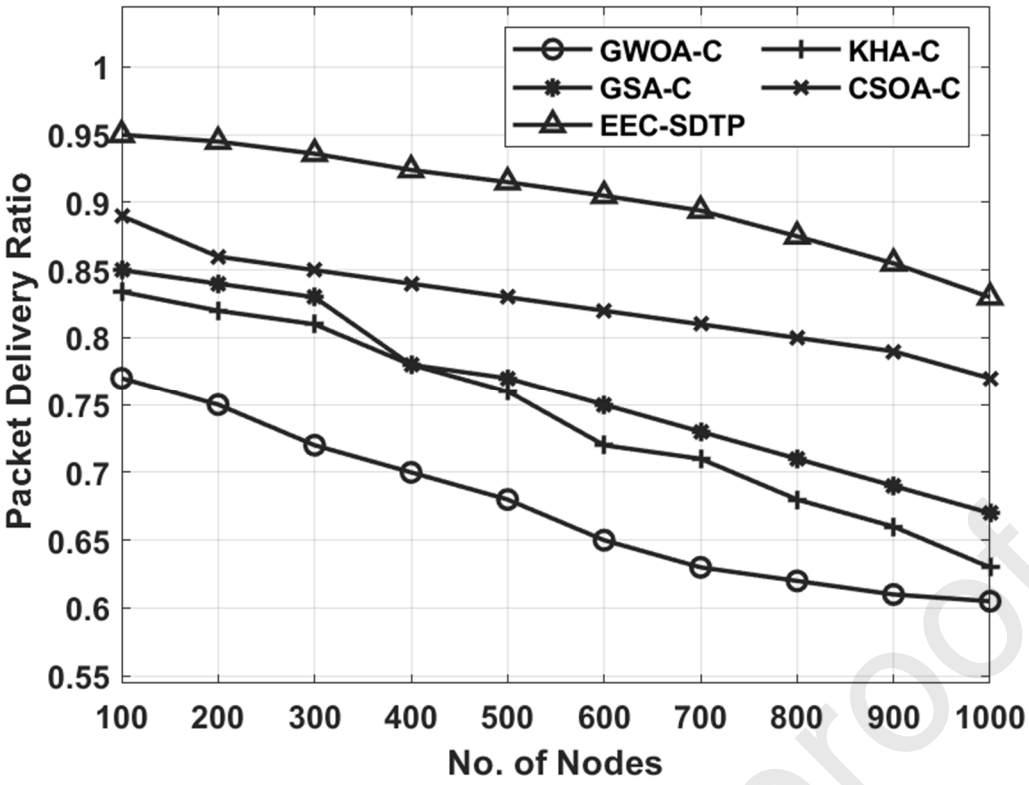


Fig. 7. PDR analysis of EEC-SDTP model

A brief PLR analysis of the EEC-SDTP model with other existing techniques takes place in Table 5 and Fig. 8. The experimental outcomes showcased that the proposed EEC-SDTP technique has offered optimum performance with the least PLR values under distinct number of nodes. For instance, with 100 nodes, the proposed EEC-SDTP model has gained an effective outcome with the PLR of 0.050 whereas the GWOA-C, KHA-C, GSA-C, and CSOA-C techniques have accomplished worse results with the PLR of 0.230, 0.166, 0.150, and 0.110 respectively. Likewise, with 500 nodes, the proposed EEC-SDTP model has accomplished better performance with the PLR of 0.085 whereas the GWOA-C, KHA-C, GSA-C, and CSOA-C techniques have accomplished worse results with the PLR of 0.320, 0.240, 0.230, and 0.170 respectively. Similarly, with 1000 nodes, the proposed EEC-SDTP approach has resulted in an effectual performance with the PLR of 0.170 whereas the GWOA-C, KHA-C, GSA-C, and CSOA-C techniques have accomplished worse results with the PLR of 0.395, 0.370, 0.330, and 0.230 respectively.

Table 5 PLR analysis of different methods

No. of Nodes	Packet Loss Rate (%)				
	GWOA-C	KHA-C	GSA-C	CSOA-C	EEC-SDTP

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

100	0.230	0.166	0.150	0.110	0.050
200	0.250	0.180	0.160	0.140	0.055
300	0.280	0.190	0.170	0.150	0.064
400	0.300	0.220	0.220	0.160	0.076
500	0.320	0.240	0.230	0.170	0.085
600	0.350	0.280	0.250	0.180	0.095
700	0.370	0.290	0.270	0.190	0.106
800	0.380	0.320	0.290	0.200	0.125
900	0.390	0.340	0.310	0.210	0.145
1000	0.395	0.370	0.330	0.230	0.170

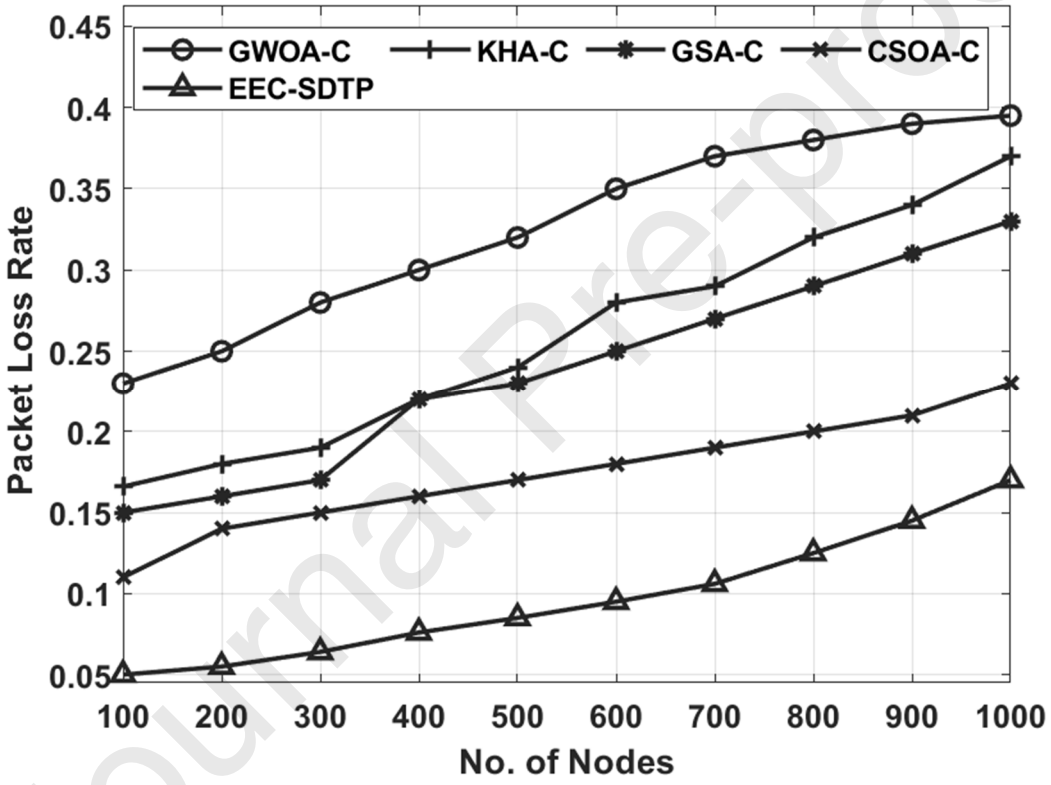


Fig. 8. PLR analysis of EEC-SDTP model

From the aforementioned outcomes analysis, it can be evident that the EEC-SDTP technique has accomplished maximum energy efficiency and lifetime of the IoT edge computing systems.

5. Conclusion

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof

This paper has developed a new EEC-SDTP technique to achieve energy efficiency and secured data transmission in intelligent IoT edge systems. The EEC-SDTP technique is mainly focused on the choice of optimal CHs and secure routes for data transmission in the network. The proposed model involves OCGOC technique for CH selection, which derives a fitness function utilizing 3 input parameters such as trust level, distance to neighbors, and energy. In addition, SRP-QSPO technique is applied to determine trustworthy routes to destination for secure data transmission in the IoT edge system. The inclusion of trust factor helps to choose the reliable nodes in the network for CH selection and data transmission. For examining the improved efficiency of the proposed EEC-SDTP technique, a detailed set of simulation analyses is made and the results are inspected in terms of distinct evaluation parameters. The experimental results highlighted the superior performance of the EEC-SDTP technique over the recent state of art methods under distinct measures. As a part of future extension, effective resource allocation and task scheduling strategies can be developed to proficiently exploit the available resources in the IoT edge systems.

Conflict of Interest

The authors declare that they have no conflict of interest. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Data Availability Statement

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

Consent to Participate

Not applicable.

ACKNOWLEDGEMENTS

This work is partially funded by FCT/MCTES through national funds and when applicable co-funded EU funds under the Project UIDB/50008/2020; and by Brazilian National Council for Scientific and Technological Development - CNPq, via Grant No. 313036/2020-9.

References

[1] Kwon, M., Lee, J. and Park, H., 2019. Intelligent IoT connectivity: deep reinforcement learning approach. *IEEE Sensors Journal*, *20*(5), pp.2782-2791.

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof

[2] V. Nazarzehi and A. V. Savkin, “Distributed self-deployment of mobile wireless 3D robotic sensor networks for complete sensing coverage and forming specific shapes,” Robotica, vol. 36, no. 1, pp. 1–18, 2018

[3] M. Erdelj, M. Krol, and E. Natalizio, “Wireless sensor networks and multi- ´ UAV systems for natural disaster management,” Computer Networks, vol. 124, pp. 72–86, 2017.

[4] Maddikunta, P.K.R., Gadekallu, T.R., Kaluri, R., Srivastava, G., Parizi, R.M. and Khan, M.S., 2020. Green communication in IoT networks using a hybrid optimization algorithm. *Computer Communications*, *159*, pp.97-107.

[5] H. Atlam, R. Walters, G. Wills, Fog computing and the internet of things: a review, big data and cognitive computing, 2, (2), 2018, p. 10.

[6] T.M. Behera, S.K. Mohapatra, U.C. Samal, M.S. Khan, M. Daneshmand, A.H. Gandomi, Residual energy-based cluster-head selection in WSNs for IoT application, IEEE Internet Things J. 6 (3) (2019) 5132–5139.

[7] D. Cacciagrano, R. Culmone, M. Micheletti, L. Mostarda, Energy-efficient clustering for wireless sensor devices in internet of things, in: Performability in Internet of Things, Springer, 2019, pp. 59–80.

[8] Almiani, M., AbuGhazleh, A., Al-Rahayfeh, A., Atiewi, S. and Razaque, A., 2020. Deep recurrent neural network for IoT intrusion detection system. *Simulation Modelling Practice and Theory*, *101*, p.102031.

[9] Yu, W., Liang, F., He, X., Hatcher, W.G., Lu, C., Lin, J. and Yang, X., 2017. A survey on the edge computing for the Internet of Things. *IEEE access*, *6*, pp.6900-6919.

[10] Ilyas, M., Ullah, Z., Khan, F.A., Chaudary, M.H., Malik, M.S.A., Zaheer, Z. and Durrani, H.U.R., 2020. Trust-based energy-efficient routing protocol for Internet of things–based sensor networks. International Journal of Distributed Sensor Networks, 16(10), p.1550147720964358.

[11] Santos, A.L., Cervantes, C.A., Nogueira, M. and Kantarci, B., 2019. Clustering and reliability-driven mitigation of routing attacks in massive IoT systems. Journal of Internet Services and Applications, 10(1), pp.1-17.

[12] Fang, W., Zhang, W., Yang, W., Li, Z., Gao, W. and Yang, Y., 2021. Trust management-based and energy efficient hierarchical routing protocol in wireless sensor networks. Digital Communications and Networks.

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof

[13] Markiewicz, R.P. and Sgandurra, D., 2020, August. Clust-IT: clustering-based intrusion detection in IoT environments. In Proceedings of the 15th International Conference on Availability, Reliability and Security (pp. 1-9).

[14] Awan, K.A., Din, I.U., Almogren, A., Guizani, M. and Khan, S., 2020. StabTrust—A stable and centralized trust-based clustering mechanism for IoT enabled vehicular ad-hoc networks. Ieee Access, 8, pp.21159-21177.

[15] Alshehri, M.D. and Hussain, F.K., 2019. A fuzzy security protocol for trust management in the internet of things (Fuzzy-IoT). Computing, 101(7), pp.791-818.

[16] Wang, T., Qiu, L., Sangaiah, A.K., Liu, A., Bhuiyan, M.Z.A. and Ma, Y., 2020. Edge-computing-based trustworthy data collection model in the internet of things. IEEE Internet of Things Journal, 7(5), pp.4218-4227.

[17] Kalkan, K., 2020. SUTSEC: SDN Utilized trust based secure clustering in IoT. Computer Networks, 178, p.107328.

[18] Sadrishojaei, M., Navimipour, N.J., Reshadi, M. and Hosseinzadeh, M., 2021. A new preventive routing method based on clustering and location prediction in the mobile Internet of Things. IEEE Internet of Things Journal.

[19] Malchi, S.K., Kallam, S., Al-Turjman, F. and Patan, R., 2021. A trust-based fuzzy neural network for smart data fusion in internet of things. Computers & Electrical Engineering, 89, p.106901.

[20] Alqahtani, F., Al-Makhadmeh, Z., Tolba, A. and Said, O., 2020. TBM: A trust-based monitoring security scheme to improve the service authentication in the Internet of Things communications. Computer Communications, 150, pp.216-225.

[21] Kaur, A., Jain, S. and Goel, S., 2020. Sandpiper optimization algorithm: a novel approach for solving real-life engineering problems. *Applied Intelligence*, *50*(2), pp.582-619.

[22] Tizhoosh, H.R. Opposition-based learning: A new scheme for machine intelligence. In Proceedings of theInternational Conference on Computational Intelligence for Modelling, Control and Automation, CIMCA2005 and International Conference on Intelligent Agents, Web Technologies and Internet, Vienna, Austria,28–30 November 2005; pp. 695–701.

[23] Lv, X., Chen, H., Zhang, Q., Li, X., Huang, H. and Wang, G., 2018. An improved bacterial-foraging optimization-based machine learning framework for predicting the severity of somatization disorder. *Algorithms*, *11*(2), p.17.

[24] Talatahari, S. and Azizi, M., 2021. Chaos Game Optimization: a novel metaheuristic algorithm. *Artificial Intelligence Review*, *54*(2), pp.917-1004.

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof

[25] Karthick, P.T. and Palanisamy, C., 2019. Optimized cluster head selection using krill herd algorithm for wireless sensor network. Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije, 60(3), pp.340-348.

Journal Pre-proof

Corresponding Authors (Deepak Gupta, Shahab Shamshirband)

Journal Pre-proof



M. Padmaa received her M.E. (VLSI Systems) from NIT, Tiruchirappalli and Ph.D. from Anna University, Chennai. Currently she is working as professor in Saranathan College of Engineering, Trichy. She has teaching experience of 32 years and has published 15 research articles in National and International journals. Her research area includes Network security and information security.



Dr. T. Jayasankar is working as an Assistant Professor(Sr.Gr) in the Electronics and Communication Engineering Department, University College of Engineering (BIT Campus), Anna University, Tiruchirappalli, Tamilnadu, India. He received the B.E. Degree in Electronics and Communication Engineering from Bharathiyar University, Coimbatore in 2001 and M.E. Degree at Madurai Kamaraj University, Madurai in 2003 and Ph.D. in Speech Processing at Anna University Chennai 2017. He is having 19 years of teaching experience. He is a Member of IEI, ISTE. He has participated in a more number of International and National level Conferences and Workshops. He has published more than 70 papers in the reputed International Journals including Science Indexed Journals and Scopus Indexed Journals, 20 papers in the International and National Conferences. His main interest is currently Speech Synthesis, Speech and Image Processing, Bio Medical Signal Processing and Wireless Networks. He is also a peer reviewer in Elsevier, Springer, SCOPUS and SCI indexed International Journals.



Dr. S. Venkatraman received his B.E. degree in Computer Science and Engineering from Bharathidasan University, Tiruchirappalli, India in 2001 and his M.E. degree in software engineering from Anna University, Chennai, India in 2004. He has completed Ph.D in CSE form National Institute of Technology

Journal Pre-proof

Puducherry in 2019. At present, he is working as an Associate Professor in School of Computer Science & Engineering, Vellore Institute of Technology, Chennai. He has published 19 research papers in leading international journals, 6 papers in an international conferences and 2 books.



Dr. Ashit Kumar is working as an associate professor in the department of computer science and information systems. His specialization is in the field of artificial intelligence and cyber security. He is a certified ethical hacker from EC Council, U.S.A. He has experience of 18 years in both national and international level of education. He has published more than 60 research papers, and most of them are ISI journals. He has completed 5 research projects in the local and international levels, respectively. He is the author of 4 books. He is serving as Managing Editor in chief and editorial member in various international journals. He is serving as multiple membership of professional bodies /societies. He is also one of the committee members of Scientific Research, curriculum development and Quality management of educational institutions. He got various meritorious award in both national and international levels.



Dr. Deepak Gupta (SMIEEE, MACM) received a B.Tech. degree in 2006 from the Guru Gobind Singh Indraprastha University, Delhi, 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 (AKTU), Lucknow, India. He has completed his Post-Doc from National Institute of Telecommunications (Inatel), Brazil in 2018. He has co-authored more than 153 journal articles including 106 SCI papers and 44 conference articles. He has authored/edited 50 books, published by IEEE-Wiley, Elsevier, Springer, Wiley, CRC Press, DeGruyter and Katsons. He has filled four Indian patents. He is convener of ICICC, ICDAM & DoSCI Springer conferences series. Currently he is Associate Editor of Expert Systems (Wiley), and Intelligent Decision Technologies

Journal Pre-proof

(IOS Press). Deepak Gupta have also been featured in the list of top 2% scientist/researcher database in the world [Table-S7-singleyr-2019]. He is also working towards promoting Startups and also serving as a Startup Consultant. He is also a series editor of “Elsevier Biomedical Engineering” at Academic Press, Elsevier, “Intelligent Biomedical Data Analysis” at De Gruyter, Germany, “Explainable AI (XAI) for Engineering Applications” at CRC Press. He is appointed as Consulting Editor at Elsevier.



SHAHAB S. BAND received the M.Sc. degree in artificial intelligence from Iran, and the Ph.D. degree in computer science from the University of Malaya (UM), Malaysia, in 2014. He was an Adjunct Assistant Professor with the Department of Computer Science, Iran University of Science and Technology. He also severed as a Senior Lecturer with UM, Malaysia, and with Islamic Azad University, Iran. He participated in many research programs within the Center of Big Data Analysis, IUST and IAU. He has been associated with young researchers and elite club, since 2009. He supervised or co-supervised undergraduate and postgraduate students (master’s and Ph.D.) by research and training. He has also authored, or coauthored papers published in IF journals and attended to high-rank A and B conferences. He is an Associate Editor, a Guest Editor, and a Reviewer of high-quality journals and conferences. He is a professional member of the ACM.



Prof. Joel J.P.C. Rodrigues [S’01, M’06, SM’06, F’20] is a professor at the Federal University of Piauí, Brazil; and senior researcher at the Instituto de Telecomunicações, Portugal. Prof. Rodrigues is the leader of the Next Generation Networks and Applications (NetGNA) research group (CNPq), an IEEE Distinguished Lecturer, Member Representative of the IEEE Communications Society on the IEEE Biometrics Council, and the President of the scientific

Journal Pre-proof

council at ParkUrbis – Covilhã Science and Technology Park. He was Director for Conference Development - IEEE ComSoc Board of Governors, Technical Activities Committee Chair of the IEEE ComSoc Latin America Region Board, a Past-Chair of the IEEE ComSoc Technical Committee (TC) on eHealth and the TC on Communications Software, a Steering Committee member of the IEEE Life Sciences Technical Community and Publications co-Chair. He is the editor-in-chief of the International Journal of E-Health and Medical Communications and editorial board member of several high-reputed journals (mainly, from IEEE). He has been general chair and TPC Chair of many international conferences, including IEEE ICC, IEEE GLOBECOM, IEEE HEALTHCOM, and IEEE LatinCom. He has authored or coauthored about 1000 papers in refereed international journals and conferences, 3 books, 2 patents, and 1 ITU-T Recommendation. His contributions received more than 21700 known citations in the literature with an H-index 73 (Google Scholar). He had been awarded several Outstanding Leadership and Outstanding Service Awards by IEEE Communications Society and several best papers awards. Prof. Rodrigues is a member of the Internet Society, a senior member ACM, and Fellow of IEEE.

Journal Pre-proof

Journal Pre-proof

Declaration of interests

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

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof

Journal Pre-proof

AUTHORSHIP STATEMENT

Manuscript title: _____

_____

_____

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript. Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication before its appearance in the *Hong Kong Journal of Occupational Therapy*.

Authorship contributions
Please indicate the specific contributions made by each author (list the authors' initials followed by their surnames, e.g., Y.L. Cheung). The name of each author must appear at least once in each of the three categories below.

Category 1
Conception and design of study: _____, _____, _____, _____;

acquisition of data: _____, _____, _____, _____;

analysis and/or interpretation of data: _____, _____, _____, _____.

Category 2
Drafting the manuscript: _____, _____, _____, _____;

revising the manuscript critically for important intellectual content: _____, _____,

_____, _____.

Category 3
Approval of the version of the manuscript to be published (the names of all authors must be listed):

_____, _____, _____, _____, _____,

_____, _____, _____, _____, _____.

Journal Pre-proof

Journal Pre-proof

Acknowledgements

All persons who have made substantial contributions to the work reported in the manuscript (e.g., technical help, writing and editing assistance, general support), but who do not meet the criteria for authorship, are named in the Acknowledgements and have given us their written permission to be named. If we have not included an Acknowledgements, then that indicates that we have not received substantial contributions from non-authors.

This statement is signed by all the authors (*a photocopy of this form may be used if there are more than 10 authors***):**

Author's name (typed)	Author's signature	Date
	shahab shamshirband	
	joel jpc rofrigues	

Journal Pre-proof

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/MCE.2022.3153276, IEEE Consumer Electronics Magazine

Palmprint Biometric vs Encrypted Hash based Digital Signature for Securing DSP Cores Used in CE systems

Rahul Chaurasia*, Aditya Anshul*, Anirban Sengupta*, Shikha Gupta*+
*Computer Science & Engineering, Indian Institute of Technology (IIT) Indore
*+ Indian Institute of Management (IIM) Indore
*Email: asengupt@iiti.ac.in

Abstract

In the prior articles of reusable intellectual property (IP) core, different facets of their threat models and defence mechanisms against IP Trojan and IP piracy have been presented. This article provides novel comparative perspective between the two security techniques viz. palmprint biometric based approach and encrypted-hash based digital signature approach for IP cores, integrated in Consumer Electronics (CE) systems. These reusable IP cores such as finite impulse response (FIR) filter used in audio devices, HD TVs etc., discrete cosine transform (DCT) core used in audio/image/video/compression, fast fourier transform (FFT) used in digital video broadcasting and JPEG-CODEC used in image/video compression of camera systems can be secured using the two security methodologies discussed in this article. These both approaches enable detective control against important security threats such as IP piracy, fraudulent claim of IP ownership, overproduction and IP counterfeiting. The article therefore contributes the following: (a) security comparison of security techniques on digital signal processing (DSP) IP cores in terms of probability of coincidence (Pc) and tamper tolerance (TT) (b) comparison between two security techniques in terms of its properties (c) end to end demonstration of the palmprint biometric security approach (d) design flow of the encrypted-hash based digital signature approach.

1. Classification of Hardware Threats/ Vulnerabilities and their Countermeasures

Ensuring the integrity and security of reusable IP cores in the modern era, where globalization of integrated circuits (ICs) supply chain involves multi party vendors to counter the supply demand and time to market, must be of high priority [1,2]. Further due to involvement of offshore design houses several security threats are identified at the level of IP designer, system on chip (SoC) integrator and at foundry level. Major security threats are IP piracy, counterfeiting, IP forgery, fraudulent claim of IP ownership, hardware trojan and threat of reverse engineering the design [3]. Protection against threat of IP ownership is to authenticate genuine IP vendor/ designer in case of SoC integrator falsely claiming the ownership of the IP core. On the other hand, protection against IP piracy is to authenticate SoC integrator/ user from dishonest IP vendor selling extra copies

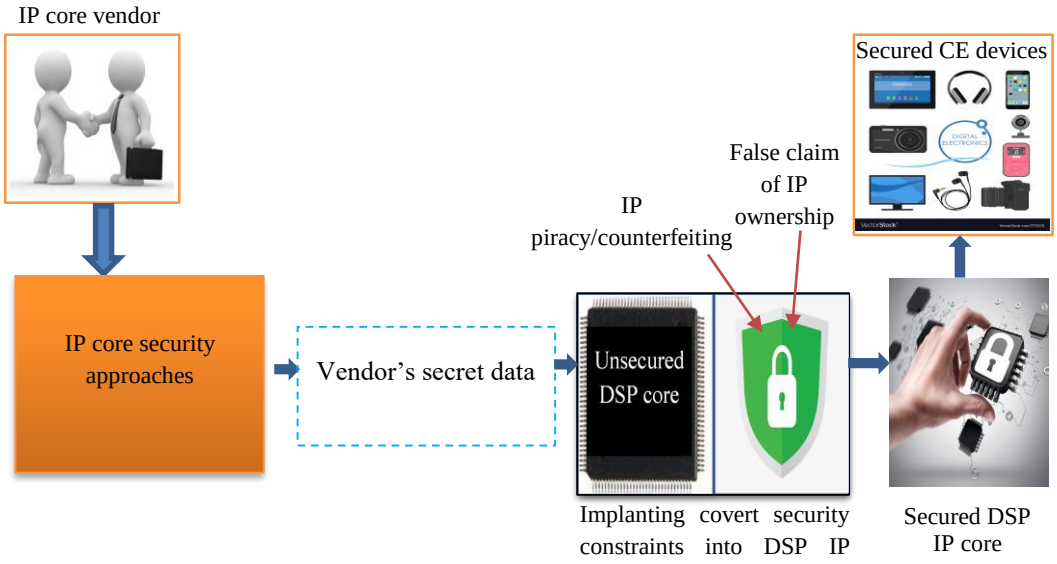


Fig. 1. Securing reusable DSP IP core used in CE systems

of IPs and blaming the user. Further, counterfeited IPs may consist malicious logic/ hardware trojan implanted through reverse engineering (that occurs by identifying the safe place to insert the malicious logic by any rogue element in the design house). Trojan can be inserted at any stage of IP design and is not easily detectable during testing phase of IP design or remains dormant until the happening of some specific triggering/ timing event. Counterfeited IPs may cause leakage of credential information/ passwords, drowning energy resources,

2162-2248 (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 March 07,2022 at 04:46:51 UTC from IEEE Xplore. Restrictions apply.

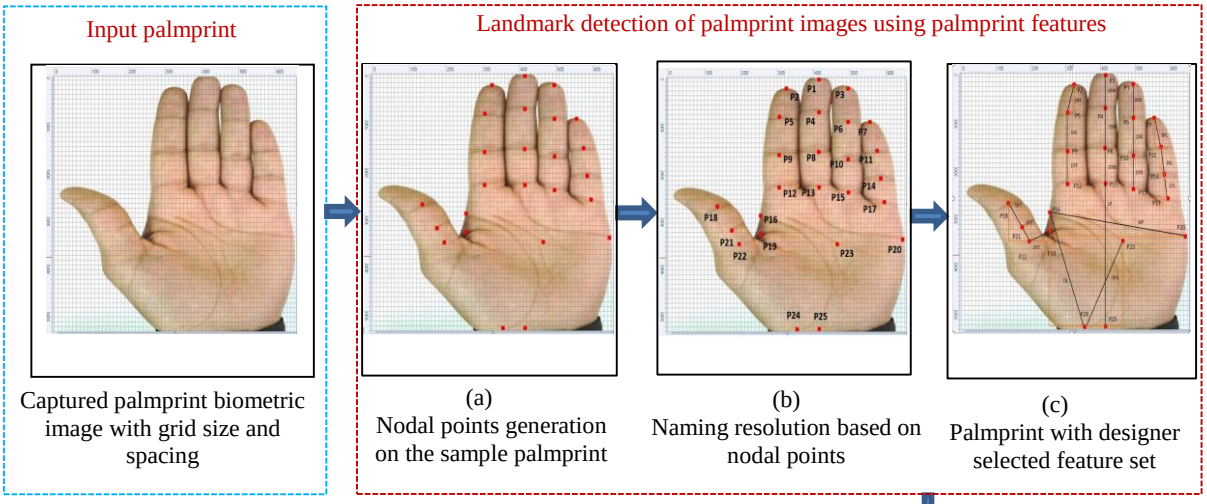
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCE.2022.3153276, IEEE Consumer Electronics Magazine

excessive heat dissipation of the IC components, and abnormal functioning or denial of service of the underling computing device [4]. Therefore, detective control of IP core from the SoC integrator's perspective must be mandatory. Fig. 1 shows a thematic representation of the targeted problem and its solution, in this article.

To countermeasure the above security threats, several security methodologies are proposed. They can be classified as authentication based (offers detective control against counterfeited versions of IP core) and obfuscation-based (offers preventive control against backdoor trojan insertion through reverse engineering) approaches. Authentication based approaches are mainly signature based approaches. Signature based approaches generates digital template to detect the counterfeited IP design and to award IP ownership to genuine IP vendor/ designer as well as helps in distinguishing between genuine and fake IP cores. Signature based approaches are IP watermarking [5, 6, 7] and encrypted-hash based digital signature [8]. Watermarking approaches can be further classified as regular watermarking, multilevel watermarking, dual/triple phase watermarking (exploiting behavioural level synthesis framework) and physical level watermarking. Moreover, biometric based approaches are facial biometric, fingerprint biometric and palmprint biometric based. Obfuscation based approaches obscures the design architecture (making the design architecture/ logic unobvious for an adversary to interpret and thus disabling the back door trojan insertion) but does not alter the true end functionality of the design. Obfuscation based approaches can be classified as structural obfuscation [9] and functional obfuscation-based approaches [10,11].

2. Exploiting Contact-less Palmprint Biometric for Securing Reusable DSP IP cores

Palmprint biometric based approach for securing reusable digital signal processing (DSP) IP core extracts the naturally unique palm features (biometric information of palm of IP designer/vendor) to generate palmprint template. Further, the security of reusable IP core is ensured by implanting the derived security constraints



Feature #	Feature name	Feature dimension	Binary representation
F1	DL	260	100000100
F2	DHL	210.95	11010010.11110011001100110011
F3	WP	393.19	110001001.00110000101000111101
F4	LP	300	100101100
F5	DFF	63.24	111111.00111101011100000101
F6	DSF	80	1010000
F7	DTF	70	1000110
F8	DFM	70	1000110
F9	DSM	80	1010000
F10	DTM	80	1010000
F11	DFR	70	1000110
F12	DSR	80	1010000
F13	DTR	70	1000110
F14	DFL	63.24	111111.00111101011100000101
F15	DSL	60.82	111100.11010001111010111
F16	DTL	50.99	110010.11111101011100000101
F17	DFT	64.03	1000000.00000111101011100001
F18	DST	36.05	100100.00001100110011001101
F19	DTT	63.24	111111.00111101011100000101

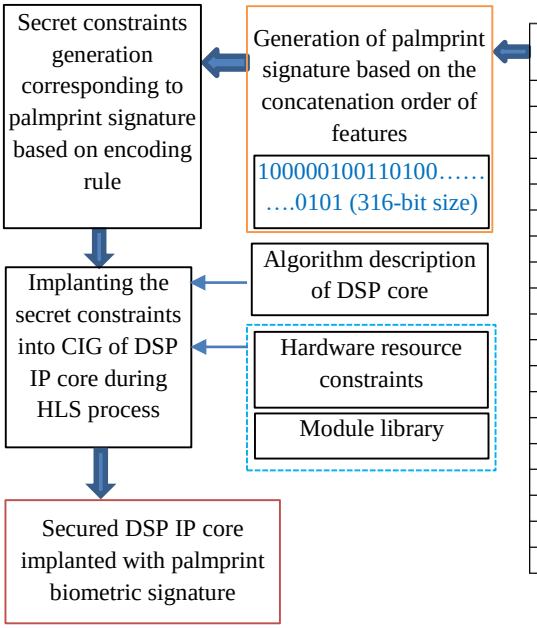


Fig. 2. Process of generation of secured reusable IP core using contact-less palmprint biometric based approach

2162-2248 (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 March 07,2022 at 04:46:51 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/MCE.2022.3153276, IEEE Consumer Electronics Magazine

(based on encoding rule devised by IP designer) into unsecured design. The process of generating secured reusable IP core through palmprint biometric is described in two modules: (i) palmprint template generation (ii) perform embedding of palmprint signature in order to obtain secured reusable IP core as shown below:

The process of palmprint template generation takes palmprint image captured using high resolution camera thus enabling contact-less palmprint template generation. The palmprint is placed with designer specified grid size and spacing thereby attenuating the effect of palm rotation and thus enabling precise nodal points generation as shown in Fig. 2 (a). Subsequently, naming conventions are made corresponding to each nodal point (P1 to P25) in order to palm feature generation as shown in Fig. 2 (b). Next, the nodal points are decided by IP designer based on the number of features and signature strength to be taken for palm signature generation by considering security design cost trade-off. Subsequently, palmprint image with designer selected feature set is generated (as shown in Fig. 2 (c). Next, feature dimensions are calculated corresponding to each feature selected by IP designer. Subsequently, in order to generate palm signature, feature dimensions are converted into binary. Finally, palmprint signature (palmprint template) is obtained by concatenating the binarize feature dimensions of each chosen feature based on the concatenation order decided by IP designer.

In the next module, initial phase generates covert secret constraints corresponding to palmprint signature strength based on the encoding rule devised by IP designer. Subsequently, constraints embedding phase takes DSP framework, resource constraints and module library as input and based on that scheduling, resource allocation and binding is performed. Consequently, scheduled data flow graph (DFG) of the DSP framework is obtained based on which colour interval graph is constructed during high level synthesis process. Next, derived secret constraints are embedded into colour interval graph (CIG) of corresponding DSP framework in order to obtain secured reusable IP core. Note: recapturing of palmprint is not required during verification/validation process in order to reproduce the palmprint for authentication purpose as stored palmprint with specified grid size/spacing and nodal points are used for the same. Finally, palmprint implanted secured reusable DSP IP core is obtained.

3. Exploiting Encrypted-Hash based Digital Signature Approach for Securing Reusable DSP IP cores

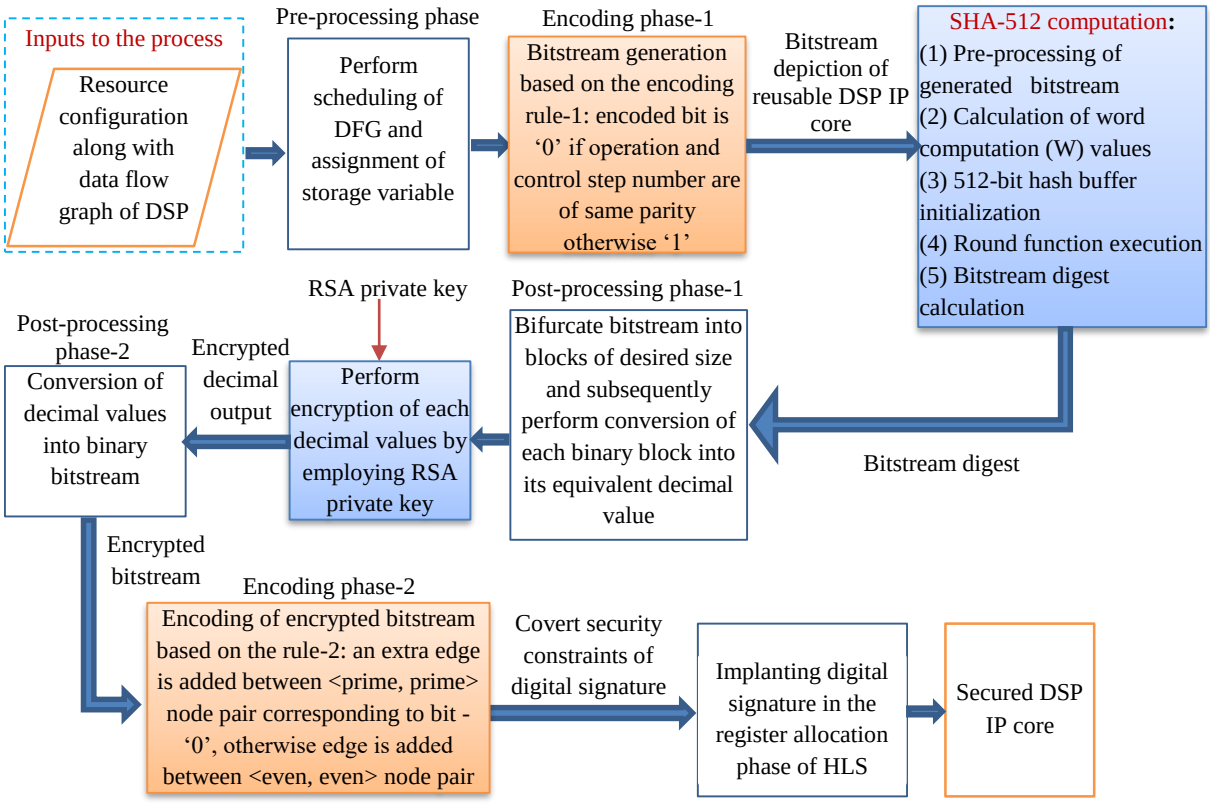


Fig. 3. Process of securing DSP IP core using encrypted-hash based digital signature

2162-2248 (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 March 07,2022 at 04:46:51 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/MCE.2022.3153276, IEEE Consumer Electronics Magazine

In order to secure the reusable IP core, encrypted-hash based digital signature approach takes DFG of the DSP application (in which the digital signature is to be embedded) and user specified resource constraints as primary input and based on which scheduling of DFG of the DSP core is performed. Subsequently, SDFG is fed as input to the encoding phase-1based on which bitstream is generated (using the encoding rule-1). Subsequently, the generated bitstream is fed into SHA-512, which generates the bitstream digest of corresponding DSP application as its output. The generation of bitstream digest involves word (W) computation process which employs the following functions: circular right shift of the 64-bit argument, left shift of the 64-bit argument and addition modulo 2^{64}. Next, in the post-processing phase-1, the generated binarize bitstream is bifurcated into desired blocks of equal size and has been converted into its equivalent decimal value. Subsequently, in the next phase encryption of each decimal value is performed using private key of the user (IP owner) through RSA encryption. Subsequently, encrypted data output is converted into binary bitstream during post processing phase-2. Thereafter, the encrypted bitstream is fed as input to the encoding phase-2 in order to generate covert security constraints corresponding to the digital signature strength (chosen by IP designer by considering the security and design cost trade-off). Please refer to Fig.3 for the diagrammatic representation. Subsequently, the covert security constraints are implanted during register allocation phase of HLS process. Further, the details of digital signature embedding process is shown in Fig.4. Moreover, in case of digital signature and key (decryption key) are compromised by an adversary, s/he cannot perform the exact regeneration of digital signature due to unknown concatenation point/order and secret encoding rules. Therefore, the encrypted-hash based digital

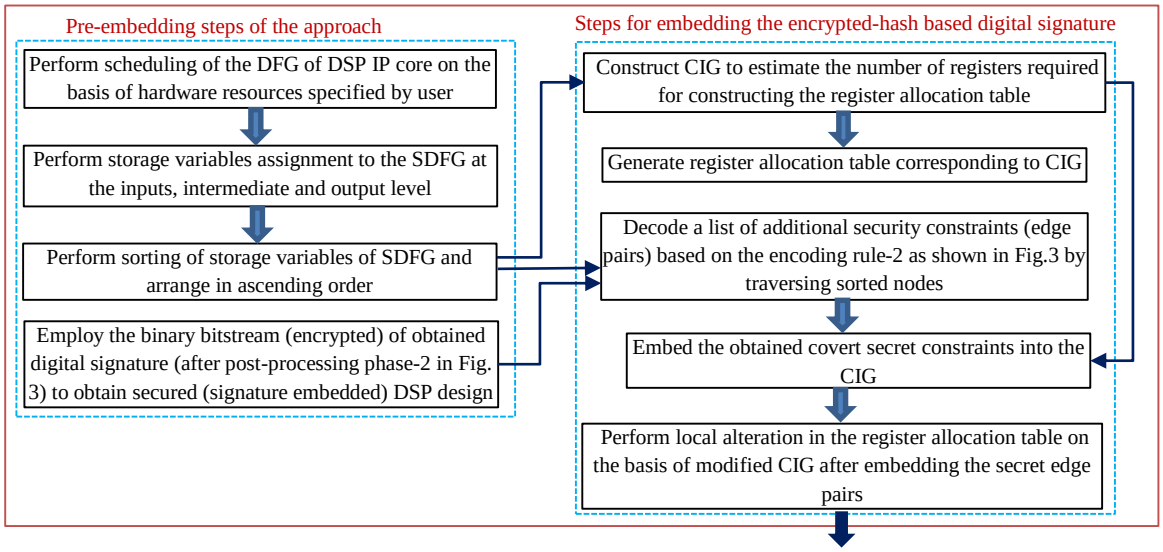


Fig. 4. Algorithm to embed encrypted-hash based digital signature into DSP IP core

TABLE I
Palmprint biometric Vs Encrypted-hash based digital signature approach for hardware security

Characteristics/ Parameters	**Palmprint biometric approach**	**Digital signature-based approach**	**Analysis of strength/ weaknesses**
Security approach	Palm features based on nodal point generation	Multilevel encoding, SHA-512 and RSA encryption	Implementation complexity in palmprint based approach is less than the encrypted hash based digital signature approach
Security strength	More unique than digital signature	Less unique than palmprint biometric	Due to naturally unique and non-replicable palm features based digital template than key based digital template generation
Digital template regeneration by an adversary	Impossible as dependent on nodal points, feature sets, grid size, feature orders etc.	Possible if encoding and RSA encryption key is compromised	Palmprint based approach is more robust than encrypted hash based digital signature approach as exact digital template regeneration is not possible without employing the same palmprint
Tamper tolerance capability	more	less	As the number of signature bits embedded are more in palmprint based approach than encrypted hash based digital signature approach
Counterfeited/fake IP core detection	Robust and seamless	More complex	Matching process of covert security constraints of DSP core in palmprint approach is much easier specifically for authentic vendor (having knowledge about the secret parameters)
Cost overhead of signature implanted design	Almost negligible	Higher	Due to signature encoding rule and palmprint based approach applies three types of encoding digits than two types of encoding digits in encrypted hash based digital signature approach
Dependence on secret keys	No	Yes	Palmprint based approach depends on the naturally unique palm feature generation for digital template than key based signature generation in encrypted hash approach

2162-2248 (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 March 07,2022 at 04:46:51 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/MCE.2022.3153276, IEEE Consumer Electronics Magazine

signature approach enables robust security against counterfeited IP cores through SHA-512 and RSA encryption. Finally, the digital signature embedded secured reusable DSP IP core is obtained which can be integrated into ICs to deploy secured CE systems.

4. Comparison between Palmprint Biometric and Encrypted-Hash based Digital Signature Approaches

Though, both the approaches such as palmprint biometric and encrypted-hash based digital signature are capable of providing robust security against reusable IP core threats, they have their own significance and applicability. The comparison based on several parameters such as implementation complexity, tamper tolerance capability, design cost and robustness etc.is shown in Table I. Therefore, the purpose of listing Table I is to analyse the strength and weaknesses of each security technique in terms of several characteristics and properties such as security approach in terms of security strength, digital template regeneration, tamper tolerance capability, counterfeited/fake IP core detection, cost overhead of signature implanted design and their dependence on secret keys.

5. Analysis and Inference

It is evident after analysing the palmprint biometric based approach and encrypted-hash based digital signature approach that both of these approaches are capable of ensuring the security of reusable DSP IP core against IP piracy, false claim of IP ownership and counterfeited design. Encrypted-hash based digital signature approach employs multi-level encoding and multi-level encryption through SHA-512 and RSA based encryption which results higher implementation complexity. On the other hand, palmprint based approach employs naturally unique palm features therefore preventing vulnerability and replicability of palm signature. Further, regeneration of exact palmprint is also impossible for an adversary as regeneration depends on a number of secret parameters such as nodal points, feature sets, feature ordering and specified grid size, all of which are unknown to an adversary. Unlike in the encrypted-hash based digital signature approach, regeneration of digital template is possible if encoding rules and RSA secret key are compromised. Though, both the approaches provide the detective control against counterfeited IP designs. However, palmprint based biometric offers stronger robustness. Fig. 5 and Fig. 6 shows the comparison of palmprint biometric approach and encrypted-hash based digital signature approach in terms of security metrics such as probability of coincidence and tamper tolerance ability. As evident from the experimental validation that palmprint biometric based approach achieves lower Pc and higher TT value due to larger signature strength causing more number of embedded encoded bits in the IP core design. This results into stronger digital evidence against IP piracy and fraudulent claim of IP ownership, than encrypted-hash based digital approach.

6. Conclusion

This article presented the following contributions for secured DSP core design in CE systems: (a) palmprint biometric signature-based technique Vs encrypted-hash based digital signature technique in terms of security parameters, characteristics and properties (b) complete demonstration of palmprint biometric based security technique from capturing input palmprint image to generating corresponding digital template to finally embedding the same in IP cores of CE systems (c) presents the flow methodology of encrypted-hash based digital signature approach for securing reusable IP cores. The palmprint biometric approach achieves number of times improvements in security and robustness over encrypted hash based digital signature approach in terms of

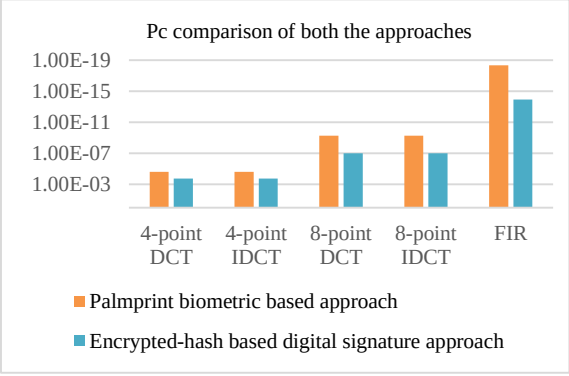


Fig. 5. Pc comparison corresponding to different DSP frameworks

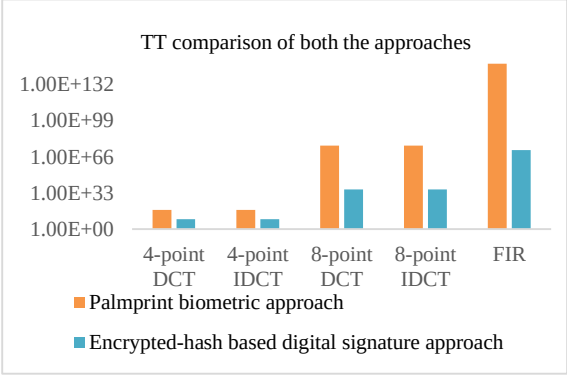


Fig. 6. TT comparison corresponding to different DSP frameworks

2162-2248 (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 March 07,2022 at 04:46:51 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/MCE.2022.3153276, IEEE Consumer Electronics Magazine

(i) Pc in the range of 10 to 10^5 and (ii) TT in the range of 10^8 to 10^{78}.

References

[1] M. Beaumont et. al "Hardware Trojans – Prevention, Detection, Countermeasures (A Literature Review)", Australian Government, Department of Defence, Defence Science and Technology Organization, DSTO-TN-1012.
[2] C. Pilato, S. Garg, K. Wu, R. Karri and F. Regazzoni, "Securing Hardware Accelerators: A New Challenge for High-Level Synthesis," in IEEE Embedded Systems Letters, vol. 10, no. 3, pp. 77-80, Sept. 2018.
[3] S. Rai, A. Rupani, P. Nath and A. Kumar, "Hardware Watermarking Using Polymorphic Inverter Designs Based On Reconfigurable Nanotechnologies," 2019 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), 2019, pp. 663-669.
[4] Y. Xie and A. Srivastava, "Anti-SAT: Mitigating SAT Attack on Logic Locking," in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 38, no. 2, pp. 199-207, Feb. 2019.
[5] Rizzo, S., Bertini, F. & Montesi, D. Fine-grain watermarking for intellectual property protection. EURASIP J. on Info. Security 2019, 10 (2019).
[6] F. Koushanfar, I. Hong, and M. Potkonjak, "Behavioral synthesis techniques for intellectual property protection," ACM Trans. Des. Autom. Electron. Syst., vol. 10, no. 3, pp. 523–545, July 2005.
[7] R. Karmakar and S. Chattopadhyay, "Hardware IP Protection Using Logic Encryption and Watermarking," 2020 IEEE International Test Conference (ITC), 2020, pp. 1-10.
[8] A. Sengupta, E. R. Kumar and N. P. Chandra, "Embedding Digital Signature Using Encrypted-Hashing for Protection of DSP Cores in CE," in IEEE Transactions on Consumer Electronics, vol. 65, no. 3, pp. 398-407, Aug. 2019.
[9] Y. Lao and K. K. Parhi, "Obfuscating DSP Circuits via High-Level Transformations," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 23, no. 5, pp. 819-830, May 2015.
[10] M. Yasin, J. J. Rajendran, O. Sinanoglu and R. Karri, "On Improving the Security of Logic Locking," in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 35, no. 9, pp. 1411-1424, Sept. 2016.
[11] J. Zhang, "A Practical Logic Obfuscation Technique for Hardware Security," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 24, no. 3, pp. 1193-1197, March 2016.

Biography of the author:

Rahul Chaurasia received the MTech. degree in computer science engineering from SOIT, RGPV Bhopal, India, in 2018. He is currently pursuing Ph.D. in CSE at Indian Institute of Technology (I.I.T) Indore.

Aditya Anshul received the M.Tech degree in CSE from National Institute of Technology, Sikkim in 2021. He is currently pursuing Ph.D. in CSE at Indian Institute of Technology (I.I.T) Indore.

Dr Anirban Sengupta received the Ph.D. degree in Electrical and Computer Engineering from Ryerson University, Toronto, Canada. He is an Associate Professor in CSE at Indian Institute of Technology Indore and author of over 245 publications including 4 books and 11 books. He is Fellow of IET (UK) and Fellow of BCS (UK).

Dr. Shikha Gupta received her Ph.D. from Delhi Technological University (DTU) in Delhi. She is currently a Visiting Researcher in CSE in "Security and Reliability of CE systems group" at Indian Institute of Technology (IIT) Indore and Visiting Faculty at Indian Institute of Management (IIM) Indore.

2162-2248 (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 March 07,2022 at 04:46:51 UTC from IEEE Xplore. Restrictions apply.

Journal of Engg. Research ICAPIE Special Issue pp. 198-207 **DOI:10.36909/jer.ICAPIE.15037**

Prioritizing Scheduling Parameters in the Automotive Industry Using Fuzzy TOPSIS-DEMATEL Model

Sumit Chawla* and Ranganath M. Singari

Department of Mechanical, Production and Industrial Engineering, Delhi Technological University
**Corresponding Author: chawlasumit4@gmail.com*

ABSTRACT

The Automotive industry is one of the biggest emerging sectors in terms of revenue. Every automotive industry has an indispensable need for optimum manufacturing scheduling systems for generating good revenues and profits. This need can be pulled off by identifying and prioritizing the scheduling parameters. Multi-criteria decision making (MCDM) is one of the best techniques of operation research in selecting the best parameters or factors among the various alternatives. This study includes the identification and prioritization of the various important scheduling parameters (SPs) in the Indian automotive industry. The twelve scheduling parameters have been identified in this study and these parameters are prioritized by the fuzzy- Technique for order preference by similarity to ideal solution (TOPSIS) and Decision Making Trial and Evaluation Laboratory Model (DEMATEL). The expert's views are gathered from the five automobile industries. Makespan, energy consumption, due date, and travel time are the crucial parameters obtained using Fuzzy TOPSIS. The least important parameters obtained using Fuzzy TOPSIS are work in process, flow time, and release date. The most influential parameters identified using the DEMATEL method are completion time and processing time. This study is valuable for every industry and research organization in the field of the automobile industry.

Keywords: Scheduling; Automotive industry; Fuzzy TOPSIS; Ideal solutions; DEMATEL.

INTRODUCTION

In today's scenario, the desideratum of every manufacturing firm is scheduling. Scheduling means the fulfilling the different performance criteria or factors by distributing the available production resources over time (Akyol et al., 2007). Production Scheduling is a very important decision-making process that includes the proper allocation of all the available resources for performing all tasks (R. Agrawal et al., 2012). On-time delivery of products or services provides customer satisfaction and scheduling helps in achieving on-time delivery (M. Pinedo, 1996). The primary objective of scheduling includes determining the job processing time, sequence, and due date of jobs (Tavakkoli-Moghaddam et al., 2010). The complexity of each production scheduling problem depends on different objectives, environmental conditions, and process constraints. The manufacturing schedule depends not only on production but also depends on scheduling parameters (Coudert et al., 2002). Prioritizing scheduling parameters is the need of every industry for optimum schedule because, without an optimum schedule, industries can't increase the profit and productivity on a full scale. Most of the industries only work in the area of improving the work culture, surrounding conditions, and performance of workers and machines, etc. But for every industry, it is important to identify and prioritize the critical scheduling parameters or most influential scheduling parameters of their respective area or field. By identifying the critical scheduling parameters, we can generate the most effective scheduling system. So, there is an immense need of prioritizing these parameters. These parameters can be effectively prioritized using the MCDM approach. This study includes the identification of scheduling parameters and finding the criticality of these parameters in the automobile industry.

Due to uncertainty and ambiguity in human judgment decisions, all values corresponding to scheduling parameters in the decision set can't be crisp. So, some linguistic variable or fuzzy variable must be taken to deals with all the criteria (Zhao et al., 2014). Therefore, fuzzy TOPSIS methodology is used in this study for prioritizing the SPs for an optimum schedule in the automotive industry, and the DEMATEL method is used for identifying the most influential parameters and also finds cause criteria and effect criteria group. DEMATEL is the structural modeling approach used in finding the relationship between the cause and effect criteria. Many researchers have used fuzzy TOPSIS and DEMATEL methodology separately in various scheduling problems as discussed in the literature part. But the first time, both these approaches are applied simultaneously in prioritizing the scheduling parameters and identify the most influential parameters for an optimum schedule in the automotive industry.

MCDM is a branch of operation research in dealing with complex multi-criteria or specifications. This technique helps in obtaining the best choice among the various alternatives in these fluctuating real-world problems. The fuzzy TOPSIS technique is one of the best techniques of MCDM used by various researchers in past studies in dealing with uncertainty (Chawla et al., 2019 & Yong, 2006). This method has been applied in reverse logistics (S. Agrawal et al., 2016), material selection (Maity et al., 2013), project selection (Taylan et al., 2014), warehouse location selection (Ashrafzadeh, 2012), plant location selection (Yong, 2006), supplier selection (Daneshvar et al., 2014), machine tool selection (Yurdakul et al., 2009) and wireless network selection (Mansouri et al., 2020). The fuzzy TOPSIS technique can be effectively used for prioritizing the critical factors for implementing the reverse logistics (S. Agrawal et al., 2016). S. Maity et al. solved the grinding material selection problem using Fuzzy TOPSIS (Maity et al., 2013). Construction projects are dynamic in nature and consist of high risks and uncertainty. Their high risks can be effectively evaluated by fuzzy TOPSIS (Taylan et al., 2014). The warehouse location selection problem also consists of uncertainty and vagueness. M. Ashrafzadeh et al. utilized this approach for solving the warehouse location problem (Ashrafzadeh, 2012).

DEMATEL method is another effective approach that can be applied for prioritizing the critical parameters or risk factors in the automobile spare industry (Wu et al., 2011), project management (Zhang et al., 2019), waste management (Chauhan et al., 2018), maintenance management (Vujanovic et al., 2012), product design (Ibrahim et al., 2018), remanufacturing (Singh et al., 2018), service quality improvement (Chu et al., 2017) and supplier selection (Chiou et al., 2011). Hsin-Hung Wu et al. identified the technological capability, organization, and service as the three most critical dimensions in the automobile spare industry using the DEMATEL method. They identified the five most influential criteria among thirty criteria for improvement (Wu et al., 2011). Lin Zhang et al. identified the critical risk factors in urban projects for controlling the flooding and water shortage using the DEMATEL method (Zhang et al., 2019). Ankur Chauhan et al. identified and prioritized the barriers of waste recycling management using this methodology (Chauhan et al., 2018). Davor Vujanovic et al. solved the vehicle fleet management problem and evaluated its indicators by ANP (Analytic Network Process) and DEMATEL method (Vujanovic et al., 2012). Jianjie Chu et al. enhanced the air travel service quality by prioritizing the key service criteria using DEMATEL and gray theory (Chu et al., 2017). The above studies show that scheduling parameter prioritization in the automobile industry is an untouched area of research and the Fuzzy TOPSIS and DEMATEL methods are effective in the identification and prioritization of the critical parameters or risk factors of various industries problems.

IDENTIFICATION OF SCHEDULING PARAMETERS

Twelve Scheduling parameters are identified from previous studies and industry expert's reviews. These parameters are discussed below. Makespan is one of the strongest performance measures in all types of scheduling problems and it represents the total time to process all the jobs (Pinedo, 1996). For minimizing the makespan, machine speed can be increased but it increases energy consumption. The tradeoff is required between these parameters (Gong et al., 2015). Makespan or tardiness is taken as the main parameter for a single optimality criteria problem. Flow time represents the size of the average inventory. Flow time can be significantly reduced by minimizing the average inventory. Lateness indicates the condition of completing the orders near the due date. Processing time depends on the starting time in machine scheduling problems. Processing time can be linear or exponential in time-dependent problems (Cheng et al., 2004).The release date shows the value before which a job cannot be processed on a machineA job must be entered at the

date and leaves at the due date. Earliness represents the negative lateness and shows the condition of completing the orders earlier to the due date (Alharkan, 2005).

FUZZY TOPSIS METHOD FOR PRIORITIZING SCHEDULING PARAMETERS

TOPSIS technique was the first time applied in the fuzzy environment for the group decision making in 1997 by Chen –Tung Chen. In fuzzy TOPSIS methodology, the linguistic variable is used with a fuzzy number on a point scale (Chen, 2000). The methodology is shown as follows:

Step 1: Fuzzy TOPSIS starts with the establishment of a committee of decision-makers.

Step 2: Define Linguistic Variable terms with their Fuzzy Number on the point scale. This scale is defined in Table 1 using the linguistic terms which are negative low (NLO), low (LO), average or arithmetic mean (AM), high (HI), and positive high (PHI).

Table 1. Linguistic Variable terms with their Fuzzy Number

Linguistic term	Fuzzy number
Negative low (NLO)	1,1,3
Low (LO)	1,3,5
Average (AM)	3,5,7
High (HI)	5,7,9
Positive high (PHI)	7,9,9

Step 3: Find the decision matrix (DM) = $[x_{ij}]_{mxn}$, $x_{ij} = (a_{ij},b_{ij},c_{ij})$ which is a fuzzy number with i =1,2,...,m number of decision maker of various automobile companies and j = 1,2,...,n number of scheduling parameters.

Step 4: Find the Normalized fuzzy decision matrix R_{ij}

Equation 1 is showing a normalized decision matrix for non-beneficial criteria. For beneficial criteria, larger R_{ij} is desirable, whereas, for cost criteria, smaller R_{ij} is desirable.

$$\text{NDM} = [R_{ij}]_{mxn}, \; R_{ij} = \left(\frac{C^*_j}{C^*_{ij}}, \frac{C^*_j}{b^*_{ij}}, \frac{C^*_j}{a^*_{ij}} \right), \; C^*_j = \min_i c \qquad (1)$$

Step 5: Determine the weightage normalized fuzzy decision matrix *V* which is obtained by multiplying the weightage w_j given to the decision makers with matrix R_{ij}.

$$V = v_{ij} = w_j \text{ x } R_{ij} \qquad (2)$$

Step 6: Calculate the Fuzzy positive from the A^+ and s^-_i separation from the A^-

$$A^+ = \{v_1^+, v_2^+ ...., v_n^+\}, \qquad A^- = \{v_1^-, v_2^- ...., v_n^-\} \qquad (3)$$

Step 7: Calculate the separation from the A^+ and s^-_i separation from the A^-

$$S_i^+ = \sqrt{\sum_{j=1}^{n} (v^+_j - v_{ij})^2}, \; s_i^- = \sqrt{\sum_{j=1}^{n} (v^-_j - v_{ij})^2}, \qquad i=1,2,...,m \qquad (4)$$

Step 8: Calculate the relative closeness coefficient C_j

$$C_j = \frac{s^+_i}{s^+_i + s^-_i} \qquad \begin{matrix} C_j = 1 \text{ if } A_i = A^+ \\ C_j = 0 \text{ if } A_i = A^- \end{matrix} \qquad (5)$$

$s^-_i \geq 0$ and s+i ≥ 0, then clearly $0 \leq C_j \leq 1$

Step 9: The ranking of alternatives is obtained using relative closeness values.

APPLICATION OF FUZZY TOPSIS METHODOLOGY

Literature review and experts view from various decision-making companies help us in identifying the main twelve scheduling parameters for the optimum schedule in automotive part manufacturing companies. The twelve scheduling parameters identified are makespan, flow time, lateness, processing time, due date, energy consumption, earliness, travel time, work in process, tardiness, completion time, and release date. Profiles of the various crankcase automotive part manufacturing companies are illustrated in Table 2.

Table 2. Profile of various Decision maker companies

Companies	**Product Manufactured**	**Material**	**Annual Turnover (USD)**	**No. of Branches in India**	**Manufacturing Facility Setup in India**
Super Auto India limited (CPY 1)	Motorcycle crankcase	Aluminum	70 Lakh	3	Faridabad, Pune
Shiv Shakti Engineering Co. (CPY 2)	Ingersoll Rand Crank Case	CIFC225	1.4 Lakh	1	Ahmedabad (Gujarat, India)
Kolben Compressor Spares (India) Private Limited (CPY 3)	Vilter 440 Crankcase	Steel	35 Lakh	1	Churchgate, Mumbai, Maharashtra
Industrial Spare Syndicate Limited (CPY 4)	ISS Crank Case For KG2	Cast Iron	4.9 Lakh	1	Mori Gate, Delhi
Shanirajeshwar Die Casting Pvt. Ltd. (CPY 5)	Black Automobile Crankcase	Aluminum Alloy	24.5 Lakh	1	Moshi, Pune, Maharashtra

Scheduling parameters for optimum schedule with their symbols are represented as makespan (Mk), flow time (Ft), lateness (Lt), processing time (Pt), due date (Dd), energy consumption (Ec), earliness (E), travel time (Tt), work in process (Wp), tardiness (T), completion time (Ct), and release date (Rd). Table 3 is showing the linguistic variable as well as the fuzzy number-based decision matrix. A linguistic variable-based matrix is developed on the five-point linguistic scale as discussed in Table 1.

Table 3. Linguistic Variable & Fuzzy number Based Decision Matrix

Linguistic Variable Based Decision Matrix						Fuzzy number based Decision matrix				
	CPY 1	**CPY 2**	**CPY 3**	**CPY 4**	**CPY 5**	**CPY 1**	**CPY 2**	**CPY 3**	**CPY 4**	**CPY 5**
$\mathbf{M_k}$	HI	HI	PHI	HI	PHI	5,7,9	5,7,9	7,9,9	5,7,9	7,9,9
$\mathbf{F_t}$	NLO	AM	LO	NLO	LO	1,1,3	3,5,7	1,3,5	1,1,3	1,3,5
$\mathbf{L_t}$	LO	AM	LO	LO	AM	1,3,5	3,5,7	1,3,5	1,3,5	3,5,7
$\mathbf{P_t}$	HI	HI	AM	HI	AM	5,7,9	5,7,9	3,5,7	5,7,9	3,5,7
$\mathbf{D_d}$	HI	PHI	AM	HI	AM	5,7,9	7,9,9	3,5,7	5,7,9	3,5,7
$\mathbf{E_c}$	HI	AM	AM	LO	AM	5,7,9	3,5,7	3,5,7	1,3,5	3,5,7
E	LO	AM	LO	LO	LO	1,3,5	3,5,7	1,3,5	1,3,5	1,3,5
$\mathbf{T_t}$	AM	HI	AM	HI	AM	3,5,7	5,7,9	3,5,7	5,7,9	3,5,7
$\mathbf{W_p}$	NLO	AM	NLO	NLO	LO	1,1,3	3,5,7	1,1,3	1,1,3	1,3,5
T	HI	AM	AM	HI	AM	5,7,9	3,5,7	3,5,7	5,7,9	3,5,7
$\mathbf{C_t}$	LO	AM	AM	LO	AM	1,3,5	3,5,7	3,5,7	1,3,5	3,5,7
$\mathbf{R_d}$	LO	AM	LO	NLO	LO	1,3,5	3,5,7	1,3,5	1,1,3	1,3,5

For calculating the normalized fuzzy decision matrix, we have assumed all criteria to be non-beneficial (cost) criteria. The matrix will be the same as the matrix because, for this study, equal weights are considered for all the decision-makers. Table 4 shows the matrix with the fuzzy ideal solutions. The distances or separation from the ideal solutions are calculated in Table 5.

Table 4. Matrix *V* with ideal solutions

	CPY 1	**CPY 2**	**CPY 3**	**CPY 4**	**CPY 5**
$\mathbf{M_k}$	0.11,0.14,0.2	0.33,0.42,0.6	0.11,0.11,0.14	0.11,0.14,0.2	0.11,0.11,0.14
$\mathbf{F_t}$	0.33,1,1	0.42,0.6,1	0.2,0.33,1	0.33,1,1	0.2,0.33,1
$\mathbf{L_t}$	0.2,0.33,1	0.42,0.6,1	0.2,0.33,1	0.2,0.33,1	0.14,0.2,0.33
$\mathbf{P_t}$	0.11,0.14,0.2	0.33,0.42,0.6	0.14,0.2,0.33	0.11,0.14,0.2	0.14,0.2,0.33
$\mathbf{D_d}$	0.11,0.14,0.2	0.33,0.33,0.42	0.14,0.2,0.33	0.11,0.14,0.2	0.14,0.2,0.33
$\mathbf{E_c}$	0.11,0.14,0.2	0.42,0.6,1	0.14,0.2,0.33	0.2,0.33,1	0.14,0.2,0.33
E	0.2,0.33,1	0.42,0.6,1	0.2,0.33,1	0.2,0.33,1	0.2,0.33,1
$\mathbf{T_t}$	0.14,0.2,0.33	0.33,0.42,0.6	0.14,0.2,0.33	0.11,0.14,0.2	0.14,0.2,0.33
$\mathbf{W_p}$	0.33,1,1	0.42,0.6,1	0.33,1,1	0.33,1,1	0.2,0.33,1
T	0.11,0.14,0.2	0.42,0.6,1	0.14,0.2,0.33	0.11,0.14,0.2	0.14,0.2,0.33
$\mathbf{C_t}$	0.2,0.33,1	0.42,0.6,1	0.14,0.2,0.33	0.2,0.33,1	0.14,0.2,0.33
$\mathbf{R_d}$	0.2,0.33,1	0.42,0.6,1	0.2,0.33,1	0.33,1,1	0.2,0.33,1
$\mathbf{A^+}$	0.33,1,1	0.42,0.6,1	0.33,1,1	0.33,1,1	0.2,0.33,1
$\mathbf{A^-}$	0.11,0.14,0.2	0.33,0.33,0.42	0.11,0.11,0.14	0.11,0.14,0.2	0.11,0.11,0.14

Table 5. Separation from each parameter to the FPIS & FNIS

	Separation from each parameter to the FPIS					Separation from each parameter to the FNIS				
	CPY 1	**CPY 2**	**CPY 3**	**CPY 4**	**CPY 5**	**CPY 1**	**CPY 2**	**CPY 3**	**CPY 4**	**CPY 5**
M_k	0.689	0.258	0.725	0.689	0.515	0	0.116	0	0	0
F_t	0	0	0.394	0	0	0.689	0.373	0.515	0.689	0.515
L_t	0.394	0	0.394	0.394	0.395	0.477	0.373	0.515	0.477	0.122
P_t	0.689	0.258	0.612	0.689	0.395	0	0.116	0.122	0	0.122
D_d	0.689	0.373	0.612	0.689	0.395	0	0	0.122	0	0.122
E_c	0.689	0	0.612	0.394	0.395	0	0.373	0.122	0.477	0.122
E	0.394	0	0.394	0.394	0	0.477	0.373	0.515	0.477	0.515
T_t	0.612	0.258	0.612	0.689	0.395	0	0.116	0.122	0	0.122
W_p	0	0	0	0	0	0.689	0.373	0.725	0.689	0.515
T	0.689	0	0.612	0.689	0.395	0	0.373	0.122	0	0.122
C_t	0.394	0	0.612	0.394	0.395	0.477	0.373	0.122	0.477	0.122
R_d	0.394	0	0.394	0	0	0.477	0.373	0.515	0.689	0.515

S^+, S^-, and C_j values are calculated using Equations 6 and 7. Based on Cj values, priority values are given. Table 6 shows the Priority Matrix Based on Closeness Coefficient Values. The most important SPs obtained is makespan and the least important SPs is work in process.

Table 6. Priority Matrix Based on Closeness Coefficient Values

	M_k	**F_t**	**L_t**	**P_t**	**D_d**	**E_c**	**E**	**T_t**	**W_p**	**T**	**C_t**	**R_d**
S^+	2.87	0.39	1.57	2.64	2.75	2.09	1.18	2.56	0	2.38	1.79	0.78
S^-	0.11	2.78	1.96	0.36	0.24	1.09	2.35	0.36	2.99	0.61	1.57	2.56
C_j	0.96	0.12	0.44	0.88	0.91	0.65	0.33	0.87	0	0.79	0.53	0.23

DEMATEL METHODOLOGY

DEMATEL methodology is based on the causal relationship and provides visual graphical relations between the complex criteria. This method also identifies the cause and effect groups among the various criteria. It is the only method that provides visualization to researchers about interrelations among criteria. For providing visualization, it uses the basics of graph theory (Ibrahim et al., 2018). The steps involved within the DEMATEL methodology are as follows.

Step 1: Set up a pairwise comparison scale of the DEMATEL method which includes the terms with numeral like no effect (0), low effect (1), medium effect (2), high effect (3), and very high effect (4).

Step 2: Compute the initial direct relation matrix which is a matrix generated from a comparison scale in terms of numerical and it shows the influences of all criteria on each another. is a square matrix in which diagonal elements are zero.

Step 3: Calculate the Normalized direct relation matrix U i.e. $U = [U_{ij}]_{nxm}$ and $0 \leq U_{ij} \leq 1$. This matrix can be determined from the Equation 6. Set of elements contains in a system is

$$S = \{s_1, s_2, ..., s_n\}.$$

$$S = \frac{1}{\max_{1 \le i \le n} \sum_{j=1}^{n} x_{ij}}, \quad U = S \times M \tag{6}$$

Step 4: Determine the total relation matrix *V* using Equations 7 and 8 in which identity matrix is denoted by 1 .

$$V = M + M^2 + ... + M^h = M(I - M^h)(I - M)^{-1} \tag{7}$$

when $\lim_{h \to \infty} M^h = [0]_{nxn}$,

$$V = (v_{ij})_{nxn} = M(I - M)^{-1} \tag{8}$$

Step 5: Defining the sum of rows (D) and sum of column(R) in the total relation matrix *V* as vector d and r through Equations 9.

$$D = (d_i)_{n \times 1} = \left[\sum_{j=1}^{n} v_{ij}\right]_{n \times 1}, \quad R = (r_i)_{1 \times n} = \left[\sum_{i=} v_{ij}\right]_{1 \times n} \tag{9}$$

Step 6: With the help of Equation 9, cause and effect diagram is generated to find the most affected criteria/parameter.

APPLICATION OF DEMATEL METHODOLOGY

The influence of each criterion on all other criteria is represented by the initial direct relation matrix in terms of numerical value as shown in Table 7. Data for Table 7 is collected from the literature review and industry experts. A total relation matrix is generated by using Equations 7, 8, and 9 as shown in Table 8. The positive and negative values of D-R represent the cause and effect criteria respectively as depicted in Table 9.

Estimate the threshold value (α=0.2689) of the total relation matrix. Then, compare all the matrix values with it and marked bold those values whose value is greater than threshold values. In the DEMATEL method, values with the highest prominence vector and relation vector are the most influential criteria. Completion time (C11) and processing time (C4) are the most influential criteria identified from the causal diagram as shown in Figure 1.

Table 7. Initial direct relation matrix (M)

	M_k	F_t	L_t	P_t	D_d	E_c	E	T_t	W_p	T	C_t	R_d
M_k	0	4	3	0	3	3	3	1	2	3	4	1
F_t	4	0	3	0	3	3	3	1	3	3	3	0
L_t	1	1	0	0	3	0	4	0	0	2	3	2
P_t	4	4	3	0	1	3	3	0	1	3	4	0
D_d	1	1	3	1	0	0	4	3	0	1	3	4
E_c	3	3	0	4	0	0	0	2	0	1	4	0
E	1	1	2	0	1	0	0	0	1	4	3	2
T_t	4	4	3	0	1	2	3	0	1	2	3	0
W_p	3	3	2	1	0	3	1	4	0	3	3	0
T	1	1	1	0	1	0	4	0	0	0	3	2
C_t	4	4	3	3	1	1	2	1	3	3	0	1
R_d	0	0	3	0	4	1	3	0	0	3	3	0

Table 8. Total relation matrix

	M_k	F_t	L_t	P_t	D_d	E_c	E	T_t	W_p	T	C_t	R_d
M_k	**0.317**	**0.446**	**0.427**	0.126	**0.344**	**0.279**	**0.489**	0.180	0.239	**0.465**	**0.568**	0.223
F_t	**0.439**	**0.310**	**0.415**	0.121	**0.333**	**0.277**	**0.475**	0.181	0.265	**0.453**	**0.526**	0.184
L_t	0.198	0.198	0.191	0.064	0.247	0.087	**0.365**	0.075	0.093	**0.281**	**0.345**	0.192
P_t	**0.456**	**0.456**	**0.425**	0.124	**0.278**	**0.288**	**0.485**	0.138	0.212	**0.468**	**0.571**	0.181
D_d	0.261	0.261	**0.357**	0.114	0.195	0.125	**0.437**	0.195	0.123	**0.318**	**0.427**	**0.285**
E_c	**0.368**	**0.368**	0.257	0.235	0.183	0.158	**0.292**	0.172	0.146	**0.314**	**0.467**	0.122
E	0.193	0.193	0.244	0.060	0.171	0.087	0.217	0.070	0.122	**0.333**	**0.331**	0.177
T_t	**0.419**	**0.419**	**0.392**	0.105	0.255	0.233	**0.446**	0.123	0.194	**0.395**	**0.489**	0.163
W_p	**0.407**	**0.407**	**0.363**	0.147	0.217	**0.279**	**0.383**	0.265	0.160	**0.430**	**0.499**	0.152
T	0.173	0.173	0.194	0.053	0.156	0.075	**0.322**	0.059	0.082	0.181	**0.304**	0.166
C_t	**0.456**	**0.456**	**0.432**	0.215	**0.283**	0.230	**0.462**	0.176	**0.276**	**0.472**	**0.442**	0.213
R_d	0.166	0.166	**0.291**	0.069	**0.279**	0.114	**0.338**	0.077	0.087	**0.310**	**0.349**	0.130

Table 9. Calculation of prominence vector and relation vector

	Diagram Notations	D	R	D+R	D-R
M_k	C1	4.1076	3.8578	7.9654	0.2498
F_t	C2	3.9853	3.8578	7.8431	0.1275
L_t	C3	2.3418	3.9952	6.3370	-1.6534
P_t	C4	4.0887	1.4397	5.5284	2.6490
D_d	C5	3.1026	2.9455	6.0481	0.1571
E_c	C6	3.0865	2.2382	5.3247	0.8483
E	C7	2.2034	4.7154	6.9188	-2.5120
T_t	C8	3.6387	1.7174	5.3561	1.9213
W_p	C9	3.7134	2.0048	5.7182	1.7086
T	C10	1.9436	4.4260	6.3696	-2.4824
C_t	C11	4.1204	5.3226	9.4430	-1.2022
R_d	C12	2.3824	2.1939	4.5763	0.1885

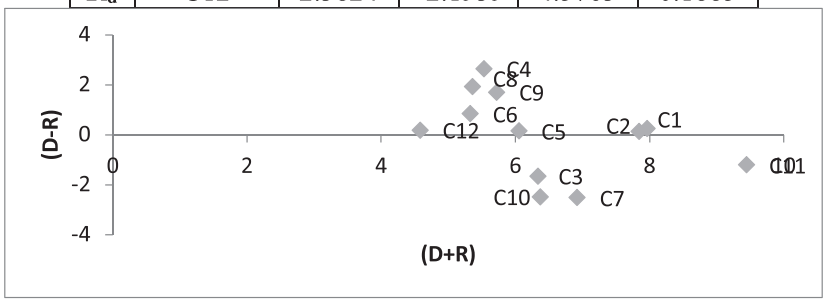


Figure 1. Causal diagram

CONCLUSION

Optimum Scheduling plays a major role in the effective manufacturing system. This scheduling can only be done by identifying the scheduling parameters. This study provides the methodology of identifying and prioritizing the twelve SPs for automotive part manufacturing. The most important four SPs identified for optimum scheduling using fuzzy TOPSIS are makespan, due date, processing time, and travel time. The least important SPs identified using fuzzy TOPSIS are work in process, flow time, and release date. Makespan (C1), flow time (C2), processing time (C4), due date (C5), energy consumption (C6), travel time (C8), work in process (C9), and release date (C12) are classified into cause criteria group, whereas effect criteria group consists of lateness (C3), earliness (C7), tardiness (C10)

and completion time (C11). Cause criteria factors are more crucial than the effect criteria factors. So, all industries and research organizations must give more attention to these factors because cause criteria group improvement has a significant effect on the improvement of the effect criteria group. Results reveal that the completion time and processing time are the most influential criteria in optimum scheduling. This study is useful for all automotive part manufacturers as well as automobile-based research organizations. This study is based on the automobile crankcase cover with five point linguistic scale.

For further research, the Fuzzy VIKOR, and fuzzy PROMETHEE can be applied and comparison analysis can be done for the same problem. Future research can be carried out by taking any other automotive part with five or more linguistic scales and for a more generalized scheduling model, the number of experts can also be increased.

REFERENCES

Agrawal, R., Pattanaik, L.N., & Kumar, S. 2012. Scheduling of a flexible job-shop using a multi-objective genetic algorithm. Journal of Advances in Management Research 9(2):178–188.

Agrawal, S., Singh, R.K., & Murtaza, Q. 2016. Prioritizing critical success factors for reverse logistics implementation using fuzzy-TOPSIS methodology. Journal of Industrial Engineering International 12: 15–27.

Akyol, D.E., & Bayhan, G.M. 2007. A review on evolution of production scheduling with neural networks. Computers & Industrial Engineering 53: 95–122.

Alharkan, I.M. 2005. Algorithms for sequencing and scheduling. Riyadha: Industrial Engineering Department, College of Engineering, King Saud University.

Ashrafzadeh, M. 2012. Application of Fuzzy TOPSIS method for the selection of warehouse location: a case study. Interdisciplinary Journal of Contemporary Research in Business 3(9): 655–671.

Chauhan, A., Singh, A. & Jharkharia, S. 2018. An interpretive structural modeling (ISM) and decision-making trial and evaluation laboratory (DEMATEL) method approach for the analysis of barriers of waste recycling in India. Journal of the Air & Waste Management Association 68(2): 100–110.

Chawla, S., Agrawal, S. & Singari, R.M. 2019. Integrated topsis-moora model for prioritization of new bike selection. In Prasad A., Gupta S., Tyagi R. (Eds) Advances in Engineering Design. Lecture Notes in Mechanical Engineering. Springer, Singapore, , 755–765.

Chen, C. 2000. Extensions of the TOPSIS for group decision-making under Fuzzy Environment. Fuzzy Sets and Systems 114: 1–9.

Cheng, T.C..E, Ding, Q. & Lin, B.M.T. 2004. A concise survey of scheduling with time-dependent processing times. European Journal of Operational Research 152: 1–13.

Chiou, C.Y., Hsu, C.W. & Chen, H.C. 2011. Using DEMATEL to explore a casual and effect model of sustainable supplier selection. 2011 IEEE International Summer Conference of Asia Pacific Business Innovation and Technology Management, Dalian: 240–244.

Chu, J., Li, W., Yang, H. & Yu, S. 2017. An approach for prioritizing key factors affecting air travel experience based on grey theory and DEMATEL. In Proceedings of the 23rd International Conference on Automation & Computing, University of Huddersfield, Huddersfield, UK, , 7–8.

Coudert, T., Grabot, B. & Archimède, B. 2002. Production/Maintenance cooperative scheduling using multi-agents and fuzzy logic. International Journal of Production Research 40(18): 4611–4632.

Daneshvar, B., Babek, R. & Eko, T. 2014. Supplier selection using integrated fuzzy topsis and mcgp: a case study. Procedia - Social and Behavioral Sciences 116: 3957–3970.

Gong, X., Pessemier, T.D., Joseph, W. & Martens, L. 2015. An energy-cost-aware scheduling methodology for sustainable manufacturing. Procedia CIRP 29: 185–190.

Ibrahim, A. & Hosam, A. 2018. A DEMATEL method in identifying design requirements for mobile environments: students' perspectives. Journal of Computing in Higher Education 30: 466–488.

Maity, S.R., & Chakraborty, S. 2013. Grinding wheel abrasive material selection using fuzzy TOPSIS method. Materials and Manufacturing Processes 28(4): 408–417.

Mansouri, M. & Leghris, C. 2020. A use of fuzzy TOPSIS to improve the network selection in wireless multiaccess environments. J Comput Networks Commun 1–12.

Pinedo, M. 1996. Scheduling: theory, algorithms, and systems. IIE Transactions 28(8): 695–697.

Singh, M. & Srivastava, R.K. 2018. Resources, conservation & recycling analysis of external barriers to remanufacturing using grey-DEMATEL Approach: an Indian perspective. Resources, Conservation & Recycling 136: 79–87.

Tavakkoli-Moghaddam, R., Javadi, B., Jolai, F. & Ghodratnama, A. 2010. The use of a fuzzy multi-objective linear programming for solving a multi-objective single-machine scheduling problem. Applied Soft Computing Journal 10(3): 919–925.

Taylan, O., Bafail, A.O., Abdulaal, R.M.S. & Kabli, M.R. 2014. Construction projects selection and risk assessment by fuzzy AHP and fuzzy TOPSIS methodologies. Applied Soft Computing Journal 17: 105–116.

Vujanovic, D., Momcilovic, V., Bojovic, N. , & Papic, V. 2012. Evaluation of vehicle fleet maintenance management indicators by application of DEMATEL and ANP. Expert Systems with Applications 39: 10552–10563.

Wu, H. & Tsai, Y. 2011. A DEMATEL method to evaluate the causal relations among the criteria in auto spare parts industry. Applied Mathematics and Computation 218(5): 2334–2342.

Yong, D. 2006. Plant location selection based on fuzzy TOPSIS. International Journal of Advanced Manufacturing Technology 28: 839–844.

Yurdakul, M. & Tansel, Y. 2009. Analysis of the benefit generated by using fuzzy numbers in a TOPSIS model developed for machine tool selection problems. Journal of Materials Processing Technology 209: 310–317.

Zhang, L., Sun, X. & Xue, H. 2019. Identifying critical risks in sponge city ppp projects using dematel method: a case study of China. Journal of Cleaner Production 226: 949–958.

Zhao, H. & Guo, S. 2014. Selecting green supplier of thermal power equipment by using a hybrid mcdm method for sustainability. Sustainability 6: 217–235

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/TAI.2022.3153593, IEEE Transactions on Artificial Intelligence

Projective Fisher Information for Natural Gradient Descent

Piyush Kaul, *Member, IEEE*, Brejesh Lall, *Member, IEEE*

***Abstract*—Improvements in neural network optimization algorithms have enabled shorter training times and the ability to reach state-of-the-art performance on various machine learning tasks. Fisher information based natural gradient descent is one such second-order method that improves the convergence speed and the final performance metric achieved for many machine learning algorithms. Fisher information matrices are also helpful to analyze the properties and expected behavior of neural networks. However, natural gradient descent is a high complexity method due to the need to maintain and invert covariance matrices. This is especially the case with modern deep neural networks, which have a very high number of parameters, and for which the problem often becomes computationally unfeasible. We suggest using the Fisher information for analysis of parameter space of fully connected and convolutional neural networks without calculating the matrix itself. We also propose a lower complexity natural gradient descent algorithm based on the projection of Kronecker factors of Fisher information combined with recursive calculation of inverses, which is computationally less complex and more stable. We finally share analysis and results showing that all these optimizations don't impact the accuracy while considerably lowering the optimization process's complexity. These improvements should enable applying natural gradient descent methods for optimization to neural networks with a larger number of parameters, than possible previously.**

***Impact Statement*—Deep Neural Networks achieve state of the art in various deep learning problems like computer vision and speech recognition. However, they are challenging to train, require extensive hyperparameter tuning, and take substantial training time running into many weeks. Hence it is important to ease the training by reducing the training time and amount of parameter tuning required. Natural gradient-based optimization methods reduce the training time and need for extensive tuning but are too complex for many networks. The training and analysis algorithm introduced here reduces the amount of time and effort required in performing the training significantly while not impacting the performance metric achieved at the given task.**

***Index Terms*—Deep Learning, Information Geometry, Natural Gradient Descent, Fisher information, Optimization**

I. Introduction

Natural gradient descent (NGD) can improve the convergence speed as well as the final performance metric achieved for neural network optimization[1][2][3][4]. NGD is based on Fisher information which conceptually indicates the information that a parameter has for a given observable random process. Fisher information has been further analyzed by using tools of mean-field theory and random weights [5][6]. The inverse of Fisher information is also linked with Cramer Rao Lower Bound, and hence indicative of the variance of the parameters estimated [7][8][9]. Thus, Fisher information is also helpful for analysis of the properties and expected behavior of neural networks and the state of training [10]. Natural Gradient Descent is an algorithm that uses Fisher information to enhance convergence speed and outcomes. It utilizes the curvature properties of loss functions (similar to other second-order techniques) to improve convergence. Conceptually while in standard stochastic gradient descent (SGD), we take a step in the direction to minimize the loss for a constant Euclidean distance between parameters, for natural gradient descent (NGD), we step in the direction that minimizes the loss for constant Kullback–Leibler divergence (K.L. divergence)[11]. However, it is a very high complexity method due to the need to invert large matrices of the size equal to the number of neural network parameters. With the help of Fisher information, we analyze and justify that the optimization step needs to be taken in only a small subset of available directions, and these can be found dynamically. We also present a method of recursive low complexity inversion of Fisher information matrices. We finally share results from empirical experiments on standard datasets showing that all these optimizations do not impact the accuracy or speed of the optimization.

Deep neural networks are the best-performing algorithms across a broad set of pattern recognition applications including computer vision, speech processing, natural language processing, graph processing, etc. [12]. Over the years specialized neural network architectures have evolved for specific application domains. Those would include convolutional neural networks (CNN)[13], recurrent neural networks (RNN)[14], generative adversarial networks (GAN) [15], transformers [16] etc. For example, convolutional neural networks, like AlexNet [17], ResNet [18], PNASNet [19] and EfficientNet [20] are especially popular as feature extractors for computer vision tasks with applications in image classification, object detection and semantic segmentation [21][22][23]. Deep learning architectures have recently been extended to new problem domains like Deep Fuzzy Hashing networks [24] for image retrieval, and reinforcement learning [25] for a variety of tasks including anomaly detection [26]. Alongside techniques like ensemble learning [27] continue to be used applications like mobile virtual resource management. There have been various attempts to utilize Fisher information

This work was supported in part by Bharti School of Telecom Technology and Management, IIT, Delhi."

Piyush Kaul is with the Department of Electrical Engineering, Indian Institute of Technology - Delhi, New Delhi, India-110016 (e-mail: eez157544@iitd.ac.in)

Brejesh Lall is with the Department of Electrical Engineering, Indian Institute of Technology - Delhi, New Delhi, India-110016 (e-mail: brejesh@iitd.ac.in).

This paragraph will include the Associate Editor who handled your paper.

2691-4581 (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 March 05,2022 at 06:03:03 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/TAI.2022.3153593, IEEE Transactions on Artificial Intelligence

matrix (FIM) and natural gradient descent (NGD) to improve the complexity and convergence speed of neural network optimization algorithms. The authors in [28] derive FIM and NGD for specific parameterizations of multivariate Gaussian distributions, and suggest that parameterization in terms of inverse variance and mean has certain advantages. In a more general setting [29] proposes a method to recast the graphical model estimation as an inference problem and try to optimize for the quality of the decision. For this, they utilize the information geometric optimization through natural gradient descent as an alternative to the expectation-maximization approach. They then show that the learned model performs better than traditional approaches and with lesser overfitting. The authors of [30] propose tackling structured parameter spaces by defining local parameter spaces which map back to global spaces and using structured natural gradient descent and show applications showing the applicability of their work to various distributions. A new algorithm called dual stochastic natural gradient descent (DSNGD) algorithm is proposed in [31] to solve the issue of high complexity and improve the convergence of natural gradient descent in comparison with SGD. This is done by first assuming dually flat manifolds (DFM) are the applicable family of distributions and then using a dual parameterization to find a bounded estimate of NGD.

Another set of optimizations include Factorized Natural Gradient and Kronecker Factored Approximate Curvature (KFAC) [32][33][34][35] which approximate local curvature by using much lower-dimensional factorized matrices instead. The authors of these papers first develop the idea for fully connected architectures and show a substantial improvement of convergence speed in training auto-encoders. They further extend these ideas in [33] to convolutional networks (KFC). More recently, the KFAC techniques have been extended to networks having batch normalization layers such that inter-dependencies introduced by batch normalization layers are accounted for. [36]. There have also been attempts at utilizing the Kronecker factoring technique, similar to KFAC, and combine it with Quasi Newton methods [37][38]. KFAC techniques have also been extended to utilize distributed computing utilizing very large batch sizes while minimizing the updates to fisher information matrices to reduce complexity [39][40]. Hence these techniques tend to utilize stale fisher information and require very large batch sizes to get accurate estimates. The techniques listed in this paper avoid these limitations, since we are able to iteratively update the Fisher information inverses over every minibatch with very little computational overhead, thus negating the need to have large intervals between updates.

We start from [32] and [33] and try the KFAC approximation for convolutional and fully connected layers. This entails estimating the Kronecker factor matrices, Γ and Ψ (section II), for each of these layers. We then propose a method for the Fisher information matrix's singular values from the Kronecker factor matrices without actually computing the Fisher matrix (given the large memory required) as explained in section III. From the analysis, we infer that the matrices are indeed low rank. We then utilize the estimated Fisher matrix Kronecker factors to find directions in which the loss metric variance is maximized. We then restrict our estimated covariance matrices to those sub-spaces and perform most of our calculations in the restricted subspace. Most implementations of the above methods don't calculate the inverses for every mini-batch, though correlations are updated. This is due to the complexity entailed in taking the inverse of large matrices. However, the periodic update of the inverse is sub-optimal. We introduce a method to recursively update inverses with low complexity, enabling us to update the weights for every mini-batch. We finally share our results for both fully connected and convolutional neural networks. We also share our implementation for all the experiments mentioned as publicly available code [41].

II. Background

A. Fisher Information and Natural Gradient

For the following description we utilize lower plain case letters for scalers (u), bold lower case for vectors ($\mathbf{u}$), and bold upper case for matrices ($\mathbf{U}$).

Let a learned model parameterized by $\boldsymbol{\theta}$ which has a distribution given by $p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta})$ or $p(\boldsymbol{z}|\boldsymbol{\theta})$, where $\boldsymbol{z}$ is a vector combining the input parameters ($\boldsymbol{x}$) and output parameters ($\boldsymbol{y}$). The score function is defined as the partial derivative of the log likelihood function given by:

$$\boldsymbol{S}(\boldsymbol{\theta}) = \frac{\partial log(p(\boldsymbol{z}|\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} = \nabla_{\theta} log(p(\boldsymbol{z}|\boldsymbol{\theta})). \tag{1}$$

It can be shown the expected value of score is 0. i.e.

$$\mathbb{E}(\boldsymbol{S}(\theta)) = \mathbb{E}\left[\frac{\partial log(p(\boldsymbol{z}|\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}}\right] = \int \frac{\frac{\partial p(\boldsymbol{z}|\boldsymbol{\theta})}{\partial (\boldsymbol{\theta})}}{p(\boldsymbol{z}|\boldsymbol{\theta})} * p(\boldsymbol{z}|\boldsymbol{\theta}) dz \tag{2}$$

$$= \frac{\partial}{\partial \boldsymbol{\theta}} \int p(\boldsymbol{z}|\boldsymbol{\theta}) dz = \frac{\partial}{\partial \boldsymbol{\theta}}(1) = 0. \tag{3}$$

The covariance of score is the Fisher information matrix defined by:

$$\boldsymbol{F} = \mathbb{E}[\boldsymbol{S}(\boldsymbol{\theta})\boldsymbol{S}(\boldsymbol{\theta})^T] = E\left[\nabla log(p(\boldsymbol{z}|\boldsymbol{\theta})) \nabla log(p(\boldsymbol{z}|\boldsymbol{\theta}))^T\right] \tag{4}$$

$$[\boldsymbol{F}]_{ij} = \mathbb{E}\left[\frac{\partial}{\partial \theta_i} log(p(\boldsymbol{z}|\boldsymbol{\theta})) \frac{\partial}{\partial \theta_j} log(p(\boldsymbol{z}|\boldsymbol{\theta}))^T\right] \tag{5}$$

Fisher information matrix gives us a measure of the accuracy of z given θ.

Since the actual distribution is unknown, the Fisher information matrix is calculated over input training data (representing by q_z). So, we replace expectation with empirical averaging.

$$\boldsymbol{F} = \frac{1}{N} \sum_{N} \left[\nabla log(p(\boldsymbol{z}|\boldsymbol{\theta})) \nabla log(p(\boldsymbol{z}|\boldsymbol{\theta}))^T\right] \tag{6}$$

where N is the number of training samples. It is also known that the negative expected Hessian of Log Likelihood is equal to Fisher information[34].

$$\boldsymbol{F} = -\mathbb{E}_{p(\boldsymbol{z}|\boldsymbol{\theta})}[\boldsymbol{H}_{log(p(\boldsymbol{z}|\boldsymbol{\theta}))}] \tag{7}$$

Fisher information is related to Cramer Rao bound (CR bound) and gives the accuracy of estimation [9][7]. For the

2691-4581 (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 March 05,2022 at 06:03:03 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/TAI.2022.3153593, IEEE Transactions on Artificial Intelligence

scaler case, the variance of unbiased estimation $\tilde{\theta}$ is bounded by the reciprocal of Fisher information. i.e.

$$var(\hat{\boldsymbol{\theta}}) = \frac{1}{\boldsymbol{F}(\boldsymbol{\theta})} \tag{8}$$

For the multivariate case, let the set of parameters of the model be $\boldsymbol{\theta} = [\theta_1, \theta_2, ..., \theta_d]^T \in \mathcal{R}^d$. Let $\boldsymbol{f_\theta}(\boldsymbol{z})$ be some function of input $\boldsymbol{z} = [z_1, z_2, ..., z_m]$ parameterized by $\boldsymbol{\theta}$. Let the expectation of $f_{\boldsymbol{\theta}}$ be ϕ_θ. The CR bound states that the covariance matrix of $\boldsymbol{f}_\theta$ satisfies.

$$Cov_\theta(\boldsymbol{f}_\theta(\boldsymbol{z})) \geq \frac{\partial \phi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} [\boldsymbol{F}(\theta)]^{-1} \left(\frac{\partial \phi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \tag{9}$$

where the inequality above is in terms of positive semi-definiteness $\frac{\partial \phi(\theta)}{\partial \theta}$ is Jacobian matrix. If $\boldsymbol{f_\theta}(\boldsymbol{x})$ is unbiased the inequality becomes.

$$Cov_\theta(\boldsymbol{f}_\theta(\boldsymbol{z})) \geq [\boldsymbol{F}(\boldsymbol{\theta})]^{-1} \tag{10}$$

We also note the following properties of Fisher information matrix [10].

1) Diagonal Elements of Cov_θ are covariance of individual parameters and give a range within which those can vary. However, it doesn't indicate learning performance.
2) Hyper Ellipsoid volume $|Cov_\theta|$ is a measure of the large number of possible combinations in weight that could generate the data with probability larger than p. So large volume could indicate the ease of learning.
3) Correlation between parameters can specifically be given by

$$C_{ij} = \frac{Cov(\boldsymbol{\theta})_{ij}}{\sqrt{Cov(\boldsymbol{\theta})_{ii} * Cov(\boldsymbol{\theta})jj}} \tag{11}$$

If large number of parameters are highly correlated i.e. $C_{ij} \approx 1$, this may indicate the solution space is banana shaped.

Let us assume that the neural network (NN) is trained for mean square error (MSE) reduction. Since the matrix is symmetric, all eigenvalues and eigenvectors exist and are real. Also, eigenvectors form a basis. The following properties can be inferred.

1) Large covariance values indicate low sensitivity on MSE for those parameters and vice versa.
2) The eigenvalues and eigenvectors define principal directions in MSE space.
3) The ratio of largest and second-largest eigenvalues $|\lambda_{max}/\lambda_{max-1}|$ gives an indication if optimization space is banana-shaped. Large value indicates slow learning in gradient descent.
4) The ratio between largest and smallest eigenvalue $|\lambda_{max}/\lambda_{min}|$ is also indicative of learning speed.
5) A large number of eigenvector correspond to directions that are very insensitive to MSE.

B. Natural Gradient

Let S be Riemannian manifold ([42][43][44]) parameterized by $\boldsymbol{\theta} \in \mathcal{R}^m$ over which function $\mathcal{L}(\boldsymbol{\theta})$ is defined. For Euclidean space, length of vector $d\boldsymbol{\theta}$ connecting $\boldsymbol{\theta}$ with $\boldsymbol{\theta} + d\boldsymbol{\theta}$ is given by

$$|d\boldsymbol{\theta}|^2 = \sum_{i,j} g_{i,j} d\boldsymbol{\theta}_i d\boldsymbol{\theta}_j \tag{12}$$

Here the matrix $\boldsymbol{G} = (g_{ij})$ is the Riemannian metric tensor. Steepest descent implies maximum change in $L(\boldsymbol{\theta})$ for unit distance over $\boldsymbol{\theta}$ i.e. minimize $L(\boldsymbol{\theta} + d\boldsymbol{\theta})$ when $d\boldsymbol{\theta}$ is fixed as

$$|d\boldsymbol{\theta}|^2 = \epsilon^2 \tag{13}$$

The steepest descent of $L(\boldsymbol{\theta})$ in Riemannian space is given by (please see [2] for proof)

$$-\tilde{\nabla} L(\boldsymbol{\theta}) = -\boldsymbol{G}^{-1}(\boldsymbol{\theta}) \nabla L(\boldsymbol{\theta}) \tag{14}$$

where $\boldsymbol{G}^{-1} = (g^{ij})$ is the inverse of the metric $G = (g_{ij})$ and ∇L is the normal gradient given as

$$\nabla L(\theta) = \left(\frac{\partial}{\partial \theta_1} L(\theta), \frac{\partial}{\partial \theta_2} L(\theta), ... \right)^T \tag{15}$$

Fisher information defines the local curvature of distribution space of which is measured through KL-divergence. More specifically, Fisher information matrix (FIM) is the Hessian of KL-divergence between two distributions $p(\boldsymbol{x}|\boldsymbol{\theta})$ and $p(\boldsymbol{x}|\boldsymbol{\theta'})$ with respect to $\boldsymbol{\theta}$ at $\boldsymbol{\theta'} = \boldsymbol{\theta}$ i.e

$$\boldsymbol{H}_{KL[p(x|\theta)||p(x|\theta')]} = \boldsymbol{F} \tag{16}$$

C. Kronecker Factored Approximate Gradient

Each layer in a multilayer perceptron (MLP) can be considered to consist of the following steps.

$$\boldsymbol{s}_l = \boldsymbol{W}_l \bar{\boldsymbol{a}}_{l-1} \tag{17}$$

$$\boldsymbol{a}_l = \phi_l(\boldsymbol{s}_l) \tag{18}$$

Here $l \in \{1..L\}$ is the layer number, $\boldsymbol{s}_l$ is the output of matrix multiplication and $\boldsymbol{a}_l$ is the output of activation function. W_l represents the matrix of weights for layer l, with bias added as one of the columns, and $\bar{a}_l$ is the same as a_l but with an appended dimension for homogeneous coordinate to accommodate bias. Also, ϕ_l is the non-linear activation that is applied per element. We can represent all parameters in a neural network with a single vector $\boldsymbol{\theta} = (vec(\boldsymbol{W}_1)^T, vec(\boldsymbol{W}_2)^T, vec(\boldsymbol{W}_3)^T ... vec(\boldsymbol{W}_L)^T)$ and the computed output as $f(\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{a}_L$

If p_{data} is the training data distribution and $R_{\boldsymbol{y}|f(\boldsymbol{x},\boldsymbol{0})}$ is the output prediction distribution, we can denote the gradient of log likelihood function as $\mathscr{D}\boldsymbol{\theta} = \nabla_\theta L(\boldsymbol{y}, f(\boldsymbol{x}, \boldsymbol{\theta}))$, and the Fisher information matrix is given by

$$\boldsymbol{F} = \mathbb{E}_{x \sim p_{data}, y \sim R_{y|f(x,\theta)}} [\mathscr{D}\theta (\mathscr{D}\theta)^T] \tag{19}$$

The simplest version of the KFAC algorithm is the block diagonal one, and it is based on two approximations. The first approximation is that the parameters across layers are

2691-4581 (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 March 05,2022 at 06:03:03 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/TAI.2022.3153593, IEEE Transactions on Artificial Intelligence

uncorrelated. And hence $\boldsymbol{F}$ becomes a block diagonal matrix such as

$$\boldsymbol{F} \approx \begin{bmatrix} \mathbb{E}[vec(\mathscr{D}\bar{\boldsymbol{W}}_1)vec(\mathscr{D}\bar{\boldsymbol{W}}_1)^T] & \dots & 0 \\ .. & \dots & \\ 0 & \dots & \mathbb{E}[vec(\mathscr{D}\bar{\boldsymbol{W}}_L)vec(\mathscr{D}\bar{\boldsymbol{W}}_L)^T] \end{bmatrix} \tag{20}$$

For the second approximation we note that

$$\mathbb{E}\left[\mathscr{D}[\boldsymbol{W}_l]_{ij}\mathscr{D}[\boldsymbol{W}_l]_{i'j'}\right] = \mathbb{E}[\mathscr{D}[s_l]_i[\bar{a}]_j\mathscr{D}[s_l]_i{}'[\bar{a}]_j{}'] \tag{21}$$

By assuming pre-activation and post-activation values to be uncorrelated, one can approximate

$$\mathbb{E}\left[\mathscr{D}[\boldsymbol{W}_l]_{ij}\mathscr{D}[\boldsymbol{W}_l]_{i'j'}\right] \approx \mathbb{E}[\mathscr{D}[s_l]_i[\bar{a}]_j]\mathbb{E}[\mathscr{D}[s_l]_i{}'[\bar{a}]_j{}'] \tag{22}$$

Substituting the $\boldsymbol{\Psi}_{l-1} = \mathbb{E}[\bar{\boldsymbol{a}}_{l-1}\bar{\boldsymbol{a}}_{l-1}^T]$ and $\boldsymbol{\Gamma}_l = \mathbb{E}[\mathscr{D}\boldsymbol{s}_l\mathscr{D}\boldsymbol{s}_l^T]$ and also considering that Kronecker product identity $(\boldsymbol{A}\otimes\boldsymbol{B})^{-1} = (\boldsymbol{A}^{-1}\otimes\boldsymbol{B}^{-1})$ we can represent inverse Fisher matrix as

$$\boldsymbol{F}^{-1} = \begin{bmatrix} \boldsymbol{\Psi}_1^{-1}\otimes\boldsymbol{\Gamma}_1^{-1} & \dots & 0 \\ .. & \dots & \\ 0 & \dots & \boldsymbol{\Psi}_L^{-1}\otimes\boldsymbol{\Gamma}_L^{-1} \end{bmatrix} \tag{23}$$

But we don't need to actually ever do the Kronecker product. We just keep track of the factors and compute the gradient $\boldsymbol{F}^{-1}\nabla h$ directly. Using the vector trick $(\boldsymbol{A}\otimes\boldsymbol{B})vec(\boldsymbol{C}) = vec(\boldsymbol{B}\boldsymbol{C}\boldsymbol{A}^T)$, we get the following

$$\boldsymbol{F}^{-1}\nabla\boldsymbol{h} = \begin{bmatrix} vec(\boldsymbol{\Gamma}_1^{-1}(\nabla_{W_1}h)\boldsymbol{\Psi}_1^{-1} \\ \dots \\ vec(\boldsymbol{\Gamma}_L^{-1}(\nabla_{W_L}h)\boldsymbol{\Psi}_L^{-1}) \end{bmatrix} \tag{24}$$

III. Projective KFAC

Even though KFAC is a considerably lower complexity than baseline natural gradient descent, the number of parameters per layer (M_l) may still be too high. Since for KFAC, we still need to invert matrices Ψ_l and Γ_l matrix, the complexity of NGD may still be too high. So, we propose PKFAC.

A. Estimating Rank

We note that we would like to obtain the singular values of the inverse Fisher information matrix $\boldsymbol{F}^{-1}$ to estimate the Rank. However, we never compute the actual Fisher information matrix and calculate only Kronecker factors $\boldsymbol{\Gamma}$ and $\boldsymbol{\Psi}$ Suppose the SVD of the two factors are respectively.

$$\boldsymbol{\Psi}^{-1} = \boldsymbol{U}_\Psi\boldsymbol{\Delta}_\Psi\boldsymbol{V}_\Psi^T \tag{25}$$

$$\boldsymbol{\Gamma}^{-1} = \boldsymbol{U}_\Gamma\boldsymbol{\Delta}_\Gamma\boldsymbol{V}_\Gamma^T \tag{26}$$

We note that utilizing the mixed product property of Kronecker products.

$$\boldsymbol{F}^{-1} = \boldsymbol{\Psi}^{-1}\otimes\boldsymbol{\Gamma}^{-1} = \boldsymbol{U}_\Psi\boldsymbol{\Delta}_\Psi\boldsymbol{V}_\Psi^T\otimes\boldsymbol{U}_\Gamma\boldsymbol{\Delta}_\Gamma\boldsymbol{V}_\Gamma^T \tag{27}$$

$$= (\boldsymbol{U}_\Psi\otimes\boldsymbol{U}_\Gamma)(\boldsymbol{\Delta}_\Psi\otimes\boldsymbol{\Delta}_\Gamma)(\boldsymbol{V}_\Psi\otimes\boldsymbol{V}_\Gamma)^T \tag{28}$$

Hence the singular value matrix of $\boldsymbol{F}^{-1}$ is given by

$$\boldsymbol{\Delta}_{\boldsymbol{F}^{-1}} = (\boldsymbol{\Delta}_\Psi\otimes\boldsymbol{\Delta}_\Gamma) \tag{29}$$

Hence, we can estimate $\boldsymbol{\Delta}_\Gamma$ and $\boldsymbol{\Delta}_\Psi$ individually and estimate the singular values using a Kronecker product of a diagonal matrices which is computationally only $M \times N$ individual multiplications only. These singular values, as shown in Figure 1 give an estimate of the rank of Fisher information.

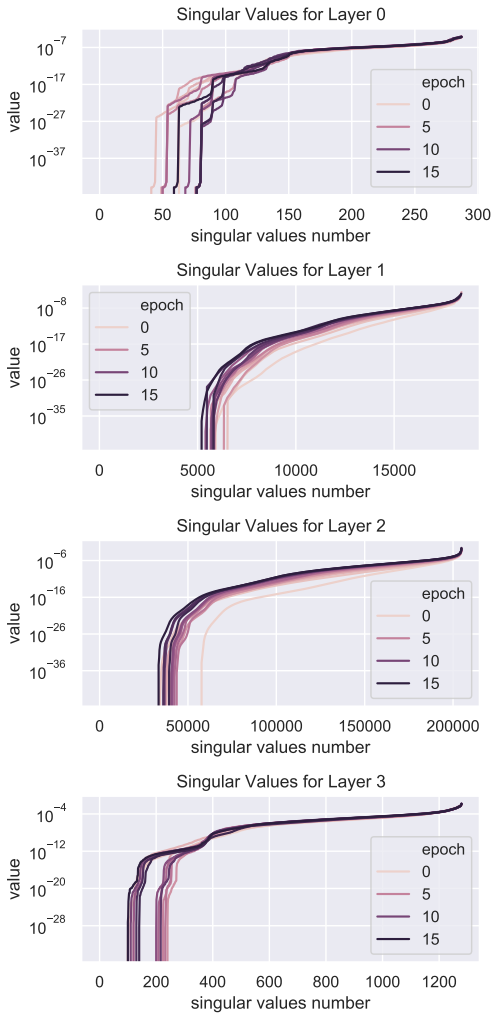


Fig. 1. Singular Values for Fashion MNIST CNN. The x-axis indicates the indices of singular values for each layer, while the y-axis has magnitude of singular values on log scale

B. Projection Algorithm

If p_{data} is the training data distribution and $R_{y|f(x,0)}$ is the output prediction distribution, we can denote the log likelihood function as $\mathscr{D}\theta = \nabla_\theta L(y, f(x,\theta))$, and the Fisher information matrix is given by

$$\boldsymbol{F} = \mathbb{E}_{\boldsymbol{x}\sim p_{data},\boldsymbol{y}\sim R_{\boldsymbol{y}|f(\boldsymbol{x},\boldsymbol{\theta})}}[\mathscr{D}\boldsymbol{\theta}(\mathscr{D}\boldsymbol{\theta})^T]$$

For Simplicity

$$\boldsymbol{F} = \mathbb{E}[\mathscr{D}\boldsymbol{\theta}(\mathscr{D}\boldsymbol{\theta})^T]$$

Then we can do an eigenvector decomposition of F

$$\boldsymbol{F} = \boldsymbol{Q}\boldsymbol{S}\boldsymbol{Q}^T$$

2691-4581 (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 March 05,2022 at 06:03:03 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/TAI.2022.3153593, IEEE Transactions on Artificial Intelligence

where, $\boldsymbol{Q}$ is the eigenvector matrix with columns corresponding to eigenvectors of $\boldsymbol{F}$, and S is eigenvalue diagonal matrix. If we set the lowest K $(K < M)$ eigenvalues in S to zero, we can set the corresponding K columns of Q to zero. Let us call this matrix $\bar{\boldsymbol{Q}}$ $(K \times M)$ with only columns corresponding to nonzero eigenvalues of S.

We can then obtain a projection of $\mathscr{D}\boldsymbol{\theta}$ to a subspace by the following matrix multiplication

$$\hat{\mathscr{D}}\boldsymbol{\theta} = \boldsymbol{Q}^T \mathscr{D}\boldsymbol{\theta} \tag{30}$$

Let $\bar{\mathscr{D}}\boldsymbol{\theta}$ be shortened vector $\mathscr{D}\boldsymbol{\theta}$ of length K obtained by

$$\bar{\mathscr{D}}\boldsymbol{\theta} = \bar{\boldsymbol{Q}}^T \mathscr{D}\boldsymbol{\theta} \tag{31}$$

We can obtain $\boldsymbol{F}$ from vector $\bar{\mathscr{D}}\theta$

$$\begin{aligned} \boldsymbol{F} &= \mathbb{E}[\mathscr{D}\boldsymbol{\theta}(\mathscr{D}\boldsymbol{\theta})^T] && (32)\\ &= \mathbb{E}[\boldsymbol{Q}\hat{\mathscr{D}}\boldsymbol{\theta}(\boldsymbol{Q}\hat{\mathscr{D}}\boldsymbol{\theta})^T] && (33)\\ &\approx \mathbb{E}[\bar{\boldsymbol{Q}}\bar{\mathscr{D}}\boldsymbol{\theta}(\bar{\boldsymbol{Q}}\bar{\mathscr{D}}\boldsymbol{\theta})^T] && (34)\\ &= \bar{\boldsymbol{Q}}\mathbb{E}[\bar{\mathscr{D}}\boldsymbol{\theta}(\mathscr{D}\bar{\boldsymbol{\theta}})^T]\bar{\boldsymbol{Q}}^T && (35) \end{aligned}$$

C. Computing Inverse

In reality, we obtain Fisher information iteratively where expectation in the above equations is replaced by empirical averaging of the form.

$$\boldsymbol{F}(t) = \lambda(\boldsymbol{F}(t-1)) + (1-\lambda)[\mathscr{D}\boldsymbol{\theta}(\mathscr{D}\boldsymbol{\theta})^T] \tag{36}$$

where λ is some averaging constant and $\boldsymbol{F}(t)$ is the estimate of Fisher information at iteration t and $\boldsymbol{F}(t-1)$ is the one available at iteration $(t-1)$. Noting that we can utilize the above simplification and project $\mathscr{D}\boldsymbol{\theta}$ first to $\bar{\mathscr{D}}\boldsymbol{\theta}$ to reduce the size of the matrices. We maintain $\bar{\boldsymbol{F}}(t)$ instead by a similar equation

$$\bar{\boldsymbol{F}}(t) = \lambda(\bar{\boldsymbol{F}}(t-1)) + (1-\lambda)[\bar{\mathscr{D}}\boldsymbol{\theta}(\bar{\mathscr{D}}\boldsymbol{\theta})^T] \tag{37}$$

Whenever we need to update weights, we utilize the following equation to obtain estimate of $\boldsymbol{F}(t)$

$$\boldsymbol{F}(t) = \boldsymbol{Q}\bar{\boldsymbol{F}}(t)\boldsymbol{Q}^T \tag{38}$$

where $\bar{\boldsymbol{F}}(t) \approx \mathbb{E}[\bar{\mathscr{D}}\boldsymbol{\theta}\bar{\mathscr{D}}\boldsymbol{\theta}^T]$ in Equation 32. For further simplification we utilize matrix inversion lemma to avoid inverting the matrix. This is given as

$$(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} - VA^{-1}U)^{-1}VA^{-1} \tag{39}$$

Substituting $C = I$ and $U = \mathscr{D}\boldsymbol{\theta}$, $V = U^T$, $\alpha = \frac{1}{\lambda}$ and $\beta = \sqrt{(1-\lambda)}$ we obtain.

$$\begin{aligned} \boldsymbol{F}(t)^{-1} &= [\lambda\boldsymbol{F}(t-1) + (1-\lambda)\mathscr{D}\boldsymbol{\theta}(\mathscr{D}\boldsymbol{\theta})^T]^{-1} && (40)\\ &= \alpha\boldsymbol{F}(t-1)^{-1} - \alpha\beta\boldsymbol{F}(t-1)^{-1}\mathscr{D}\boldsymbol{\theta}(1 && (41)\\ &+ \beta^2\alpha(\mathscr{D}\boldsymbol{\theta})^T\boldsymbol{F}(t-1)^{-1}\mathscr{D}\boldsymbol{\theta})\beta\alpha\mathscr{D}\boldsymbol{\theta}\boldsymbol{F}(t-1)^{-1} && (42) \end{aligned}$$

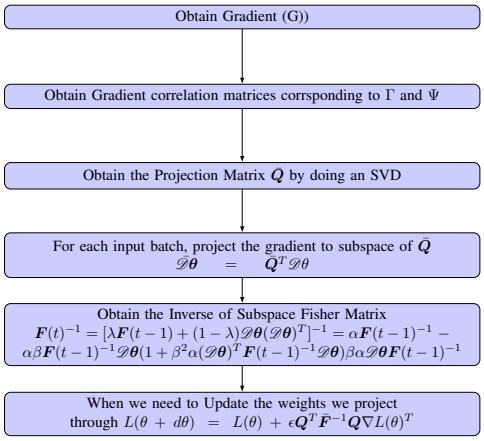


Fig. 2. Algorithm for projective KFAC

D. Extending to Convolutional Layers

For the case of convolutional layers, besides the assumption of independence between the activations and pre-activation derivatives, we need to make two more assumptions. First is that the first-order statistics for activations are independent of the spatial location, while for both activations and pre-activations, second-order statistics are only dependent on the spatial distance between any two spatial locations. Secondly, the pre-activation derivatives across spatial locations are uncorrelated.

Besides this we define the *im2col* operator denoted by $[\![\cdot]\!]$. This operator extracts a patch of activation equal to the size of the kernel for each spatial location in the input activation map and converts them to vectors. This stack of vectors for all spatial locations can then be formed into a matrix, and convolution operations can be represented by a matrix multiply similar to Equation 18 for Linear Layers i.e.

$$\begin{aligned} \boldsymbol{s}_l &= \boldsymbol{W}_l[\![\bar{\boldsymbol{a}}_{l-1}]\!] && (43)\\ \boldsymbol{a}_l &= \phi_l(\boldsymbol{s}_l) && (44) \end{aligned}$$

Here $[\![\bar{a}_{l-1}]\!]$ is the activation matrix after patches have been extracted with the *im2col* operator, and a single one has been appended as a homogeneous coordinate to the end so that bias value can be added within the matrix multiply operation.

E. Damping

Due to weight decay and approximation in calculation of Fisher matrix, we usually like to add a damping factor by way of multiple of identity matrix to Fisher information matrix i.e. we would like to calculate $[\bar{\boldsymbol{F}} + \lambda\boldsymbol{I}]^{-1}$ instead of $[\bar{\boldsymbol{F}}]^{-1}$. However, since we utilize the factorized matrices and never use the Fisher information matrix directly, [32] suggests using the following factorization structure.

$$\boldsymbol{F}_l + \lambda\boldsymbol{I} \approx (\boldsymbol{\Psi}_l + \pi_l\sqrt{\lambda}\boldsymbol{I}) \otimes (\boldsymbol{\Gamma}_l + \frac{1}{\pi_l}\sqrt{\lambda}\boldsymbol{I}) \tag{45}$$

Here π_l is calculated as follows.

2691-4581 (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 March 05,2022 at 06:03:03 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/TAI.2022.3153593, IEEE Transactions on Artificial Intelligence

$$\pi_l = \sqrt{\frac{||\Psi_{l-1} \otimes \boldsymbol{I}||_v}{||\boldsymbol{I} \otimes \Gamma_l||_v}} \tag{46}$$

Where the norm $||\cdot||_v$ in practice the trace norm.

F. Random Projections

The Johnson-Lindenstrauss lemma (JL Lemma)[45] states that a set of points in high dimensional space can be embedded into lower dimensional space with random projections while approximately maintaining the pair-wise distance between points. The JL lemma states that that for any $0 < \epsilon < 1$, and given set of N points in set $V \in R^d$ there exists a linear map f from original dimension d to reduced dimension m, $f : R^N \mapsto R^m$, such that for all points $u, v \in V$ the following relation holds.

$$(1-\epsilon)||u-v||^2 \leq ||f(u)-f(v)||^2 \leq (1+\epsilon)||u-v||^2 \tag{47}$$

More specifically the JL lemma suggests that we can find the reduced dimension m from the number of samples N by the following relation.

$$m \geq 4log(N)/(\epsilon^2/2 - \epsilon^3/3) \tag{48}$$

This implies that we could possibly utilize random projections instead of orthogonal projections derived from the SVD for our projection step by keeping to the constraint above. We explore this option in some of our experiments using Gaussian projections.

G. Complexity

We analyze the complexity of Fisher information based techniques based on single fully connected layer and assume the following variables.

- p: size of the mini-batch
- q: input feature maps of the fully connected layer.
- r: output feature maps of the fully connected layer.

We note from (48) that dimension of matrices after projection can be assumed to be $klog(q)$ and $klog(r)$ respectively, for some constant k. We note the following matrix sizes for clarity.

- input feature map: $p \times q$
- filter : $q \times r$
- output fature map: $p \times r$
- Fisher information matrix $\boldsymbol{F}$: $(qr) \times (qr)$
- $\boldsymbol{\Psi}$ matrix : $q \times q$
- $\boldsymbol{\Gamma}$ matrix : $r \times r$
- $\hat{\boldsymbol{\Psi}}$ matrix = $klog(q) \times klog(q)$
- $\hat{\boldsymbol{\Gamma}}$ matrix = $klog(r) \times klog(r)$

Then the Table I gives the complexity.

We note from Table I that $O(q^2log(q)^3)$ is less than $O(q^3)$ and hence the computational complexity is greatly reduced by projection. We also note that the order of complexity of the recursive version is the same. However, the total compute will usually be less since the batch size is usually much less than the number of input or output features i.e., $p << q$ and $p << r$.

Algorithm	Projection	Inverse	Total Compute	Order
Fisher information	-	$(qr)^6$	$(qr)^6$	$O((qr)^6))$
KFAC	-	$q^3 + r^3$	$q^3 + r^3$	$O(q^3) + O(r^3)$
PKFAC	$q^2 * (klog(q)^3 + r^2 * (klog(r))^3$	$klog(q)^3 + klog(r)^3$	$q^2 * (klog(q)^3 + r^2 * (klog(r))^3 + klog(q)^3 + klog(r)^3$	$O(q^2log(q)^3) + O(r^2 * (log(r)^3)$
PKFAC (recursive)	$q^2 * (klog(q)^3 + r^2 * (klog(r))^3$	$klog(q)^3 + klog(r)^3$	$q^2 * (klog(q)^3 + r^2 * (klog(r))^3 + 2 * klog(p)^3$	$O(q^2log(q)^3) + O(r^2 * (log(r)^3)$

TABLE I
COMPLEXITY ESTIMATES

IV. EXPERIMENTAL MEASUREMENT

A. Autoencoder

We first try to utilize an autoencoder as described in [46]. This autoencoder is used to reduce the dimensionality of input in a manner similar to PCA. It then reconstructs the low dimensional vector back to the original input size. The training is done against a mean square error criterion. We utilize MNIST dataset for our experiments [47]. The number of features per layer is given by 784-1000-500-250-30-250-500-1000-784. This is an extremely difficult network to train. We first utilize a Restricted Boltzmann machine (RBF) to pre-train the network one pair of layers at a time. We utilize the contrastive divergence algorithms as described in [48] to train all the RBFs and utilize those weights in the encoder. We then utilize fixed learning rate of 0.001 for the entire network and run the autoencoder training on the MNIST dataset. Starting from the same set of weights, we utilize both the standard SGD algorithm as well as NGD KFAC and NGD PKFAC variants. The loss and MSE plots obtained by training are shown in Figure 3. We notice that both KFAC and PKFAC perform much better than SGD in the final MSE metric. We also note that there is no noticeable difference in KFAC and PKFAC convergence speed or final metric achieved when measured w.r.t. epochs. However, when we plot the MSE w.r.t. time plot, we notice that the total time taken by PKFAC is lower than that for KFAC, even though the final metric achieved is the same.

B. Multi Layer Perceptron with Fashion MNIST

We next experiment with a three-layer MLP network. All layers are fully connected with features 784 - 256 - 256 - 10. We first estimate approximate KFAC factor matrices Γ and Ψ and then use the singular value estimation method mentioned above for finding singular values of Fisher Matrices associated with these layers. We then generate projection matrices of various subspace sizes and train the networks over Fashion MNIST [49] dataset. We fix the number of epochs to 100 for all experiments. We use grid search to find the optimal learning rate and gamma parameters for the learning hyperparameters.

2691-4581 (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 March 05,2022 at 06:03:03 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/TAI.2022.3153593, IEEE Transactions on Artificial Intelligence

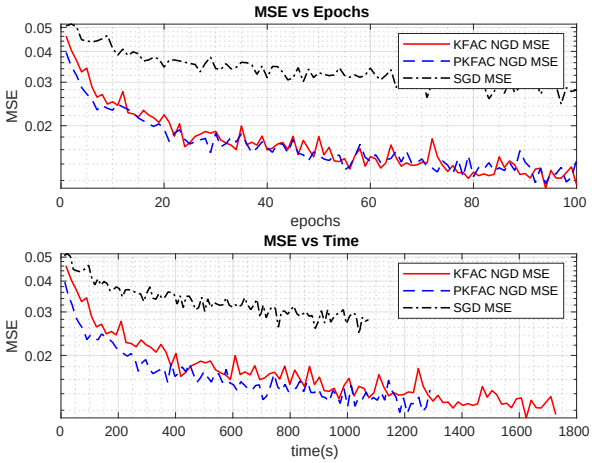


Fig. 3. Autoencoder with MNIST: SGD vs NGD

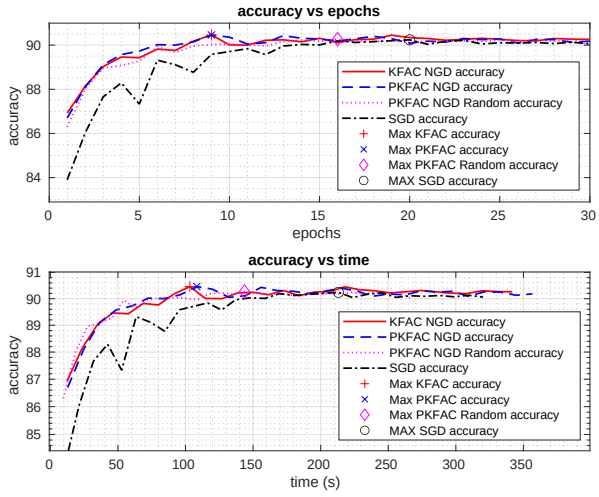


Fig. 4. Loss and accuracy over epochs for SGD and NGD for MLP with Fashion MNIST

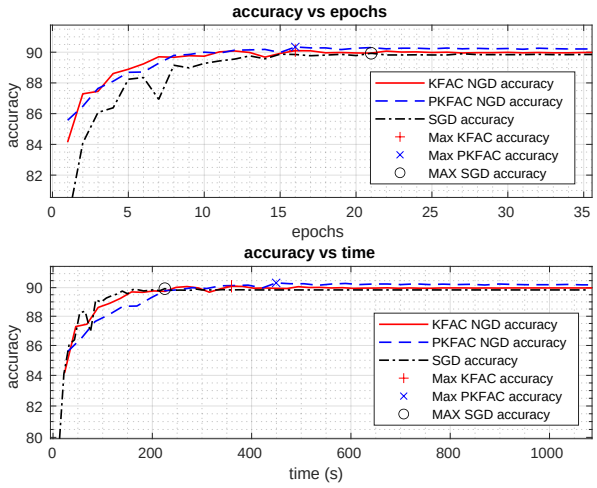


Fig. 5. Loss and accuracy over epochs for SGD and NGD for LeNet-5 with Fashion MNIST

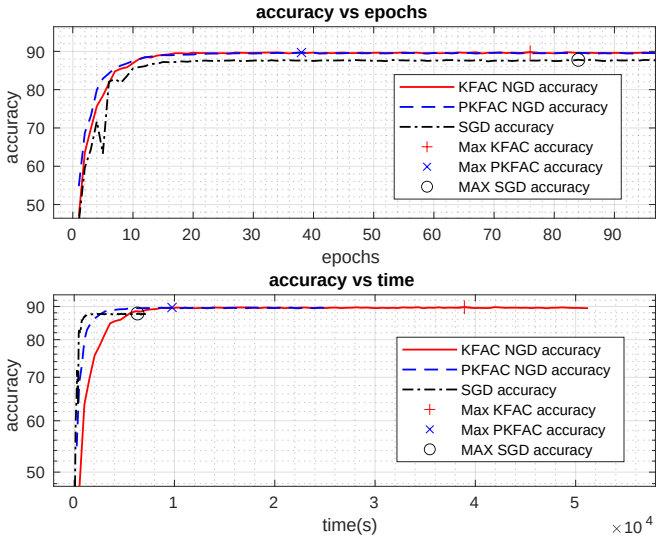


Fig. 6. Loss and accuracy over epochs with ResNet18 and CIFAR10

We also measure the impact of varying the periodicity of projection matrix update and inverse estimation. We finally utilize the recursive inverse update algorithm above and see the impact of final accuracy after 100 epochs. The decrease in learning rate is done by the parameter gamma, which is the factor (≤ 1) which we multiply to the learning rate every epoch. The optimum value for gamma is also found through grid search. We get the plots show in Figure 4 from our experiments. We again note that both the speed and final convergence are much better in PKFAC and KFAC compared to SGD. From the "accuracy vs. epochs" (top) plot, we can see that both KFAC and PKFAC reach the maximum accuracy in the same epoch. Contrasting this with the accuracy vs. time (bottom) plot, we notice that PKFAC reaches the identical maximum much earlier compared to KFAC. This is due to the inherently lower complexity of PKFAC.

The hyperparameters listed in Table III are the ones for which the best performance was obtained for a particular gradient descent scheme. We don't list some hyperparameters like periodicity, for which much fine-tuning doesn't seem to have an impact.

We note that final accuracy is best for Natural Gradient Descent methods and better than SGD. Besides this, we note that PKFAC doesn't seem to have any negative impact on the final accuracy reached. .

We perform damping as per the modified formula derived above. However, we don't notice any noticeable impact of damping for the projective method.

Lastly, we also tried random projections for the given experiments and noticed a very small impact on the final achieved accuracy. This is listed as "PKFAC Random" in the plots show in Figure 4.

C. LeNet-5 with Fashion MNIST

The CNN we start with the simple LeNet-5 architecture [50], which has two repetitions of Conv-Relu-MaxPool followed by two repetitions of Fully Connected-Relu, followed by a single fully connected layer to arrive at a ten-way classification. The number of features used is 6-16-400-120-

2691-4581 (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 March 05,2022 at 06:03:03 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/TAI.2022.3153593, IEEE Transactions on Artificial Intelligence

84-10. Similar to MLP, we use grid search to find optimal values of gamma and initial learning rate for each optimization algorithm separately. We repeat the above experiments but with modifications to the algorithm required for convolutional layers as described in III-D. For estimating covariance, we utilize the im2col type technique for creating the vectors from activations but with a randomized subset. The reason for utilizing a randomized subset is that utilizing the whole set requires too much memory and a subset seems adequate in practice for providing a good estimate of the statistics required. Once again, we repeat experiments with both NGD, NGD, and SGD. For NGD, we try the experiment with both standard and projective variants. As shown in the Figure 5, the KFAC NGD and PKFAC NGD converge much faster than SGD when measured in terms of epochs and give better final accuracy. However, when measured against wall clock time (accuracy vs. time (s)), we note that the SGD is the fastest to reach its maximum, followed by PKFAC, with KFAC being slowest. However, given the maximum accuracy reached with PKFAC is better than SGD, still choosing PKFAC may be a reasonable tradeoff.

D. Resnet and with CIFAR10

We also wanted to see the applicability of the techniques mentioned above to more modern neural network architecture. We extend the same method to Resnet-18 [18] architectures with the CIFAR10 dataset [51] . The number of features is standard, and we again use grid search to obtain optimal hyper-parameters. Please see section VII for hyper-parameter selection details. The resulting plots and results are show in Figure 6 and Table IV. We note that here also NGD performs better than SGD, though the difference in final performance is not as much as in the case of the autoencoder. However, the rate of convergence with KFAC NGD and PKFAC NGD is much faster in terms of epochs. Even in terms of wall clock time, PKFAC takes only a little more time in comparison to SGD and much less than KFAC.

E. Super-Resolution with ESPCN

Finally, we try to so the impact of the techniques on some other problem domain besides classification and autoencoders. We apply the same set of techniques on the problem of super-resolution as described in [52] for efficient sub-pixel convolutional network (ESPCN). We use the Berkley Segmentation dataset BSD300 [53] for training and test images. For training a super-resolution network of factor r, an high resolution (HR) image from the training set is gaussian filtered and then subsampled by $1/r$ to generate a low resolution (LR) image. This image is then fed to a series of four convolution layers of kernel size (3×3) and feature sizes (1-64-64-32-r^2). ReLU is used as the activation function after each conv layer. After the final convolution layer the feature maps of dimensions $H \times W \times C * r^2$ are generated which are then fed to a pixel shuffle layer that shuffles the pixels to create an output image of dimensions $H * r \times W * r \times C$. This is the super-resolution (SR) image of the same size as the original HR image and can be used as a target for training the network according to the mean square loss criterion. For these experiments, we again check for optimal hyperparameters and then run a set of experiments as shown in Figure 7 and VI. The validation is carried out by measuring the PSNR (dB) between the SR image generated by the network and the original HR image. We note from the results that KFAC and PKFAC perform much better in terms of final accuracy compared to SGD. The accuracy improvement across epochs with PKFAC and KFAC is similar. However, when considered against wall clock time, PKFAC is much faster in reaching high PSNR compared to KFAC.

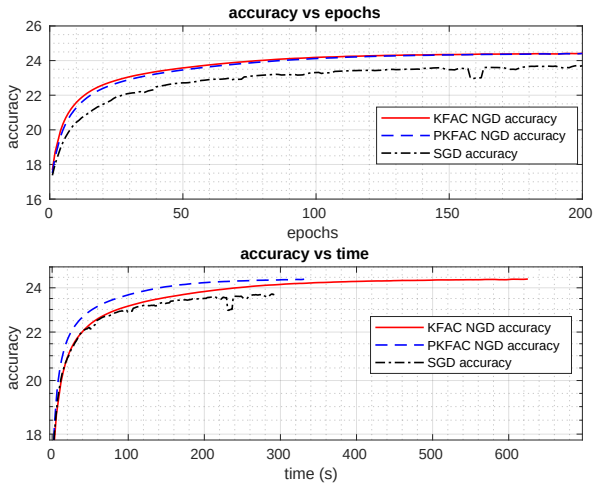


Fig. 7. Super-resolution PSNR variation over Epochs for ESPCN with BSD300

TABLE II
ACCURACY RESULTS WITH MLP AND FASHION MNIST

optimizer	projective	test acc	train acc	stdev (test acc)
SGD	No	90.19%	91.20%	0.11%
NGD	No	90.49%	92.23%	0.10%
NGD	SVD	90.48%	92.15%	0.10%
NGD	Random	90.35%	92.11%	0.10%

Mean Accuracy results with MLP and Fashion MNIST over 30 independent randomly intialized runs. Learning rate of 0.1 was used for all SGD and 0.01 for all NGD simulations. Gamma is 0.8 in all cases. Standard deviation of test accuracy is indicated in last column. Significant performance improvement observed in NGD compared to SGD (p-value < 0.05 with two sample t-test)

V. CONCLUSION AND FUTURE WORK

We note from classification experiments done over the Fashion MNIST, CIFAR10 and BSD300 datasets that the natural gradient descent performs better than plain SGD. In the case of autoencoder, which is an inherently tougher optimization problem, the accuracy gap (MSE) between using SGD and NGD is very large. We also note that the projective step is theoretically justified by analyzing the Kronecker factors' singular values and doesn't lead to any loss in performance. We further note that we can use

2691-4581 (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 March 05,2022 at 06:03:03 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/TAI.2022.3153593, IEEE Transactions on Artificial Intelligence

TABLE III
ACCURACY RESULTS WITH LENET-5 AND FASHION MNIST

optimizer	projective	test acc	train acc	stdev (test acc)
SGD	No	91.49%	96.50%	0.10
NGD	No	91.81%	96.43%	0.12
NGD	SVD	91.81%	96.43%	0.12
NGD	Random	91.79%	96.43%	0.12

Mean accuracy results with Lenet-5 and Fashion MNIST over 30 randomly intialized runs. Learning rate of 0.5 was used for all SGD and 0.01 for all NGD simulations. Gamma is 0.8 in all cases. Standard deviation of test accuracy is indicated in last column. Significant performance improvement observed in NGD compared to SGD (p-value $<$ 0.05 with two sample t-test)

TABLE IV
MEAN ACCURACY RESULTS WITH RESNET18 AND CIFAR10

optimizer	projective	test acc	train acc	stdev (test acc)
SGD	No	88.77%	92.13%	0.11
NGD	No	90.18%	95.808%	0.14
NGD	SVD	90.00%	95.00%	0.14

Mean accuracy results with Resnet18v1 and CIFAR10 over 10 randomly intialized runs. Learning rate of 0.1 was used for all SGD and 0.01 for all NGD simulations. Gamma is 0.7 for SGD and 0.8 for all NGD runs. Standard deviation of test accuracy is indicated in last column. Significant performance improvement observed in NGD compared to SGD (p-value $<$ 0.05 with two sample t-test)

random projections instead of orthogonal projections for the case of a very large number of parameters without much impact on the accuracy achieved. We finally perform the detailed complexity calculation and note that the algorithm described is asymptotically less complex than using the full rank KFAC algorithm. The utilization of the derived recursion technique further reduces the complexity for layers with a smaller number of parameters (non-asymptotic case). Hence a combination of projection and the recursive inverse can make it possible to utilize NGD based optimization for the larger class of networks with convolutional and fully connected layers with a large number of parameters. We also note that while we have utilized the random projection method for analyzing the complexity and some of the experiments, we have only touched upon this method. We foresee that there could be further implications and interactions of random projections with learning behavior which might be fruitful to explore in the future. We also derived a tractable method for obtaining the singular values of FIM without actually forming the Fisher matrix. We note that this method extends the possibility of statistical analysis of neural networks utilizing the tool of fisher information.

ACKNOWLEDGMENT

This work was supported by the Bharti School of Telecommunication Technology and Management, IIT Delhi.

TABLE V
PSNR RESULTS USING ESPCN AND BSD300 DATASE

optimizer	projective	PSNR	train loss	stdev (PSNR)
SGD	No	23.61%	3.9×10^{-3}	0.12
NGD	No	24.19%	3.9×10^{-3}	0.14
NGD	SVD	24.15%	3.9×10^{-3}	0.14

Mean PSNR results using ESPCN and BSD300 dataset over with 30 independently initialized runs. Learning rate of 0.1 and gamma of 0.7 used for all cases. Variance of PSNR show in last column. Significant performance improvement observed in NGD compared to SGD (p-value $<$ 0.05 with two sample t-test)

TABLE VI

VI. APPENDIX A: DETAILS OF COMPLEXITY CALCULATION

We assume the following variables.

- p: size of the minibatch
- q: input feature maps of the fully connected layer.
- r: output feature maps of the fully connected layer.

We note from (48) that dimension of matrices after projection can be assumed to be $klog(q)$ and $klog(r)$ respectively, for some constant k. We note the following matrix sizes for clarity.

- input feature map : $p \times q$
- filter : $q \times r$
- output fature map: $p \times r$
- Fisher information matrix : $(qr) \times (qr)$
- Ψ matrix : $q \times q$
- Γ matrix : $r \times r$
- $\hat{\Psi}$ matrix = $klog(q) \times klog(q)$
- $\hat{\Gamma}$ matrix = $klog(r) \times klog(r)$

The complexity calculation is as follows.

- Calculation of $\hat{\Psi}$ is of complexity: $(klog(q) * klog(q))$.
- Calculation of $\hat{\Gamma}$ is of complexity: $(klog(r) * klog(r))$.
- Complexity of projection is: $q^2(klog(q)^3 + r^2(klog(r))^3$
- Complexity of inversion of (assumed to be N^3 for square matrix of size N) :
  - $\hat{\Psi}$: $(klog(q))^3$
  - $\hat{\Gamma}$: $(klog(r))^3$
  - Total: $(klog(q))^3 + (klog(r))^3$
- Total Computations for calculating $\hat{\Psi}$, $\hat{\Gamma}$, projection and inversion is: $q^2 * (klog(q)^3 + r^2(klog(r))^3 + klog(p)^3 + klog(p)^3))$
- We note that for $f_1 = O(g_1)$ and $f_2 = O(g_2)$, $f_1 + f_2 = O(max(g1, g2))$. Utilizing this to simplify the order of complexity we get the following asymptotic complexity: $O(q^2 log(q)^3) + O(r^2 * (log(r)^3))$
- We note that when we use recursive inverse calculation, rest of the estimation remains same except that the Inverse calculation is of the order of mini-batch size. Hence The total compute is given by $q^2 + (klog(q))^3) + r^2 + (klog(r))^3 + 2 * klog(p)^3$ and order is same as for non recursive case.

2691-4581 (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 March 05,2022 at 06:03:03 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/TAI.2022.3153593, IEEE Transactions on Artificial Intelligence

VII. Appendix B: Hyper-parameter Selection

Neural network training performance can be dependent on hyperparameters used for training. And these training hyperparameters may need to be different for different optimization algorithms, e.g., using the same starting learning rate for SGD and NGD would give invalid performance metrics for comparing these algorithms. So we perform a grid search to select the optimal parameters for each of the optimization algorithms and use those for the particular algorithm. The parameters are selected in the following manner.

- We try all starting learning rates between 0.5 and 0.01 in multiplicative steps of 0.5 for each of the experiments. The best performing learning rate is selected. It is usually the maximum one for which loss does not diverge.
- We use a factor of gamma for step-wise reduction of learning rate after each epoch. This step is carried out by multiplying the learning rate by gamma after each epoch. We try various gamma values between 0.5 (halving) and 1 (no change) in additive increments of 0.1. There are other methods to decrease learning rates systematically, but we felt this method is adequate for comparison between algorithms.
- We usually update the projection matrix after several mini-batches. This we call projection period, and we use a value of 50 but also try to increase it to 100 and decrease it to 1. We notice that there is a minor impact on performance, with the smaller period being better for performance.
- Similar to the projection period, we also have a period (in terms of mini-batches elapsed) for inverse updates, and we check performance with an inverse period of 1, 10, 50, and 100 mini-batches, which has a minor impact on performance, with the smaller period being better. However, for PKFAC, we assume the recursive version, which effectively has an inverse period of 1, without a corresponding increase in computational complexity.
- We also utilize various batch sizes to see the impact on performance. With increased batch size, we need to increase the learning rate correspondingly, and the convergence is slower. However, we note that with a larger mini-batch, the projective algorithm is better in estimating the Fisher matrices. We try batch sizes of 50, 100, 300, and 600, finally selecting 100 for all of our results.

References

[1] S.-i. Amari, Neural learning in structured parameter spaces-natural riemannian gradient, in: Advances in neural information processing systems, 1997, pp. 127–133.
[2] S.-I. Amari, Natural gradient works efficiently in learning, Neural computation 10 (2) (1998) 251–276.
[3] S.-i. Amari, H. Nagaoka, Methods of information geometry, Vol. 191, American Mathematical Soc., 2007.
[4] L. Bottou, F. E. Curtis, J. Nocedal, Optimization methods for large-scale machine learning, Siam Review 60 (2) (2018) 223–311.
[5] S.-i. Amari, R. Karakida, M. Oizumi, Fisher information and natural gradient learning of random deep networks, arXiv preprint arXiv:1808.07172 (2018).
[6] S.-i. Amari, R. Karakida, M. Oizumi, Statistical neurodynamics of deep networks: geometry of signal spaces, Nonlinear Theory and Its Applications, IEICE 10 (4) (2019) 322–336.
[7] A. Papoulis, S. U. Pillai, Probability, random variables, and stochastic processes, Tata McGraw-Hill Education, 2002.
[8] J. Sohl-Dickstein, B. Poole, S. Ganguli, Fast large-scale optimization by unifying stochastic gradient and quasi-newton methods, in: International Conference on Machine Learning, 2014, pp. 604–612.
[9] H. L. Van Trees, Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory, John Wiley & Sons, 2004.
[10] W. Schmidt, R. Duin, Feed forward networks and the cramer-rao bound, in: [Proceedings 1992] IJCNN International Joint Conference on Neural Networks, Vol. 1, IEEE, 1992, pp. 646–651.
[11] T. M. Cover, Elements of information theory, John Wiley & Sons, 1999.
[12] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT press, 2016.
[13] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[14] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (8) (1997) 1735–1780.
[15] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Advances in neural information processing systems 27 (2014).
[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in neural information processing systems, 2017, pp. 5998–6008.
[17] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.
[18] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[19] C. Liu, B. Zoph, M. Neumann, J. Shlens, W. Hua, L.-J. Li, L. Fei-Fei, A. Yuille, J. Huang, K. Murphy, Progressive neural architecture search, in: Proceedings of the European conference on computer vision (ECCV), 2018, pp. 19–34.
[20] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: International Conference on Machine Learning, PMLR, 2019, pp. 6105–6114.
[21] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.
[22] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, in: Advances in neural information processing systems, 2015, pp. 91–99.
[23] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, G. Chen, et al., Deep speech 2: End-to-end speech recognition in english and mandarin, in: International conference on machine learning, 2016, pp. 173–182.
[24] H. Lu, M. Zhang, X. Xu, Y. Li, H. T. Shen, Deep fuzzy hashing network for efficient image retrieval, IEEE transactions on fuzzy systems 29 (1) (2020) 166–176.
[25] R. S. Sutton, A. G. Barto, Reinforcement learning: An introduction, MIT press, 2018.
[26] H. Lu, Y. Li, S. Mu, D. Wang, H. Kim, S. Serikawa, Motor anomaly detection for unmanned aerial vehicles using reinforcement learning, IEEE internet of things journal 5 (4) (2017) 2315–2322.
[27] H. Lu, Y. Zhang, Y. Li, C. Jiang, H. Abbas, User-oriented virtual mobile network resource management for vehicle communications, IEEE transactions on intelligent transportation systems (2020).
[28] T. D. Barfoot, Multivariate gaussian variational inference by natural gradient descent, arXiv preprint arXiv:2001.10025 (2020).
[29] E. Benhamou, D. Saltiel, B. Guez, J. Atif, R. Laraki, From forecast to decisions in graphical models: A natural gradient optimization approach, Available at SSRN 3813403 (2021).
[30] W. Lin, F. Nielsen, M. E. Khan, M. Schmidt, Tractable structured natural gradient descent using local parameterizations, arXiv preprint arXiv:2102.07405 (2021).
[31] B. Sánchez-López, J. Cerquides, On the convergence of stochastic process convergence proofs, Mathematics 9 (13) (2021) 1470.
[32] J. Martens, R. Grosse, Optimizing neural networks with kronecker-factored approximate curvature, in: International conference on machine learning, 2015, pp. 2408–2417.
[33] R. Grosse, J. Martens, A kronecker-factored approximate fisher matrix for convolution layers, in: International Conference on Machine Learning, 2016, pp. 573–582.

2691-4581 (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 March 05,2022 at 06:03:03 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/TAI.2022.3153593, IEEE Transactions on Artificial Intelligence

[34] J. Martens, New insights and perspectives on the natural gradient method, arXiv preprint arXiv:1412.1193 (2014).
[35] T. George, C. Laurent, X. Bouthillier, N. Ballas, P. Vincent, Fast approximate natural gradient descent in a kronecker factored eigenbasis, in: Advances in Neural Information Processing Systems, 2018, pp. 9550–9560.
[36] J. Lee, H. G. Hong, D. Joo, J. Kim, Continual learning with extended kronecker-factored approximate curvature, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 9001–9010.
[37] D. Goldfarb, Y. Ren, A. Bahamou, Practical quasi-newton methods for training deep neural networks, arXiv preprint arXiv:2006.08877 (2020).
[38] Y. Ren, D. Goldfarb, Kronecker-factored quasi-newton methods for convolutional neural networks, arXiv preprint arXiv:2102.06737 (2021).
[39] K. Osawa, Y. Tsuji, Y. Ueno, A. Naruse, R. Yokota, S. Matsuoka, Large-scale distributed second-order optimization using kronecker-factored approximate curvature for deep convolutional neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 12359–12367.
[40] K. Osawa, Y. Tsuji, Y. Ueno, A. Naruse, C.-S. Foo, R. Yokota, Scalable and practical natural gradient for large-scale deep learning, arXiv preprint arXiv:2002.06015 (2020).
[41] P. Kaul, B. Lall, Projective kronecker factored approximate curvature : Implementation in pytorch, https://github.com/piyushkaul/information_geometry (2020).
[42] Y. Choquet-Bruhat, C. deWitt Morette, M. Dillard-Bleick, Analysis, manifolds and physics (revised edition), Elsevier 35 (1982) 36.
[43] J. M. Lee, Riemannian manifolds: an introduction to curvature, Vol. 176, Springer Science & Business Media, 2006.
[44] P. Kaul, B. Lall, Riemannian curvature of deep neural networks, IEEE Transactions on Neural Networks and Learning Systems 31 (4) (2019) 1410–1416.
[45] S. Dasgupta, A. Gupta, An elementary proof of the johnson-lindenstrauss lemma, International Computer Science Institute, Technical Report 22 (1) (1999) 1–5.
[46] G. E. Hinton, R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks, science 313 (5786) (2006) 504–507.
[47] Y. LeCun, C. Cortes, MNIST handwritten digit database, http://yann.lecun.com/exdb/mnist/ (2010) [cited 2016-01-14 14:24:11]. URL http://yann.lecun.com/exdb/mnist/
[48] G. E. Hinton, A practical guide to training restricted boltzmann machines, in: Neural networks: Tricks of the trade, Springer, 2012, pp. 599–619.
[49] H. Xiao, K. Rasul, R. Vollgraf, Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms, arXiv preprint arXiv:1708.07747 (2017).
[50] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (11) (1998) 2278–2324.
[51] A. Krizhevsky, G. Hinton, et al., Learning multiple layers of features from tiny images (2009).
[52] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1874–1883.
[53] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, Vol. 2, IEEE, 2001, pp. 416–423.



Piyush Kaul received the B.E. degree in electronics and communication engineering from the Delhi College of Engineering, Delhi University, Delhi, in 1999. He has been studying for a Ph.D. degree in Electrical Engineering from the Indian Institute of Technology, Delhi, since 2016. Since 2013, he has been working as an architect in the system engineering team (IPG) at Cadence Design Systems for developing machine learning and artificial intelligence systems based on Deep Neural Networks with a specific focus on computer vision. His primary research interest has been in the areas of mathematical characterization and understanding of Deep Neural Networks.



Brejesh Lall (received the B.Tech. degree in electronics and communication engineering, the master's degree in signal processing from the Delhi College of Engineering, New Delhi, India, and the Ph.D. degree in signal processing from the Indian Institute of Technology (IIT) Delhi, New Delhi, India, in 1999. He is currently a Professor with the Department of Electrical Engineering, IIT Delhi. In 1997, he joined Aricent. In 2005, he joined IIT Delhi as an Assistant Professor. He is currently a Professor at IIT Delhi, and actively working in the area of image processing, signal processing, and communication. His research interests include object representation, tracking and classification, odometry, depth map generation, representation and rendering, vector sensor-based underwater acoustic communications, characterization, and analysis of Neural Networks, and performance issues in molecular communications.

2691-4581 (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 March 05,2022 at 06:03:03 UTC from IEEE Xplore. Restrictions apply.

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.

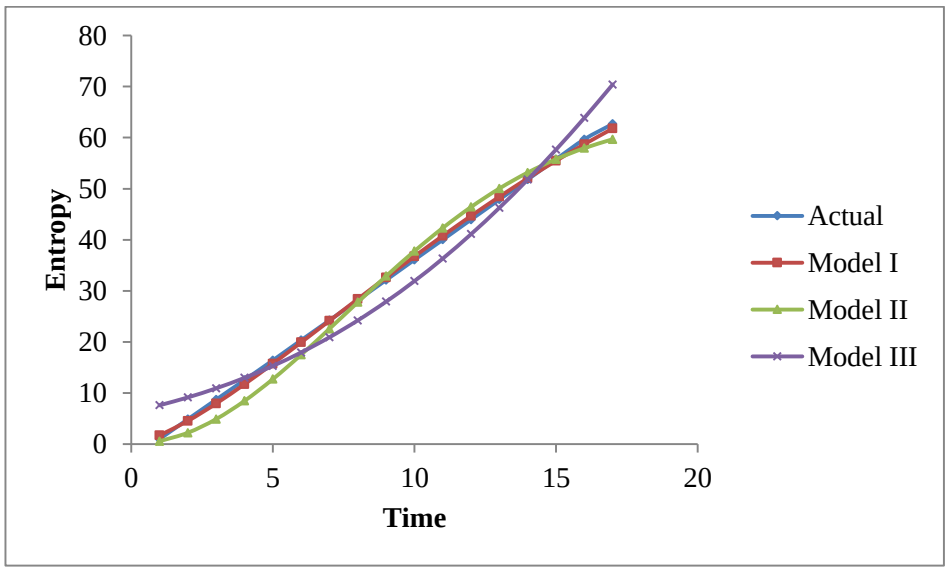


Figure 1. Goodness fit curve for Data Set-I.

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

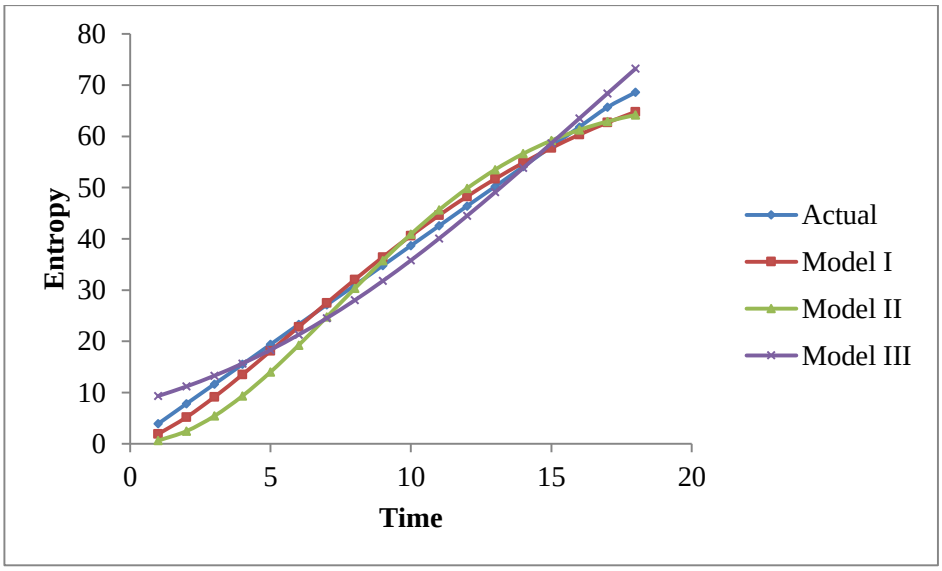


Figure 2. Goodness fit curve for Data Set-II.

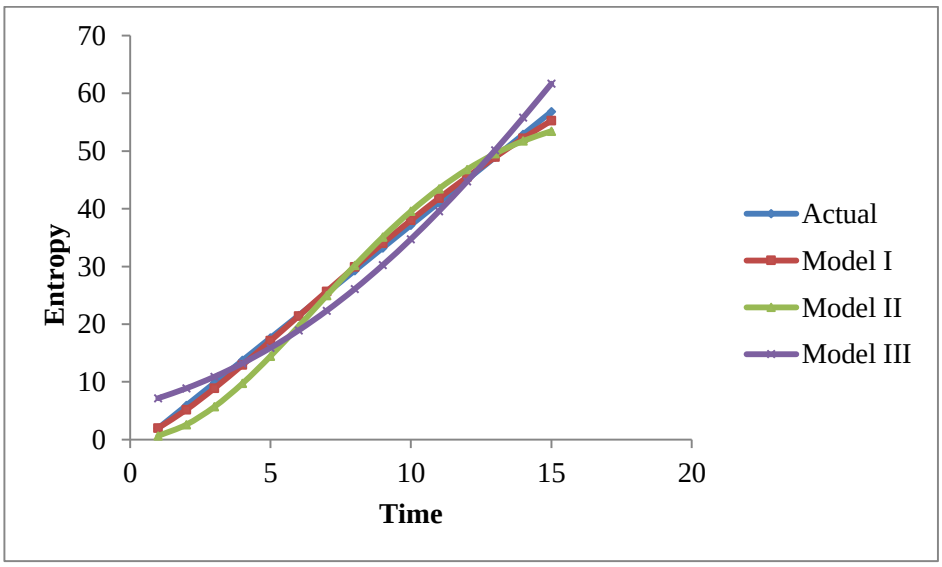


Figure 3. Goodness fit curve for Data Set-III.

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/

© 2021. This work is published under https://creativecommons.org/licenses/by/4.0/ (the “License”). Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/358899239

Sentiment Analysis using Scikit Learn: A Review

Preprint · February 2022

DOI: 10.13140/RG.2.2.26189.10720

CITATIONS
0

READS
25

2 authors:



Divij Gera
Delhi Technological University
1 PUBLICATION **0** CITATIONS
SEE PROFILE



Amita Kapoor
University of Delhi
71 PUBLICATIONS **193** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Pattern Recognition Using Neural Networks View project

Exploration of Quantum Neural Networks View project

All content following this page was uploaded by Amita Kapoor on 28 February 2022.

The user has requested enhancement of the downloaded file.

Sentiment Analysis using Scikit Learn: A Review

Divij Gera[1], Amita Kapoor[2]

Delhi Technological University, Delhi, India. SRCASW ,University of Delhi, Delhi, India.

geradivij@gmail.com[1], dr.amita.kapoor@ieee.org[2]

***Abstract*— Sentiment Analysis is defined as an approach to Natural Language Processing that aims to identify the emotional tone behind a body of text. This paper covers various traditional as well as newer state-of-the-art machine learning techniques to perform sentiment analysis on movie reviews. It covers various pre-processing and classification techniques used and compares their performance on binary and multi-class movie review datasets from popular movie review websites. It thereby concludes that Support Vector Classifiers with word embedding vectorization outperforms the other classification techniques considered.**

***Keywords*— Sentiment Analysis, NLP, Classification, Reviews, ML**

I. INTRODUCTION

Sentiment Analysis is a common text classification tool used to analyze text and classify its underlying sentiment into categories such as positive, negative, and neutral. It has a variety of applications across industries like marketing, customer support, product analysis and political campaigns.

Sentiment analysis systems can be applied in almost every business and social domain because opinions are central to almost all human activities and are key influencers of our behavior. Moreover, we are living in an age where we have the availability of large amounts of opinionated data available for analysis through various websites and social media platforms.

This paper summarizes the various Machine Learning and Deep Learning techniques used to classify text sentiment in movie reviews from popular websites IMDb and Rotten Tomatoes. The techniques have been implemented on one of two datasets: Stanford's Large Movie Review Dataset (A binary classification dataset from IMDb) and The Rotten Tomatoes Movie Review Dataset (A multi-class dataset from Rotten Tomatoes), both of which are covered in detail in the following section.

II. DATASETS USED

Two movie review datasets were used to classify movie review sentiment. The two were in stark contrast to each other:

A. Stanford's Large Movie Review Dataset

The large movie review dataset from Stanford's AI lab [1] was used for binary text classification. The dataset contains 50,000 distinct and highly polar movie reviews for training and testing from the popular movie review website, IMDb. The reviews were equally split into positive and negative reviews. Reviews that are accompanied by a rating of 7 or higher were considered as Positive and the remaining were taken to be Negative.
Target labels: 0 represents negative reviews and 1 represents positive reviews

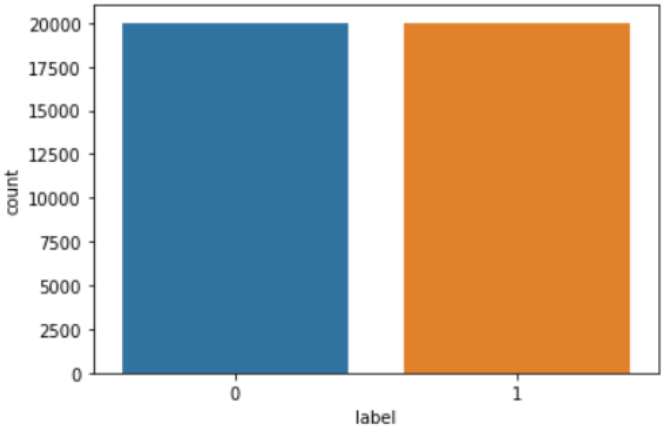


Figure 1: Sentiment Distribution of Binary Data

Classification Techniques Implemented: Naive-Bayes Classification, Logistic Regression, Support Vector Classification

B. Rotten Tomatoes Movie Review Dataset

Dataset for Multi-class Text classification, originally collected by Pang and Lee [2]. Classifies text into 5 categories: Negative, Somewhat Negative, Neutral, Somewhat Positive and Positive

- Consists of 8544 reviews from the Rotten Tomatoes website divided into 156060 phrases
- The dataset is very unevenly split between the 5 categories with most phrases being classified as Neutral
- Target labels are numbered 0-4 from Negative to Positive

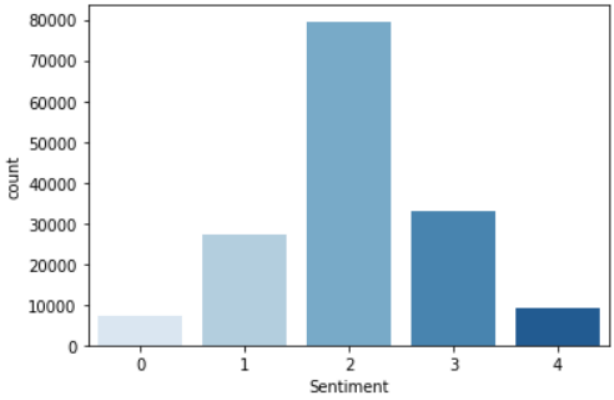


Figure 2: Sentiment Distribution of Multi-Class Data

Classification Techniques Implemented: Naive-Bayes Classification, Support Vector Classification, Word embeddings, Recurrent Neural Network, Bidirectional Encoder Representation of Transformer (BERT)

III. EXPLORATORY DATA ANALYSIS AND PRIMARY PREPROCESSING

A. IMDb Large Movie Review Dataset

The dataset consisted of 2 columns, namely the 'review' and the 'label'. No null values were present in the data. The data was relatively clean with the exception of line breaks and punctuation marks. There was a perfectly even distribution of data between the two classes. On average, the movie reviews were around 230 words long with an upper extreme of 1000 words. The length of the review seemed to have no correlation with the review sentiment. On the other hand, there did exist some level of correlation between the polarity of the reviews and the underlying sentiment (Polarity was obtained using Python's TextBlob library).

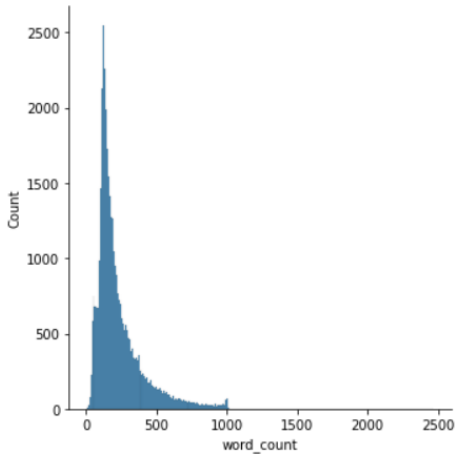


Figure 3: Word Count of IMDb Movie Reviews

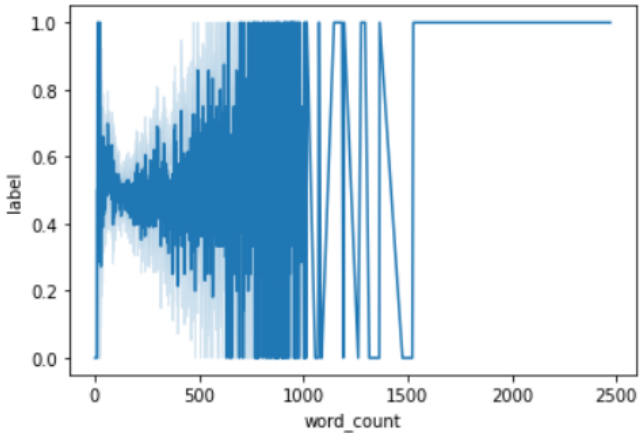


Figure 4: Word Count vs Sentiment for IMDb Reviews

	polarity	label
polarity	1.000000	0.559325
label	0.559325	1.000000

Figure 5: Sentiment vs Polarity correlation on IMDb Data

In order to pre-process the text for model training, line breaks and punctuation marks were removed using RegEx functions.

Functions were defined and used to expand contractions and remove numbers from the text. This was done because numbers do not influence the underlying sentiment of the text, so removing the numbers and expanding on the contractions reduces the text complexity as well as the size of the corpus. The text was also converted to lowercase for the latter purpose.

Since all 3 techniques used on this dataset treat words as independent features rather than with semantic context as part of a whole sentence, stopwords (Commonly used words that do not add much to the sentence such as 'and', 'not', 'it', 'the' etc.) were removed from the reviews.

B. Rotten Tomatoes Movie Review Dataset

The original dataset consisted of 4 columns, namely 'SentenceId', 'PhraseId', 'Phrase' and 'Sentiment'. No null values were present in the data. From the data distribution, the mean sentiment was 2 (neutral) and the standard deviation is 0.89 (<1) which implied an uneven distribution between sentiments with a higher proportion of reviews and phrases near neutral as compared to extreme values.

The average word count of reviews in this dataset was 7.02, with the longest phrase comprising 52 words. This is significantly lower than that of the IMDb dataset. There seemed to be no real relationship between the word count and the resulting sentiment, except for the fact that phrases with extremely low word counts did skew towards a more negative sentiment. Similarly, no correlation was observed between the polarity and the phrase sentiment.

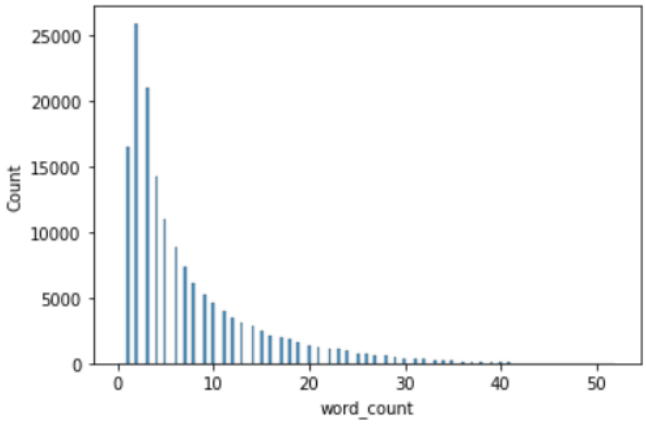


Figure 6: Word Count of Rotten Tomatoes Dataset

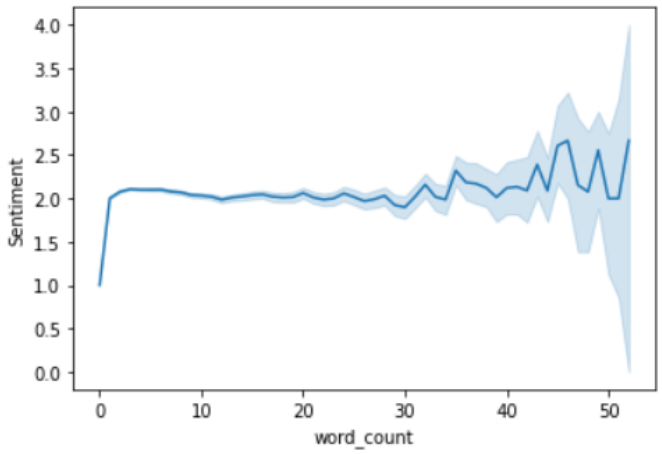


Figure 7: Word Count vs Sentiment for Rotten Tomatoes Reviews

	polarity	Sentiment
polarity	1.000000	0.373277
Sentiment	0.373277	1.000000

Figure 8: Sentiment vs Polarity correlation on Rotten Tomatoes Dataset

Due to the uneven distribution and existence of multiple classes, we could not rely on absolute accuracy as an evaluation metric and fell back on F1 score instead, which is calculated as:

$$F_1\text{–score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

where,
TP = True Positive
FP = False Positive
FN = False Negative
Similar to the previous dataset, numbers and punctuations were removed from the dataset and contractions were expanded. Stratification was introduced in the test train split to ensure equal class proportion in training and validation sets.

IV. CLASSIFICATION TECHNIQUES ON BINARY DATA

Naive Bayes Classifier on Binary Data

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. Bayes Theorem is a mathematical formula for determining the likelihood of an outcome occurring, based on a previous outcome. Bayes Theorem gives conditional probability as:

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)}$$

Naive Bayes classifier holds the fundamental assumption that each feature makes an independent and equal contribution to the outcome. The conditional probability formula varies with the type of Naive Bayes classifier used.

1. Gaussian Naive-Bayes assumes a normal distribution of features and conditional probability is given by:

$$p(x = v \mid C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}}\, e^{-\frac{(v-\mu_k)^2}{2\sigma_k^2}}$$

2. Multinomial Naive-Bayes has feature vectors which represent events generated by a multinomial distribution. These are generally used for document classification. Conditional probability is given by

$$p(\mathbf{x} \mid C_k) = \frac{(\sum_{i=1}^{n} x_i)!}{\prod_{i=1}^{n} x_i!} \prod_{i=1}^{n} {p_{ki}}^{x_i}$$

3. Bernoulli Naive-Bayes includes features as independent booleans describing inputs. This is also a popular choice for document classification. Conditional probability is given by:

$$p(\mathbf{x} \mid C_k) = \prod_{i=1}^{n} p_{ki}^{x_i}(1 - p_{ki})^{(1-x_i)}$$

All of the reviews in the dataset were converted to a document format. Each review/document is a tuple wherein the first item is a list of words in the review in order and the second item is the target label of the review.

Additionally, a corpus of all words was built and their frequencies in the dataset and generated the following WordCloud to demonstrate the most commonly occurring words (Not including stop-words).

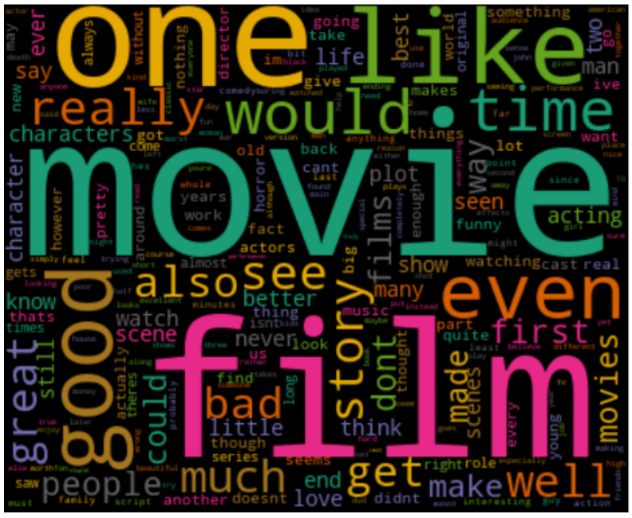


Figure 9: WordCloud for IMDb Data

The aforementioned corpus of words was then used to convert each word of the corpus to a feature of the reviews. These act as the data points for classification which would have True/False values.

These were then used to form training and testing sets and hence build the Naive-Bayes Classifier from Python's NLTK library.

Following this, the original training and testing review arrays were converted to vectors using the TF-IDF Vectorizer. This was done because the vectorizer takes into account the most commonly used words without giving a huge importance to stop words like 'and' and 'the' so it picks up the important features from the text as well as their relative importance.

These vectors were then used to train and validate a Bernoulli Naive-Bayes Classifier, which is a variation of the classifier that works with binary data.

Logistic Regression on Binary Data

Logistic Regression [3] is a classification algorithm used for predicting the categorical dependent variable using a given set of independent variables. The classifier provides the probability of an event taking place, given an input value. The probability of a given dependent variable being 1 is given using the logit function.

$$Y_{pred} = P(y = 1|X = x) = \frac{1}{1 + exp(-(b + W^T X))}$$

The training data is used to determine optimum weight and bias in a logit (or sigmoid) function. This function is then used to determine the probabilities of the new input belonging to each of the possible classes. A threshold value is chosen, and the input is classified to the class having event probability greater than the threshold value.

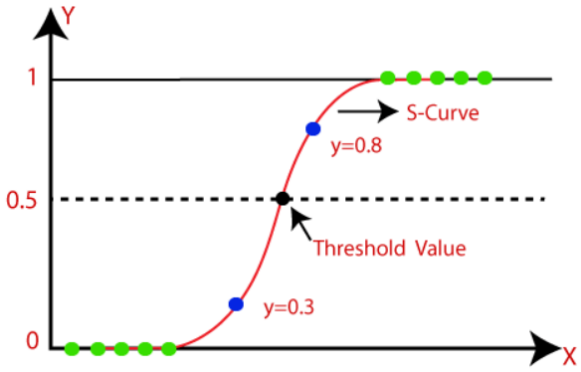


Figure 10: Logistic Regression Classifier [4]

Scikit-Learn library's Logistic Regression Classifier was trained using the documentized featuresets built for the Naive-Bayes Classifier while setting the max-iteration limit to 1000 to save training time.

Support Vector Classifier on Binary Data

Support Vector Machine (SVM) [5] is a non-parametric binary classification algorithm that creates a best separating hyperplane between the two classes. It works as a 'maximum-margin separator' ie. it works to create a decision boundary with the largest possible distance to example points from either class. Fig 11 represents an example of an SVM and its hyperplane.

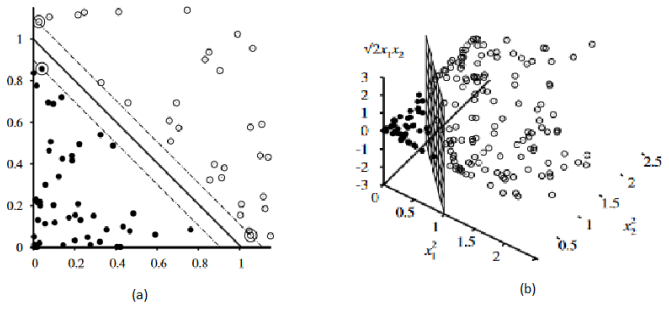


Figure 11: SVM with its hyperplane[5]

Using Python's Scikit-Learn library, a Linear Support Vector Classifier was trained using the independent word featuresets, setting max iterations to 5000.

Summary of Binary Classifiers

The following table represents the performance of various classification and Sentiment Analysis techniques on the IMDb Dataset:

TABLE 1: SUMMARY OF CLASSIFIER ACCURACY ON BINARY DATA

S. no.	**Classifiers on Binary Data**				
	Classifier	**Accuracy (%)**	**Precision**	**Recall**	**F1 Score**
1.	NLTK Naive-Bayes Classifier	85.44	0.84	0.87	0.85
2.	Scikit-Learn Bernoulli Naive-Bayes	86.06	**0.87**	0.84	0.86
3.	Scikit-Learn Logistic Regression (with 'saga' optimizer)	**87.16**	0.86	**0.88**	**0.87**
4.	Scikit-Learn Linear Support Vector Classifier	86.20	0.85	0.87	0.86

Accuracy has been chosen as the metric for evaluation since there are only 2 classes and data is equally divided among them.

Naive-Bayes Classifier on Multi-Class Data

Similar to the Binary Naive-Bayes Classifier, data was cleaned, stopwords were removed and a WordCloud was generated to show the most commonly occurring features in the corpus. (Note: This dataset had a far smaller corpus size). This WordCloud was almost identical to the previous one.


one
film
movie
like
story
nt
even
much
rrb
good
time
love
made
bad
never
flrb
enough
two
make
life
characters
little
comedy
funny
people
action
could
makes
films
director
way
movies
something
new
would
work
plot
performances
character
great
cast
another
many
may
best
us
really
still
well
better
drama
minutes
every
humor
sense
audience
world
history
without
nothing
big
hard
look
first
least
seen
watch
see
entertainment
romantic
thriller
hollywood
human
man
feel
real
bit
kids
women
show
cinema
performance
almost
actually
long
comes
rather
point
takes
music
comic
art
place
kind
interesting
material
documentary
dialogue
dark
family
fun
keep
star
moving
sweet
back
american
original
moments
years
direction
making
worth
lot
less
feels
young
subject
piece
special
watching
go
works
thriller
horror
genre
actors
screen
full
though
might
ever
give
pretty
visual
scenes
script
know
going
old
seen
also
think
gets
find
tale
quite
acting
seems
times
whole
part
come
ga
documentary
anything
cgi
effects
lot
scene
moment
sometimes
interesting
fun
keep
star
moving
big
hard
kind
dialogue
dark
family
material
another
many
nothing
art
place
hollywood
comic
music
takes
point
rather
full
screen
least
first
seen
watch
see
entertainment
romantic
thriller
horror
genre
actors
worth
making
direction
years
sweet
back
american
original
moments
often
humor
sense
audience
picture
world
history
subject
piece
special
watching
go
works
young
feels
lot
less
minutes
every
want
humor
sense
romantic

An example of a pre trained word vector is:

```
embeddings_dict['man']
```

```
array([-0.094386,  0.43007 , -0.17224 , -0.45529 ,  1.6447  ,  0.40335 ,
       -0.37263 ,  0.25071 , -0.10588 ,  0.10778 , -0.10848 ,  0.15181 ,
       -0.65396 ,  0.55054 ,  0.59591 , -0.46278 ,  0.11847 ,  0.64448 ,
       -0.70948 ,  0.23947 , -0.82905 ,  1.272   ,  0.033021,  0.2935  ,
        0.3911  , -2.8094  , -0.70745 ,  0.4106  ,  0.3894  , -0.2913  ,
        2.6124  , -0.34576 , -0.16832 ,  0.25154 ,  0.31216 ,  0.31639 ,
        0.12539 , -0.012646,  0.22297 , -0.56585 , -0.086264,  0.62549 ,
       -0.0576  ,  0.29375 ,  0.66005 , -0.53115 , -0.48233 , -0.97925 ,
        0.53135 , -0.11725 ], dtype=float32)
```

From the pre-trained vector file, a dictionary was built to store the word and the corresponding vector.

This dictionary was then used to convert each phrase in the dataset to the sum of vectors of its words. Stopwords were included in this data since they do add a level of semantic meaning to the overall text. A simple SVC was then trained using these text vectors.

SVC was chosen for this purpose since it had shown to be the best performing classifier on this dataset on traditionally vectorized data from the Rotten Tomatoes Dataset.

One of the problems faced by this model was that a number of words did not contain the necessary whitespace and hence could not be included in the text vectors since they were unrecognizable. For example, 'selfglorification', 'comedydrama' and 'martialarts' amongst others.

Recurrent Neural Networks with Embeddings

Recurrent Neural Networks (RNNs) [7] are a kind of neural network that can incorporate sequential dependence in data by having a hidden state or memory that holds the essence of what has been seen by the model so far.

In each layer of an RNN, we consider the current input with the output associated with the previous state. This process is repeated while input is fed to the network and hence allows it to retain temporal data.

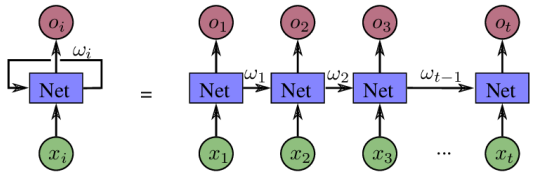


Figure 13: Unfolded Recurrent Neural Network [8]

A drawback of RNNs is the vanishing gradient problem. In the case where the individual gradients of a hidden state with respect to the previous stage is less than 1, the product of these gradients gets smaller and smaller, leading to a vanishing gradient.

Long short-term memory networks (LSTMs) are a variant of recurrent neural networks. While simple RNNs combine the hidden state from the previous layer and the current input through a *tanh* activation, LSTMs implement recurrence using four layers interacting in a specific manner (rather than a single LSTM layer). This is demonstrated below:

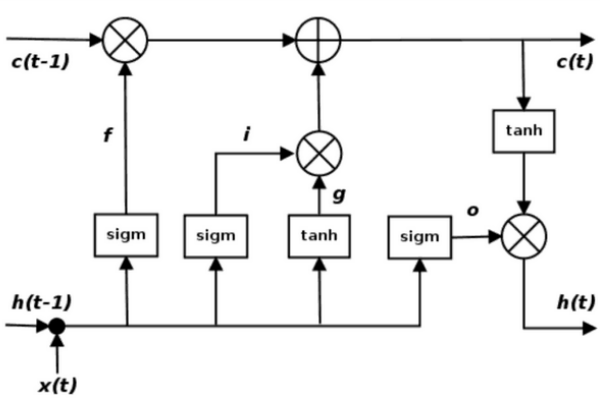


Figure 14: LSTM Architecture[8]

For this model, spaCy's [9] pre-trained word embeddings were used. spaCy was chosen in particular due to it's relatively faster speed of transfer learning with its pre-trained models. 96-Dimensional word vectors were used for this purpose.

The Keras text vectorization layer was placed before the embedding layer. It allows documents of various sizes to be passed to the model. The TextVectorization layer tokenizes, vectorizes, and pads sequences representing those documents to be passed to the embedding layer. One max pooling layer and 2 hidden dense layers were added to the model. The number of layers and nodes per layer chosen were random.

The network uses the 'Adam' optimizer and relies on 'categorical cross-entropy' as the loss function.

The phrases were passed into the model with stopwords intact and the model was trained over 45 epochs.

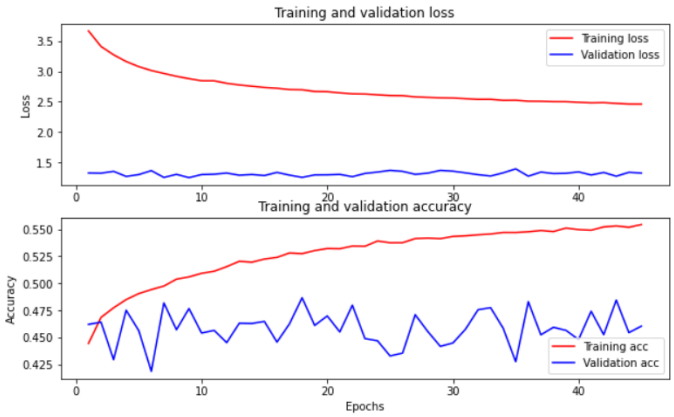


Figure 15: Accuracy and Loss on RNN

The network architecture is demonstrated below:-

Layer (type)	Output Shape	Param #
text_vectorization (TextVect	(None, None)	0
embedding (Embedding)	(None, None, 96)	1574784
lstm_2 (LSTM)	(None, None, 25)	12200
global_max_pooling1d_2 (Glob	(None, 25)	0
dropout_4 (Dropout)	(None, 25)	0
dense_6 (Dense)	(None, 32)	832
dropout_5 (Dropout)	(None, 32)	0
dense_7 (Dense)	(None, 32)	1056
dense_8 (Dense)	(None, 5)	165

Total params: 1,589,037
Trainable params: 14,253
Non-trainable params: 1,574,784

BERT Model on Multi-Class Data

BERT makes use of Transformer encoder architecture [10], based on an attention mechanism that learns contextual relations between words (or sub-words) in a text. Therefore, it is considered bidirectional, though it would be more accurate to say that it's non-directional. This characteristic allows the model to learn the context of a word based on all of its surroundings (left and right of the word).

Once a sentence is fed as input to the BERT model, it understands the context of each word in the sentence using its multi-head attention mechanism and returns its contextual representation in that sentence as vector embedding.

In order to build a Sentiment Classifier using BERT encoding, TensorFlow's pretrained BERT layer [11] was included in a feed-forward neural network in combination with a text pre-processing layer and a Dense output layer. Uncased small BERT model was used for this purpose with the following details:

No. of encoder layers(L): 4
Hidden(H): 512
No. of attention heads(A): 8

For training, the model utilizes the 'Adam' optimizer and 'Categorical Cross-entropy' as loss and was trained over 15 epochs.

The following is the structure of the feed forward neural network built:

Layer (type)	Output Shape	Param #
text (InputLayer)	[(None,)]	0
preprocessing (KerasLayer)	{'input_word_ids': (	0
BERT_encoder (KerasLayer)	{'pooled_output': (N	28763649
dropout (Dropout)	(None, 512)	0
hidden32 (Dense)	(None, 32)	16416
dropout_1 (Dropout)	(None, 32)	0
hidden32-2 (Dense)	(None, 32)	1056
classifier (Dense)	(None, 5)	165

Total params: 28,781,286
Trainable params: 17,637
Non-trainable params: 28,763,649

Summary of Models on Multi-Class Data

The following table represents the performance of various models and techniques on the multi-class Rotten Tomatoes Movie Review Dataset. The default metric used to test the classifiers for this dataset was the F1 score (formula mentioned below). This was chosen because the uneven nature of the dataset made accuracy an untrustworthy metric of classification, since most algorithms would be biased in favour of the majority class:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{\text{TP}}{\text{TP} + \frac{1}{2}(\text{FP} + \text{FN})}$$

Here, TP = No. of True Positive results
FP = No. of False Positive results
FN = No. of False Negative results

TABLE 2: SUMMARY OF CLASSIFIER METRIC ON MULTI-CLASS DATA

S. no.	Classifiers on Multi-Class Data			
	Classifier	**F1 Score**	**Precision**	**Recall**
1.	Scikit-Learn Bernoulli Naive-Bayes (without class weights)	[0.31 0.39 0.73 0.43 0.34]	[0.40 0.50 0.66 0.54 0.43]	[0.24 0.35 0.85 0.40 025]
2.	Scikit-Learn Multinomial Naive-Bayes (without class weights)	[0.06 0.35 0.73 0.45 0.09]	[0.61 0.51 0.62 0.53 0.59]	[0.05 0.28 0.89 0.41 0.06]
3.	Scikit-Learn Bernoulli Naive-Bayes (with class weights)	[0.39 0.46 0.70 0.49 0.43]	[0.29 0.46 0.75 0.51 0.34]	[0.59 0.40 0.70 0.43 0.62]
4.	Scikit-Learn Multinomial Naive-Bayes (with class weights)	[0.32 0.42 0.44 0.45 0.36]	[0.22 0.38 0.75 0.38 0.26]	[0.61 0.48 0.31 0.56 0.58]

5.	Scikit-Learn Support Vector Classifier (without class weights)	[0.32 0.48 0.76 0.54 0.35]	**[0.55 0.52 0.68 0.56 0.61]**	[0.21 0.42 0.85 0.49 0.23]
6.	Scikit-Learn Support Vector Classifier (with class weights)	**[0.41 0.46 0.70 0.48 0.48]**	[0.30 0.44 0.80 0.51 0.37]	**[0.66 0.49 0.63 0.46 0.67]**
7.	Scikit-Learn Support Vector using GloVe word embeddings	[0.33 0.42 0.65 0.44 0.40]	[0.19 0.35 0.78 0.42 0.28]	[0.58 0.37 0.60 0.31 0.63]
8.	Recurrent Neural Network (with LSTM)	[0.39 0.40 0.64 0.42 0.46]	[0.30 0.33 0.79 0.38 0.38]	[0.58 0.45 0.49 0.48 0.55]
9.	Feed forward neural network with BERT embeddings	[0.29 0.36 0.66 0.31 0.40]	[0.17 0.32 0.80 0.50 0.31]	[0.62 0.37 0.60 0.29 0.58]

VI. CONCLUSIONS

Completing Sentiment Analysis testing and validation on various classifiers allows us to derive the following conclusions:

- Since the IMDb dataset was evenly split and consisted of relatively clean text data, all classification models performed well on it and gave consistent results with a high degree of accuracy.
- In contrast, the Rotten Tomatoes dataset contained text that needed more pre-processing, and the data was very unevenly divided amongst the classes. Hence, most classification models gave poor to moderate results on this data. This dataset is closer to a real-world situation where sentiment analysis is to be used.
- Models that considered data semantics and the relationships between words performed much better than those which considered the words to be independent features.
- Among modern classifiers RNN and BERT models gave reasonably accurate results, with the Neural Network being the better performer.
- Support Vector Classifiers proved to be the best performing classifiers on multi-class data while Logistic Regression Classifier had the best performance on binary data.

AUTHOR'S CONTRIBUTIONS

The work was done as an undergraduate project by Divij Gera (DG) under the supervision of Amita Kapoor (AK). AK conceived and designed the analysis, provided consistent and valuable feedback and approved the final manuscript. DG collected data, performed the analysis and wrote the paper.

REFERENCES

[1] Association for Computational Linguistics. Meeting. (2011). *The 49th annual meeting of the association for computational linguistics: Human language technologies : ACL HLT 2011 : Proceedings of tutorial abstracts : 19 June 2011, Portland, Oregon, USA*.

[2] Pang and L. Lee. 2005. *Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales*. In ACL, pages 115–124.

[3] Kapoor, A. (2019). *Hands-on artificial intelligence for IoT: Expert machine learning and deep learning techniques for developing smarter IoT systems*. Packt Publishing.

[4] Logistic regression in machine learning - javatpoint. www.javatpoint.com. (n.d.). Retrieved November 21, 2021, from https://www.javatpoint.com/logistic-regression-in-machine-learning.

[5] Russell, S., & Norvig, P. (2016). *Artificial intelligence: A modern approach* (3rd ed.). Createspace Independent Publishing Platform.

[6] Pennington, J., Socher, R., & Manning, C. (n.d.). *GloVe: Global Vectors for Word Representation*. Glove: Global vectors for word representation. Retrieved November 9, 2021, from https://nlp.stanford.edu/projects/glove/.

[7] Gulli Antonio, et al. "Recurrent Neural Networks." Deep Learning with Tensorflow 2 and Keras Regression, Convnets, Gans, RNNS, NLP, and More with Tensorflow 2 and the Keras API, Packt, Birmingham, 2019.

[8] *Tai, Lei & Liu, Ming. (2016) "Fig 3". Deep-learning in Mobile Robotics - from Perception to Control Systems: A Survey on Why and Why not.*

[9] *Spacy 101: Everything you need to know · spacy usage documentation*. spaCy 101: Everything you need to know. (n.d.). Retrieved November 9, 2021, from https://spacy.io/usage/spacy-101.

[10] Ravichandiran, S. (2021). *Getting Started with Google BERT*. Packt Publishing.

[11] *Classify text with BERT | Text | TensorFlow*. TensorFlow. (2021). Retrieved 9 November 2021, from https://www.tensorflow.org/text/tutorials/classify_text_with_bert.

View publication stats

Proceedings of the Fourth International Conference on Smart Systems and Inventive Technology (ICSSIT-2022)
IEEE Xplore Part Number: CFP22P17-ART; ISBN: 978-1-6654-0118-0

Stator Inter-Turn Fault Diagnosis by Motor Current Signature Approach

Kuldeep Kumar Swarnkar
Ph.D Scholar
Delhi Technological University Delhi, India

Prof. J.N.Rai
Professor, EED
Delhi Technological University Delhi, India

Dr. Sulochana Wadhwani
Professor, EED
Madhav Institute of Technology & Science Gwalior, India

2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT) | 978-1-6654-0118-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICSSIT53264.2022.9716329

Abstract

Induction machines are widespread workhorse of industries because of their toughness, smaller size, efficiency, low upkeep, and cost-effective performance. Condition monitoring of these motors is essential to avoid production shutdown, personnel injuries, and wastage of materials. Induction motors are subjected to a variety of environmental conditions in industries. During their lifetime, they need to expose themselves to different abnormal conditions. The safe and secure operations of the industries rely on the protection of these Induction motors. The abnormal condition may be due to accidental short circuit and/or the deterioration of organic insulation due to heat and aging. The focus of research in this paper lies in identifying the suitable methodology for stator inter-turn fault in induction machines. So these faults are created in real-time and the stator current signatures are taken for further analysis. Various signal processing techniques are effectively used in many engineering applications. Hence, in this paper, application of signal processing techniques on motor current signatures is also being taken up. This method can be effectively implemented for various types of faults to prevent shutdowns.

Keywords: Motor Current Signature Analysis, Stator winging Fault.

1. Introduction

Electrical machines play a very essential role in the manufacturing and production industries. Among various electrical machines, the induction motor is a versatile workhorse of industrial prime movers. They are preferred because of their rugged design, small size, reliability, and low cost. Induction motors are subjected to a variety of environmental conditions in industries. During their lifetime, they need to expose themselves to different abnormal conditions. The safe and secure operations of the industries rely on the protection of these electrical machines. Figure 1 shows 3-Φ Induction motors faults. Figure 2 shows all types of faults of 3-Φ Induction motors and their percentage.

The abnormal condition may be due to the accidental short circuit and/or the deterioration of organic insulation due to heat and aging. The damages to insulation will lead to incipient faults in the system; also, if left unnoticed consequently it results in the breakdown of the machine. The cost of induction motor and its repair and refurbishment expenses might not be substantial, but cost associated with downtime of the motor is enormous for an industry. Therefore, to prevent motor failures and potential hazards, early detection of fault becomes a critical and essential need of the industry.

Decision-making support is useful if it is implemented to identify the motor fault at the incipient stage so as to avoid the unwanted shutdown of the processing time / or to bring the regular operation of the industry in a limited capacity of time. The reliable operation of the induction motors is very much dependent on condition monitoring.

Fault diagnosis of Induction mchines includes identification of the fault occurrence, its severity, and predicting fault at an early stage. Fault prediction at the incipient stage avoids costlier damages and downtime.

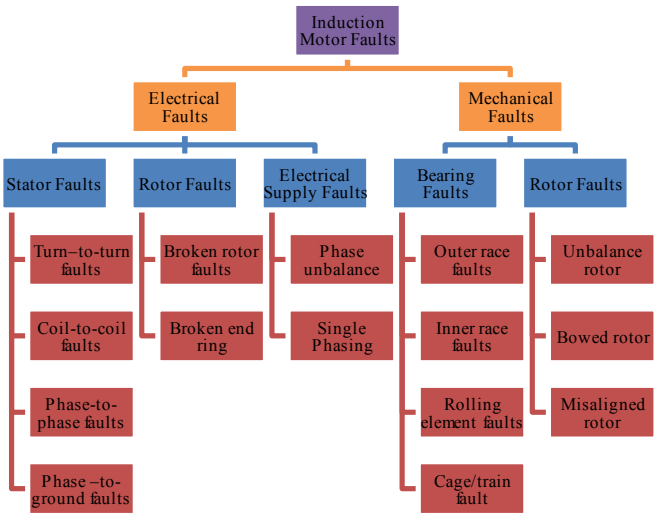


Figure 1.1: 3-Φ Induction motors faults

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 05,2022 at 05:51:17 UTC from IEEE Xplore. Restrictions apply.

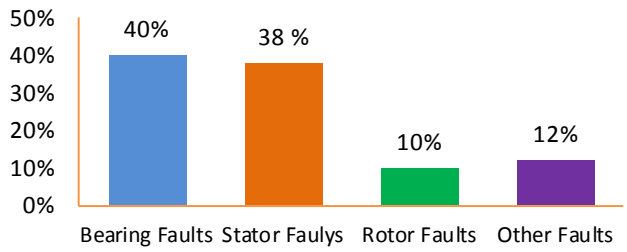


Figure 1.2: Important faults of 3-Φ Induction motors and their percentage

2. Stator Fault

Nearly 38 % of induction motor faults are on stator side [5]. The possible faults that can occur on the stator side are single phasing stator fault (open circuit fault), inter-turn short circuit fault, the line to line fault, and line to ground fault. Among all these faults, inter-turn short circuit fault is most challenging fault as other faults are the consequences of inter-turn short circuit fault. Also, 27% of faults occurring in induction motors are short circuit faults that occur due to insulation failures [12]. The stator side winding failures are initiated as insulation failure and spread as the inter-turn short circuit fault [11]. It takes only very few seconds for a short circuit fault to extend into a severe fault which may cause costlier damage to machine. Considered stator fault for analysis in this paper is inter-turn incipient short circuit fault as it occurs very often and creates quick damage to the machine.

Any abnormalities in the motor reflect in the current pattern. Also, with no doubt, the fault created in the stator has an impact on the stator current. Therefore the fault harmonic frequencies will appear in stator spectrum and are the representatives of fault. The frequencies appearing in the spectrum due to stator fault (f_{st}) can be estimated using equation 1.

$$f_{st} = f * [(\frac{m}{p}(1-s)) \pm k \quad (1)$$

Where:

f_{st}= frequency related to stator fault
f = fundamental or primary frequency,
s = slip,
p = number of poles pairs,
k = 0, 1, 3, 5…… &
m = 1, 2, 3….

Using the above equation, the expected lower order fault harmonics in the spectrum that represents stator fault are 25, 75, 100, 125, 175, 200, 225, 275,300, 325Hz...

3 Condition Monitoring's Need

Condition monitoring is the course of action of continuous evaluation of a health parameters state in machinery (vibration, speed, temperature etc.), in order to discover a considerable changes which is pinpointing of a budding faults. It is a key component of prognostic maintenance in all machineries. This is called as a just beginning failure finding (Peter Vas 1993). The use of condition monitoring permits the routine preservation to be planned or further dealings to be appropriated to prevent failure and avoid its penalty. Condition monitoring has a distinctive profit in that conditions that would shorten usual lifespan can be addressed prior to enlarge into a foremost breakdown. In accretion, it is also realistic to program upcoming defensive preservation and overhaul occupations. This can effect in slightest down time and best possible preservation agenda (Tavner & Penman 1987). A status monitoring and liability finding system allows the machinist to have indispensable spare parts before motor is stripped down, in this manner dropping blackout times. Therefore, effectual condition monitoring of industrial machines is serious in humanizing the consistency, safety, and production.

4 Current Condition Monitoring Techniques

1. Temperature Monitoring: The temperature monitoring of industrial machines is consummate either by estimating the thin or massive temperatures of motors, or by constraint assessment.

2. Torque Monitoring: All kind of machine faults creates sidebands at singular frequencies in air-gap torque.

3. Sound Monitoring: Sound monitoring is done by investigating and measuring audio noise band. Sound noise from air gap strangeness in induction machines can be utilized for fault illuminating.

4. Vibration Monitoring: Noise and vibrations are created by all-electric machines, and examination of twisted vibration and noise can be utilized to give information on state of machine.

5. Motor Current Signature Analysis

The Motor Current Signature Analysis consumes the line current spectrum of the given machine for tracing characteristic liability frequencies. When an error is present in the machine stator windings in any form, after that the line current element frequency spectrum becomes curious from energetic motor. Such a fault attuned air-gap of the machine and creates rotating frequency harmonics in self and mutual inductances of device. It depends upon finding explicit harmonic factor in machine line current. For that reason, it recommends momentous implementation and profitable benefits to diagnose the faults in the machine. The Motor Current Signature Analysis methodology offers to diagnose the commonly classified Induction Motor faults such as stator faults, rotor faults, air-gap eccentricity fault, bearing fault and fault due to load. In this study, MCSA based process is used to spot out the regular stator electrical faults of induction motor such as open circuit fault, short winding fault (line to line, coil to coil and turn to turn), and leakage fault.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 05,2022 at 05:51:17 UTC from IEEE Xplore. Restrictions apply.

6. Experimental & Results

In this proposed stipulation monitoring of Induction Machine by means of superior signal processing practices under simulation environment for determining stator winding electrical faults. Stator line current restrains distinctive slipup frequency parts that can be worn for Stator Inter-Turn fault revealing. These processes are used for unremitting supervision of the machine fitness.

An Induction Motor breakdown due to stator Inter-Turn fault may cause the blackout of a generating division or production. One most important causes of the malfunction are the breakdown of the insulation of the winding. The premature finding of stator short circuit faults in the winding during machine action may eradicate subsequent injure to neighboring stator coils. It minimizes the revamp expenditure and outage time of the machine. In addition to the profit earned from before time revealing of stator winding insulation failure, considerable compensation may accumulate by tracing the faulted portion contained by the windings.

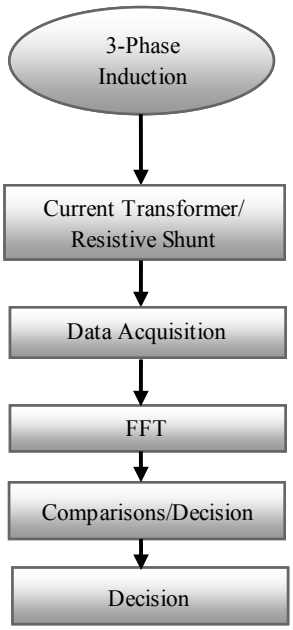


Table: 6.1 Parameter of Experimental Machine

S.No	Parameter	Data
1.	Power	0.5 Hp
2.	Frequency	50 Hz
3.	No. of Phases	03
4.	Speed	1490 rpm
5.	Volts	415 V
6.	Number of stator slots	24
7.	Number of pole pair	02

Table: 6.2 Slip & Sideband frequency

Condition	Speed	Slip	k=1	
			LSB	MSB
No-Load	1490 rpm	0.06	26 Hz	75.5 Hz
Full-Load	1410 rpm	0.06	23 Hz	76.5 Hz

Table: 6.3 Different Inter-Turn Short Circuit Condition

Experiments	Faults (%)	Load Condition
1.	0 %	No Load
2.	2 %	
3.	5 %	
4.	7 %	
5.	0 %	Full Load
6.	2 %	
7.	5 %	
8.	7 %	

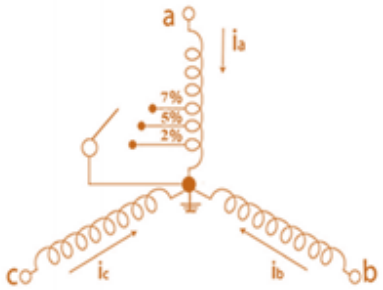


Figure 6.0: Stator winding Fault in an Induction Machine in phase '*a*'

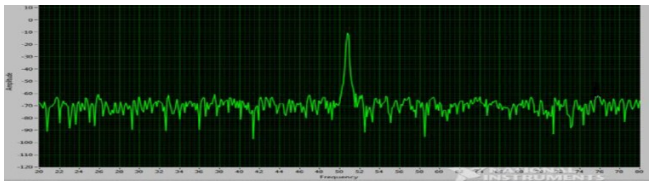


Figure 6.1: FFT Analysis spectrum of healthy motor under no load condition

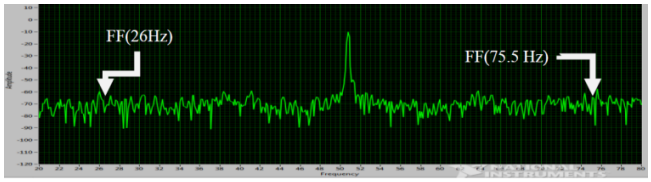


Figure 6.2: FFT Analysis spectrum of Motor with 2% Inter-turn Short-Circuit Fault under no load condition

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 05,2022 at 05:51:17 UTC from IEEE Xplore. Restrictions apply.

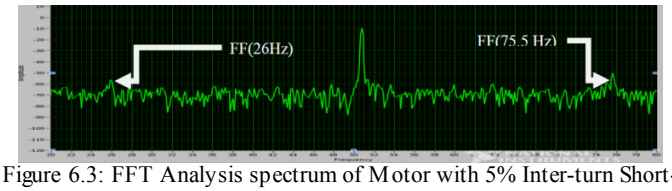
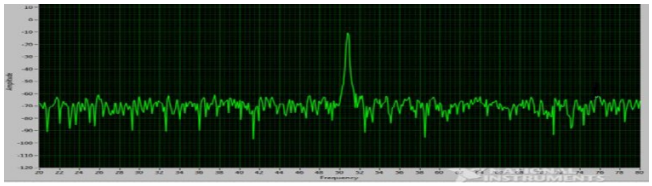


Figure 6.3: FFT Analysis spectrum of Motor with 5% Inter-turn Short-Circuit Fault under no load condition



Figure 6.4: FFT Analysis spectrum of Motor with 7% Inter-turn Short-Circuit Fault under no load condition


Figure 6.5: FFT Analysis spectrum of healthy motor under full load condition

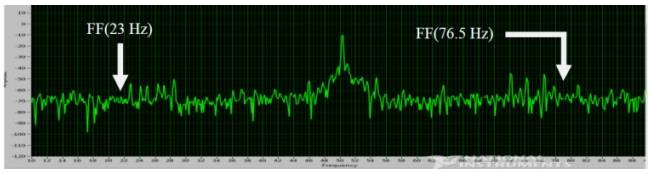


Figure 6.6: FFT Analysis spectrum of Motor with 2% Inter-turn Short-Circuit Fault under full load condition

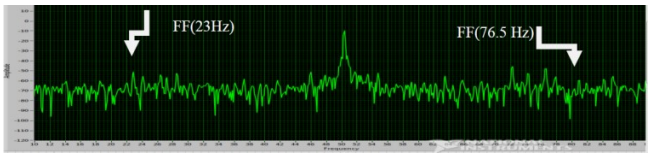


Figure 6.7: FFT Analysis spectrum of Motor with 5% Inter-turn Short-Circuit Fault under full load condition

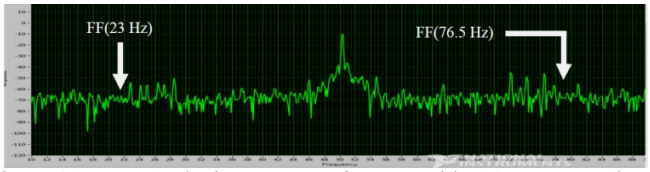


Figure 6.8: FFT Analysis spectrum of Motor with 7% Inter-turn Short-Circuit Fault under full load condition

Table: 6.4 Results

Figure Number	Stator Inter-turn Short-Circuit	Load Condition	Fault Frequency				Observation
			LSB (Lower Side band)		USB (Upper Side Band)		
			FF (Hz)	Magnitude (dB)	FF (Hz)	Magnitude (dB)	
6.1	0 %	No Load	26	-60	75.5	-60	Visible
6.2	2 %		26	-60	75.5	-58	Visible
6.3	5 %		26	-56	75.5	-50	Visible
6.4	7 %		26	-54.5	75.5	-48	Visible
6.5	0 %	Full Load	23	-67	76.5	-66	Visible
6.6	2 %		23	-54	76.5	-46	Visible
6.7	5 %		23	-50	76.5	-48	Visible
6.8	7 %		23	-48	76.5	-47	Visible

7. Observation and Discussions

Faults experimental analysis was carried out on a proto type induction motor with a rating of 0.5 Hp KW, 3 Phase, 415 V, 2 Pole 50 Hz and the stator electrical parameters were recorded in a power analyzer with a time interval of 200 milli seconds. The investigation result yielded a clear picture about the performance of the motor under various faults. The variation of current in fault phase and the presence of harmonic content in the stator current were easily predicted. Based on the results the nature of fault was easily determined.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 05,2022 at 05:51:17 UTC from IEEE Xplore. Restrictions apply.

8. REFERENCES

1. J. Lu, P. Wang, S. Duan, L. Shi, and L. Han, "Detection of broken rotor bars fault in induction motors by using an improved MUSIC and least-squares amplitude estimation," Mathematical Problems in Engineering, vol. 2018, Article ID 5942890, 12 pages, 2018.
2. Y. H. Liu, J. W. Qiao, D. Y. Huo et al., "Development of mineused flameproof three-phase asynchronous motor with high voltage and high efficiency," Industry and Mine Automation, vol. 43, no. 7,pp. 32–36, 2017.
3. X. Sun, M. Wu, G. Lei, Y. Guo, and J. Zhu, "An improved model predictive current control for PMSM drives based on current track circle," IEEE Transactions on Industrial Electronics, p. 1, 2020.
4. X. Sun, L. Chen, Z. Yang, and H. Zhu, "Speed-sensorless vector control of a bearingless induction motor with artificial neural network inverse speed observer," IEEE/ASME Transactions on Mechatronics, vol. 18, no. 4, pp. 1357–1366, 2013.
5. X. Sun, C. Hu, G. Lei, Y. Guo, and J. Zhu, "State feedback control for a PM hub motor based on gray wolf optimization algorithm," IEEE Transactions on Power Electronics, vol. 35, no. 1, pp. 1136–1146, 2020.
6. G. D. Cheng, H. P. Wang, and J. J. Xia, "Study on rotor-side frequency conversion speed regulation technology for high voltage wound-rotor motor," Industry and Mine Automation, vol. 44, no. 8, pp. 81–86, 2018.
7. Y. Xie, X. T. Shan, J. P. Guo et al., ".e relationship study between field changes and faulty condition in squirrel-cage induction motor with broken bars fault," in Proceedings of the CSEE, vol. 37, no. 14, pp. 4222–4231, 2017.
8. C. J. Aileen, S. Nagarajan, and S. R. Reddy, "Detection of broken bars in three phase squirrel cage induction motor using finite element method," in Proceedings of the International Conference on Emerging Trends in Electrical and Computer Technology, pp. 249–254, Nagercoil, India, March 2011.
9. V. Climente-Alarcon, A. Arkkio, and J. Antonino-Daviu, "2-D magnetomechanical transient study of a motor suffering a bar breakage," IEEE Transactions on Industry Applications, vol. 54, no. 3, pp. 2097–2104, 2018.
10. P. A. Panagiotou, I. Arvanitakis, N. Lophitis, J. A. Antonino- Daviu, and K. N. Gyftakis, "FEM approach for diagnosis of induction machines' non-adjacent broken rotor bars by shorttime Fourier transform spectrogram," >e Journal of Engineering, vol. 2019, no. 17, pp. 4566–4570, 2019.
11. Y. G. Lv, W. Y. Zhang, and J. C. Liu, "Analysis and diagnosis of broken rotor bars in induction motors based on Maxwell," in Proceedings of the 33rd Chinese Control Conference, pp. 3065–3069, Nanjing, China, July 2014.
12. N. Lahouasnia, M. F. Rachedi, and T. Deghboudj, "Detection of broken rotor bar defect in squirrel cage induction machine," in Proceedings of the 2019 International Conference on Advanced Electrical Engineering, pp. 1–5, Algiers, Algeria, November 2019.
13. Amanuel, Thomas, Amanuel Ghirmay, Huruy Ghebremeskel, Robel Ghebrehiwet, and Weldekidan Bahlibi "Comparative Analysis of Signal Processing Techniques for Fault Detection in Three Phase Induction Motor" Journal of Electronics 3, no. 01 (2021): 61- 76.
14. Amanuel, Thomas, Amanuel Ghirmay, Huruy Ghebremeskel, Robel Ghebrehiwet, and Weldekidan Bahlibi. "Design of Vibration Frequency Method with Fine-Tuned Factor for Fault Detection of Three Phase Induction Motor." Journal of Innovative Image Processing (JIIP) 3, no. 01 (2021): 52-65.
15. Vinothkanna, R. "Design and analysis of motor control system for wireless automation." Journal of Electronics 2, no. 03 (2020): 162-167.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on March 05,2022 at 05:51:17 UTC from IEEE Xplore. Restrictions apply.

mathematics

MDPI

Article

Techno-Economic and Environmental Analysis of Grid-Connected Electric Vehicle Charging Station Using AI-Based Algorithm

Mohd Bilal [1,*], Ibrahim Alsaidan [2], Muhannad Alaraj [2], Fahad M. Almasoudi [3] and Mohammad Rizwan [1]

[1] Department of Electrical Engineering, Delhi Technological University, Delhi 110042, India; rizwan@dce.ac.in
[2] Department of Electrical Engineering, College of Engineering, Qassim University, Buraydah 52571, Saudi Arabia; alsaidan@qu.edu.sa (I.A.); muhannad@qu.edu.sa (M.A.)
[3] Department of Electrical Engineering, Faculty of Engineering, University of Tabuk, Tabuk 47913, Saudi Arabia; falmasoudi@ut.edu.sa
* Correspondence: bilal_2k17phdee05@dtu.ac.in

Abstract: The rapid growth of electric vehicles in India necessitates more power to energize such vehicles. Furthermore, the transport industry emits greenhouse gases, particularly SO_2, CO_2. The national grid has to supply an enormous amount of power on a daily basis due to the surplus power required to charge these electric vehicles. This paper presents the various hybrid energy system configurations to meet the power requirements of the electric vehicle charging station (EVCS) situated in the northwest region of Delhi, India. The three configurations are: (a) solar photovoltaic/diesel generator/battery-based EVCS, (b) solar photovoltaic/battery-based EVCS, and (c) grid-and-solar photovoltaic-based EVCS. The meta-heuristic techniques are implemented to analyze the technological, financial, and environmental feasibility of the three possible configurations. The optimization algorithm intends to reduce the total net present cost and levelized cost of energy while keeping the value of lack of power supply probability within limits. To confirm the solution quality obtained using modified salp swarm algorithm (MSSA), the popularly used HOMER software, salp swarm algorithm (SSA), and the gray wolf optimization are applied to the same problem, and their outcomes are equated to those attained by the MSSA. MSSA exhibits superior accuracy and robustness based on simulation outcomes. The MSSA performs much better in terms of computation time followed by the SSA and gray wolf optimization. MSSA results in reduced levelized cost of energy values in all three configurations, i.e., USD 0.482/kWh, USD 0.684/kWh, and USD 0.119/kWh in configurations 1, 2, and 3, respectively. Our findings will be useful for researchers in determining the best method for the sizing of energy system components.

Keywords: artificial intelligence; salp swarm algorithm; hybrid optimization of multiple energy resources; renewable energy; electric vehicle charging station

check for updates

Citation: Bilal, M.; Alsaidan, I.; Alaraj, M.; Almasoudi, F.M.; Rizwan, M. Techno-Economic and Environmental Analysis of Grid-Connected Electric Vehicle Charging Station Using AI-Based Algorithm. *Mathematics* **2022**, *10*, 924. https://doi.org/10.3390/math10060924

Academic Editor: Alessandro Niccolai

Received: 2 February 2022
Accepted: 10 March 2022
Published: 14 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Energy is a critical component of long-term development and poverty eradication. In India, electricity is accessible to 97.2 percent of the population. Natural gas and coal are the most common fuels used in the electricity sector. India has a total power generation capacity of 388.134 GW (including off-grid renewable energy), with renewable energy (RE) accounting for 21.26 percent [1]. The scarcity of natural gas and coal reserves could pose a serious threat to electricity production. To meet the country's rising electricity consumption, the Indian government has announced plans to install 275 GW of sustainable power by 2027. The government has a strategy to satisfy demand with 356.681 GW (Ministry of power, Delhi, India) of electricity generation from coal and nuclear power plants, yet these methods are not eco-friendly and are eroding our environment. RE-based power generation, which emits fewer greenhouse gases (GHGs) than other traditional power generation systems, is required in this framework. India, as a developing country, must provide quality and

reliable energy for all of its citizens to achieve its long-term development goals. As per the information revealed by World Resources Institute, India is responsible for emitting 7.1 percent of worldwide emissions and has total emissions of 2.47 tCO_2e compared to a global average of 6.45 tCO_2e. According to World Bank data, per capita consumption of electricity in India has grown from 920 kWh in 2012–2013 to 1210 kWh in 2019–2020.

The transportation sector of India contributes significantly to GHGs emissions. CO_2 accounts for the majority of these GHGs emissions. Furthermore, the energy and agriculture industries are hastening CO_2 emissions. The rapid growth in the number of vehicles needed to serve the country's massive population is a worrisome symptom of pollution and fuel consumption. Aside from these circumstances, the use of electric vehicles (EVs) such as e-rickshaws is increasing regularly. These EVs emit fewer pollutants and no exhausts. The Ministry of Road Transport and Highways does not have any data on these EVs. Some of these EVs are certified, and data reveal that there are 517,322 registered EVs out of a total of 295,800,000 automobiles. However, the rapid expansion of these EVs necessitates a massive quantity of electricity every day from India's national grid. EVs are now charged in residential neighborhoods, with residential customers paying the electric bills. In such circumstances, charging those EVs does not generate any revenue for the electricity sector. Meanwhile, these EVs are creating a lot of stress on the grid network of India. A profitable strategy to generate electrical energy is needed to relieve the strain on the national system. However, to the best of the authors' knowledge, there are no endeavors in India to establish a hybrid-energy-based electric vehicle charging station (EVCS), which prompted us to conduct this study described here.

India has various sources of RE, which include solar photovoltaic (SPV), wind turbines (WT), biogas generators, and biomass. As a result, power might be generated using existing resources. The use of renewables can reduce harmful emissions while also lowering operating costs and maintaining cleanliness. In these systems, paying a bill every month is not an option. EV charging with RE resources enhances commercial profitability, reduces air pollution, and decreases noise in metropolitan regions.

Due to the enormous solar energy availability, it is the finest alternative for generating electrical energy to serve EVs. The solar-powered EVCS at Karna Lake Resort in Karnal is one of Bharat Heavy Electricals Limited's (BHEL) 20 such stations along the Delhi-Chandigarh expressway. It has chargers for all models of electric cars already on the road in the country, including AC001 (33.3 kW), DC001 (15 kW), and 72 kW (50 kW DC + 22 kW AC).

Solar irradiation is missing during rainy days and in foggy surroundings, and with fluctuations in wind energy, no energy is generated at that time. As a result, relying solely on solar and wind energy may compromise system stability and cause a project to fail. Hence, diesel generators (DGs), which can be a decent substitute for generating power even though solar and wind power are unavailable, are required.

1.1. Present Scenario of EVs and EVCS

In the early phases, companies put forward hybrid EV technology as the most reliable and eco-friendly technology. Even though this technology has made major developments, it still mixes traditional fuel practice with an electric engine, which is used when applications require minimal power capacity. Later, technological innovations and more effective battery storage led to the establishment of improved EV systems that completely rely on electrical energy for supply. These devices started to run using electrical energy provided by RE sources, such as SPV and WT. Some research was carried out for enhancing energy efficiency. For example, in Ref. [2], a complete investigation for the forecasting of effective charging profile for EVs by using multiple run-out of charging conditions is provided. Artificial neural networks to build algorithms that perform the optimized energy management of EVs are used [3]. The authors demonstrated how to use a smart nano grid system and plug-in electric automobiles to manage the optimized integration of hybrid SPV systems [4].

According to annual surveys, the number of EVs and charging stations has increased exponentially over the last decade. Battery-based EVs were first started in India in 2001.

Now, these are running in nearly every part of the region. There are three varieties of EVs comprising electric rickshaws, which carry between four and five passengers, auto-rickshaw for three passengers, and electric vans for delivering products. A fully charged electric rickshaw can go about 80–110 km and costs roughly one lakh seventy-five thousand rupees in the market, whereas auto-rickshaws and electric vans have a range of 50–80 km and cost around one lakh rupees. These were formerly imported from China, but today they are made in the country itself. An EV that is completely recharged can travel 80–110 km a day while consuming 8–12 kWh. The cost of power per unit is Rs. 6.90 on the commercial tariff, hence the cost incurred per km run is around Rs. 75.9. One of the e-rickshaw charging stations installed in Delhi is shown in Figure 1.

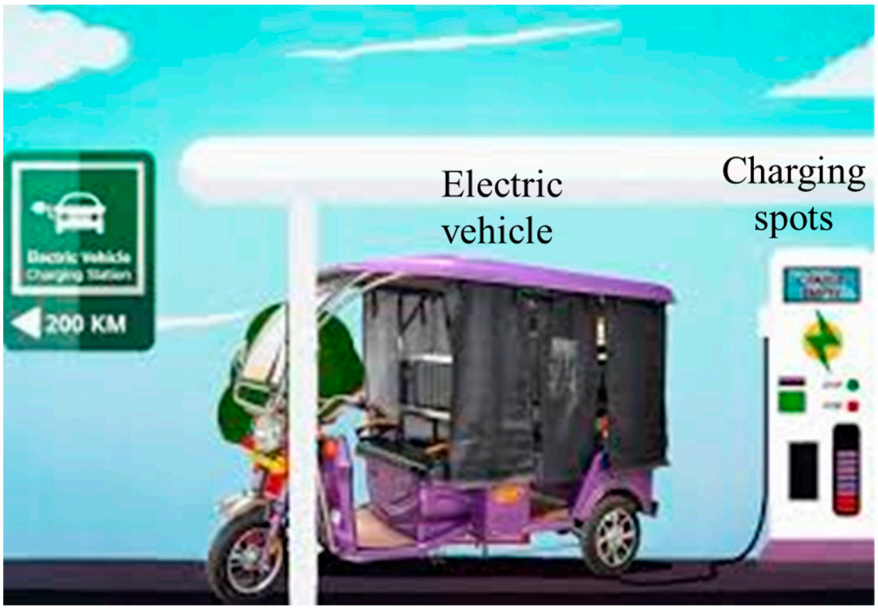


Figure 1. Electric vehicle charging station in Delhi.

The features of e-rickshaw in India are shown in Table 1. This mode of transportation is beneficial to the environment, emitting nearly no fumes and creating significantly fewer CO_2 emissions. It has brought a cultural shift to the transportation industry, becoming well known for its lower fares when related to other means of road transit, thus enhancing the satisfaction of low-income individuals, including vehicle drivers [5]. Some private ventures have launched electric-based Easy Bikes and Auto-rickshaw, which have gained traction in village areas and near metropolitan zones. Such EVs consume a significant amount of grid energy; however, there are a lot of measures underway to relieve the strain on the system. However, because power rickshaws are said to utilize a significant quantity of electricity from grid networks, the changes have yet to be widely implemented.

Table 1. Specification of EVs (e-rickshaw) in Delhi. Data from Ref. [6].

Particulars	E-Rickshaw
Dimension	2790 × 975 × 1730 mm
Carrying capacity	Four passengers + driver
Weight	210 kg (without battery)
Battery	4 Nos./48 V
Battery capacity	80 Ampere/hour
Approx. mileage	70 km per full charge
Charging time	7–8 h per day
Maximum speed	25 km/h
Power consumption	8–11 kWh

The authors performed a survey in Delhi's northwestern city for this study. Every day, more than 200 different varieties of electric rickshaws run here. There were just 20 manually operated rickshaws recorded. The majority of electric rickshaws are powered using electricity installed at home, which is not permitted by the Indian government. Additionally, power outages occur very often in the neighborhood. Therefore, charging EVs during that time is challenging.

It would be easier for those individuals to drive their battery-powered EVs under these conditions if services were created to charge EV batteries using other trustworthy energy sources in the surrounding part. EVs have had a seismic impact on road transport, particularly in rural and sub-urban regions, despite significant charging issues.

1.2. Literature Review

EVs adoption rate is rapidly increasing all over the world, promoting this energy usage industry as a potential area of research. This subsection focuses on the existing EV charging infrastructure, hybrid renewable power production technologies for EV charging, costs of electricity generation, and the reduction in GHG emissions by the charging system. The power and transport industries account for the majority of carbon footprints on the planet. RE can help to reduce harmful emissions in the electrical sector, while grid-powered EVs can greatly decrease pollution in the transportation sector. When RE sources and EVs are combined, they have a lot of potential to solve ecological and financial issues. EVs and RE sources have been discussed in terms of the operation of the system in several kinds of literature.

Optimal systems of various hybrid combinations, such as SPV/wind/diesel/battery systems, have been documented in several articles. On the other hand, SPV systems are less expensive and more applicable than WT [7]. Since the growth and deployment of hybrid RE systems have increased dramatically in the past few years [8], the proper design of these systems is critical for achieving a high degree of reliability at low prices. Under-sizing the system can result in operational limits and energy shortages, while over-sizing the system can lead to high expenses. Hence, sophisticated optimization methods are essential to tackle the optimal sizing problem. The components of the hybrid power system must be designed and integrated as part of the sizing technique to anticipate their efficiency based on meteorological conditions; this comprises predicting the energy output of renewable sources and the battery charge state.

A range of sizing strategies has been reported in the literature, which includes recursive approaches [9] and software applications, such as HOMER [10]. The use of traditional procedures based on recursive, quantitative, or statistical methods has fallen dramatically [11], whilst approaches utilizing optimization algorithms have attracted a lot of attention. Heuristic algorithms are quite efficient and have a lot of potential when it comes to handling optimization challenges. Several types of research have been performed to apply such algorithms to the challenge of sizing RE plants [12]. The practical limits of the EV market in China are investigated, and the influence of incentives on an auto manufacturer's optimal production and pricing decisions is analyzed [13]. A feasibility assessment for a solar-energy-based EVCS in Shenzhen, China, is evaluated using the HOMER software tool. This anticipated strategy addresses the grid power-related issues by incorporating SPV, as well as achieving the expected demand for EVs [14]. There is no other option for generating electricity on wet or foggy days. Another feasibility analysis based on solar-powered EVCS was undertaken in Bulgaria [15]. Many studies envision the incorporation of RE sources, such as a PV system, into charging stations, as the best solution for maximizing the financial and environmental advantages of EVs while also promoting the idea of a smart grid [16]. In Ref. [17], authors have described multiple aspects of the problem formulation, including the choice of design variables and the definition of a probability distribution function to improve efficacy and effectiveness. In Ref. [18], the authors investigated an operational strategy for EVs that employs sustainable RE sources, such as SPV, and gives an optimization method for charging park energy management.

The radial distribution network is affected by the uncertainty of the charging load of the EVCS and SPV unit. A study used a hybrid of genetic algorithm (GA)- and particle swarm optimization (PSO)-based optimum scheduling of EV load to minimize system losses, variations in voltage, cost of charging, and battery costs of EV [19]. The elastic demand was developed using either feedback from crowded travel and congested stations on route selections or the presumption that charging demand between origin and destination pairs follows a nonlinear inverse cost function without taking into account pre-generating pathways and charging configurations [20]. Authors envisioned a PSO for the optimal sizing of SPV and battery energy storage system (BESS) in a grid-connected EVCS, with a financial model as the goal [21]. In Ref. [22], authors suggested a hybrid solar and wind charging station that was deployed in Tangshan, China, and proposed an LPSP technique to evaluate the component capabilities. In Ref. [23], a cost-effective EV charging station is developed by comparing a standalone EV charging station to a grid-connected charging station while taking emissions into account. However, there was no cost evaluation for the hybrid system. The authors utilized linear programing to determine the optimum integration of CS, restricting the search area under the assumption that CSs can only be positioned in parking areas and service stations. The optimization technique considers the accessibility of CS within a reasonable driving range [24]. In Ref. [25], a charging station based on solar- and grid-connected stations for a worksite in the Netherlands is presented. Various charging possibilities for EVs were proposed. However, the work did not contain an economic analysis. The demand response systems and EV charging is investigated by integrating them into radial distribution network buses [26]. Particle swarm optimization is being used to resolve the issue of charging station placement [27]. A comprehensive design of an EVCS is proposed, in which EVCS are integrated with renewable sources of energy and a BESS to ease the burden on the utility grid. To optimize the design, a new technique is proposed, which is a hybrid of the crow search algorithm and PSO [28]. Another study was conducted at the University of Palermo in Italy to build a long-term recharging facility for EVs. It results in a low levelized cost of energy, implying that energy storage systems will require a significant amount of initial capital [29]. The GA is used to optimize the adoption of CS. The authors of this study specify some settlements in the urban environment. Settlements are intended to be the locations where EVs are parked for the majority of the time. GA defines CS positions by minimizing distances to settlements [30]. Based on local resource allocation, the author combined SPV and biogas-based resources with EV load in Bangladesh [5]. SPV energy is widely available because PV units can be easily installed on the rooftop. In Ref. [31], the author provides an optimization strategy for grid-connected SPV/BESS/EVCS to scale SPV, BESS, and establish the charging and discharging characteristics of BESS. This problem is implemented using the multi-agent PSO technique, which integrates the multi-agent system (MAS) and the PSO mechanism. The authors of Ref. [32] consider charging station implementation, taking into account the repercussions on travelers, cab drivers, electricity retail outlets, transport and distribution networks, and electricity consumers. A multi-objective optimization was suggested to handle the developed framework. A hybrid SPV and wind-based EVCS are designed and simulated in HOMER software [33]. An eco-friendly alternative for EV charging based on two RE sources, namely SPV and biogas, is presented. The HOMER package is used to assess the effectiveness and features of SPV and biogas-based EVCS. An SPV system, two biogas generators, and a bi-directional converter with BESS are all part of the proposed system. HOMER software analyzes the alteration of various costs, such as net present value, fixed cost, and energy cost for various SPV configurations [6]. The authors of Ref. [34] presented an enhanced version of chicken swarm optimization (CSO) to optimally locate the EVCS in an IEEE 33 node test distribution network. They first evaluated the network impact of projected EV charging load at the EVCSs in terms of voltage levels, average voltage deviation index, voltage stability index, and power loss, and then, using a feed-forward neural network, they investigated the solar power required to energize the EVCSs. In Ref. [35], the author presented a grid-integrated solar- and wind-based energy system to

meet the power requirements of a shopping mall, as well as EVCS load. The components are sized to attain the lowest electricity cost while lessening the deficiency of power supply probability, using artificial colony bee and PSO. The optimum strategy and operation of a renewable-powered EVCS will be presented in this study, to lower cost and emissions. The proposed hybrid system comprises SPV, WT, biogas, batteries, and DG units [36]. In Ref. [37], the authors evaluate the technical viability of implementing a standalone fast charging station in the State of Qatar, which includes WT, SPV system, and a bio-generator as RE source, as well as multiple storage devices. The envisaged design is constructed, modeled, and simulated using the HOMER software to determine the best techno-economic configuration for fast charging 50 EVs on a daily basis. Space constraints, the deterministic nature of EV demand, and metrological conditions of the considered site are all taken into account. The goal of the research undertaken is to use HOMER pro software to deliver SPV power to EVs and feed excess power to the grid network. The author used GA to develop a modeling framework for an SPV/wind/battery for a deserted landmass [38]. A multi-objective PSO to size and analyze a PV/wind/hydroelectric power station with a pumped-storage connected energy system is suggested [39]. The authors investigated a probabilistic modeling approach of an EVCS with SPV, battery, and transformer, with an emphasis on charging demand and SPV power generation complexities. To begin, a detailed EV charging requirement model is developed by projecting the coupling between the EV dynamic charging pattern, charger specifications, and EV charging assignment model. The planning model is then developed to minimize the total cost of the charging station, with the uncertainties associated with EV charging being acknowledged through various limits [40]. A multi-objective moth flame optimization (MFO) that was used to optimize a hybrid microgrid system with SPV/wind/BESS/DG is designed [41]. In Ref. [42], the author focuses on a hybrid energy system that includes an SPV, WT, a biogas generator, and BESS for providing reliable electricity to a deserted island in Bangladesh called Saint Martin. The energy system's component is sized in accordance with energy cost and life cycle emissions under a certain level of reliability. The objectives are optimized by considering two well-known multi-objective optimization methods: non-dominated sorting GA-II and infeasibility-driven evolutionary algorithm. In Ref. [43], the authors presented the technical economic viability of solar- and wind-based EVCS at five different areas in the southern region of Tamil Nadu. The locations are selected such that wind and solar possibilities are of contrary nature in the required places. In Ref. [44], a comprehensive quantitative model is designed to identify the suitable SPV/WT/BESS sizes for distant locations. The suggested method is tested at Ras-Shaitan, a rural area in Egypt-Sinai. The goal of the optimization method is to accommodate power requirements by lowering the cost of energy under various power supply failure scenarios. To get the optimum objective value, the gray wolf optimizer is used to determine the SPV, WT, and BESS units.

Several heuristic methodologies for tackling the sizing issue of various system components have been suggested in the research publications. Because of their excellent ability to tackle difficult optimization issues, researchers are increasingly looking into these strategies. However, the efficiency of each optimization technique, when used for evaluating the optimal size, may vary in terms of precision and convergence rate. Although there is a range of sizing methodologies available, academics working on this problem are constantly looking for a robust and highly precise approach with a small implementation time. This is in order to develop and formulate energy systems that are both reliable and cost effective.

The control and power management of EVs in grid-connected systems are the primary focus of researchers. However, one of the important aspects that must be addressed is an economic analysis that takes into account the power exchange with the grid. The fast adoption of EVs poses both constraints and opportunities for the current electricity system. A small grid-connected SPV and DG-based hybrid system with EVs are presented in this article for a charging station in the northwest region of Delhi, India. The main objective is to formulate a statistical model of a solar and diesel generator-based hybrid system with EVs and a backup grid. Furthermore, the purpose of this research is to reduce

power interchange with the grid. The salp swarm algorithm (SSA) and the modified salp swarm algorithm (MSSA) have never been introduced before for satisfying the EV load requirement using hybrid SPV and DG with utility grid as a backup. To our knowledge, no previous study has compared the performance of so many different sizing methodologies on a single energy system. Hence, the assessment described in this article will be a useful guide for researchers looking to select a technique for their sizing concern.

1.3. Contribution

A solar PV module, diesel generator, battery storage, and backup grid are all part of the hybrid energy system being investigated. The system is designed as a simple micro grid to test the algorithms' performance against a conventional framework. The system's total net present cost (TNPC) is minimized, while the lack of power supply probability (LPSP) is taken into account as a reliability restriction. In addition, the levelized cost of energy cost (LCOE) is used to evaluate the hybrid energy system. The optimization technique is based on real-time solar irradiation and ambient temperature data collected over one year.

In light of the preceding discussions, the following are the key objectives of this paper:

- Identify the optimal EVCS design using realistic inputs on physical, operating, and economic features of RE sources, such as SPV and DG, with the utility grid.
- To create a novel idea that combines a photovoltaic module with DG to provide electricity for charging plug-in electric vehicles.
- To assess the technological, financial, and ecological viability of the proposed model.
- Two novel methods, SSA and MSSA, are discussed and employed in the sizing problem of the system component for the first time. The efficacy of such methods is assessed and equated in terms of consistency, convergence cycle, and processing time.

The rest of this article is laid out as follows. The modeling of the various components of a hybrid energy system is performed in Section 2. The problem formulation is elaborated in Section 3. The proposed optimization methodology is discussed in Section 4. The results are thoroughly discussed in Section 5. Future challenges and opportunities are presented in Section 6. Finally, conclusions are stated in Section 7.

2. Hybrid System Components and Their Mathematical Modeling

The mathematical modeling of the hybrid energy system to meet the charging demand of EVs is described in this section. An SPV panel, DG, battery, inverter, and backup grid are all part of the hybrid energy system under consideration. To enhance load supply reliability, the battery is utilized to control the variation of RE generation. A graphical representation of the proposed SPV/DG/battery storage system is shown in Figure 2. The methodology for conducting this research work is shown in Figure 3.

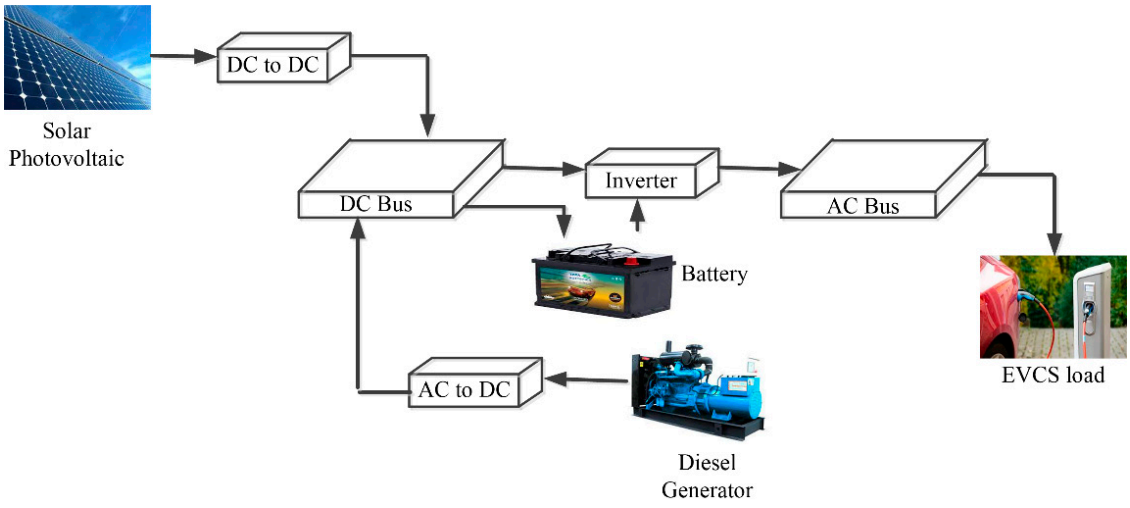


Figure 2. Schematic representation of the proposed hybrid energy system.

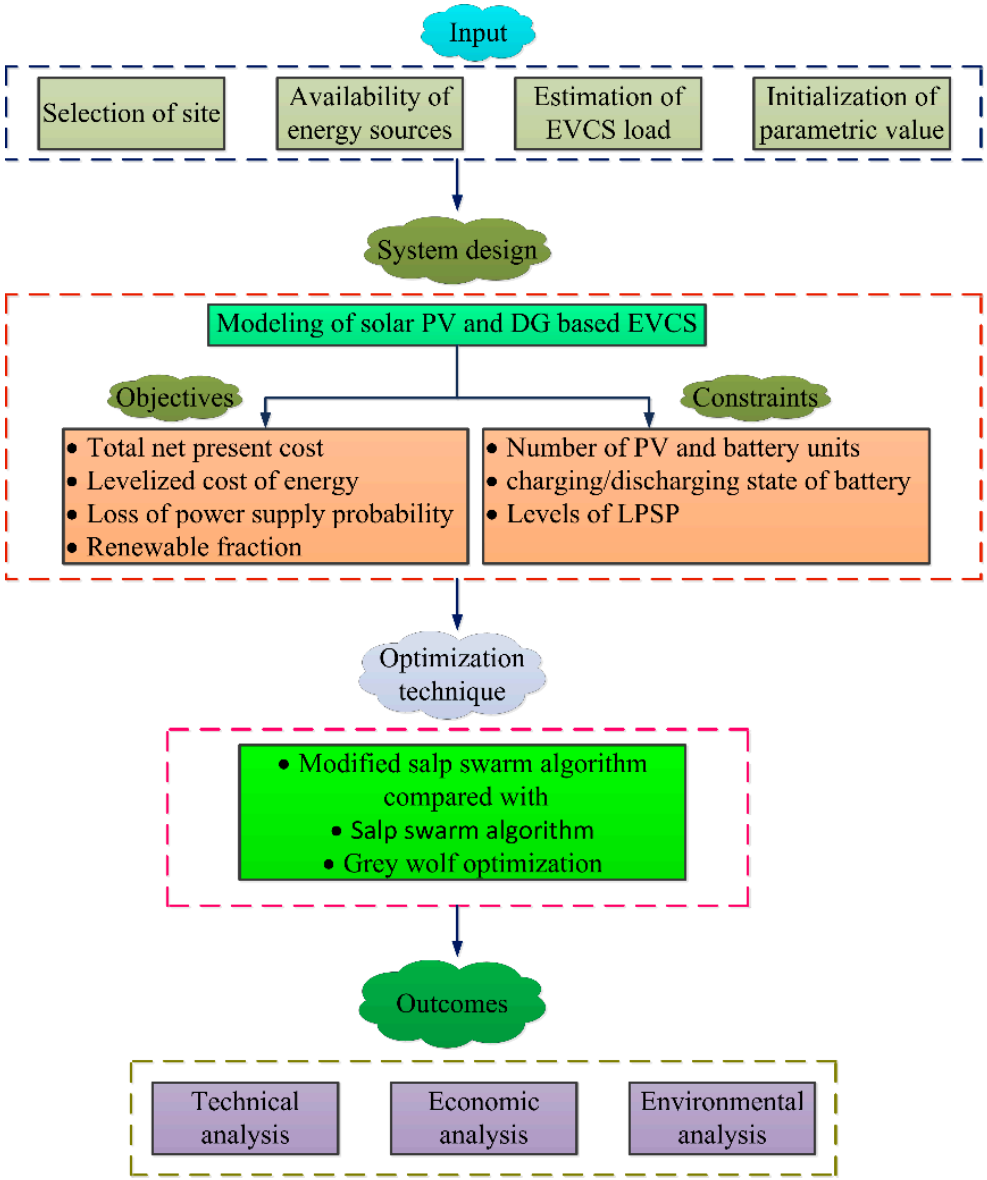


Figure 3. A schematic representation of the system analysis and design procedure.

2.1. SPV Panel Output

Solar radiation is the most prevalent component on the planet, as well as being the most robust RE source. It is comparatively simple to execute and can be made affordable in all remote areas of the country. Concerning the maximum output pattern of SPV, modules are widely used to estimate the quality of an SPV generation system to its peak power. SPV panels are used to generate electricity by collecting solar energy. Only a small portion of the solar radiation that penetrates the SPV panels is transformed to electricity, with the remainder being converted to heat. Several factors can influence the quality of the SPV power output. The power output obtained from SPV is dependent on the area of the panel, solar irradiance incident on the surface of the panel, temperature of the SPV cells, and geographical parameters, i.e., latitude and longitude of the panel location, using Equation (1):

$$P_{SPV\text{-}out}(t) = \eta_{SPV} A_{SPV} G_h(t)\{1 - 0.005(T_{cell} - 25)\} \quad (1)$$

where η_{SPV} is the efficiency of solar panels, A_{SPV} represents the surface area of the solar panel, $G_h(t)$ denotes the hourly solar irradiance falling on the surface of the solar panel,

and T_{cell} is the temperature of the cell. The parameters taken for the selected PV panel are tabulated in Table 2.

The temperature of the cell and the fluctuation of power as a function of temperature are given by

$$T_{cell} = T_a + \frac{NOCT - 20}{0.8} G_h(t) \quad (2)$$

where T_a is the ambient temperature.

The rated maximum power output of solar panel can be expressed as

$$P_{max} = V_{max} I_{max} \quad (3)$$

where V_{max} denotes the maximum value of voltage, and I_{max} represents the maximum current, respectively.

The overall power generated using SPV modules can be estimated by

$$P^{total}_{SPV\text{-}out}(t) = N_{SPV} P_{SPV\text{-}out}(t) \quad (4)$$

where N_{SPV} is the number of the installed solar PV panels.

2.2. Diesel Generator (DG)

The quantity of energy storage necessary for isolated villages and rural enterprises can be lowered by using DGs, resulting in a profitable and reliable solution. In the event of battery degradation during peak loads, diesel can be used as a backup source of energy.

The DG's efficiency and hourly fuel usage should be accounted for when planning a hybrid system and may be represented using the equation below [34]:

$$u(t) = a \times P_{DG}(t) + b \times P_{rated\text{-}DG} \quad (5)$$

where u(t) is the fuel consumption in liters per hour, $P_{DG}(t)$ is produced power in kW, $P_{rated\text{-}DG}$ is the rated power, a (0.246) and b (0.08415) are constant parameters in liter per kW, which characterize the coefficients of fuel consumption.

The efficiency of DG can be computed by

$$\eta_{overall} = \eta_{brake\ thermal} \times \eta_{DG} \quad (6)$$

where $\eta_{overall}$ is the overall efficiency of DG, whereas $\eta_{brake\ thermal}$ represents the brake thermal efficiency of DG.

2.3. Modeling of Battery Energy Storage System (BESS)

Due to the erratic nature of SPV output, optimum battery sizing is essential to meet the load demand. The existing state of charge (SOC) is the most important decision variable for monitoring the charging/discharging states of the battery. Overcharging of the battery takes place when the hybrid model generates too much power or when the load demand is too low. When the battery's SOC reaches its maximum value, i.e., SOC (max), the control system gets involved and stops the charging mechanism, whereas when it acquires its minimum value, i.e., SOC (min), the control system disables the load to avoid the battery from being drained [45].

There are instances when the state of the battery varies depending on the power output and load demand. The battery performs an important role in the system's energy management by adjusting power demands and power supply. The battery is said to be charging when the generated power is higher than the required energy, and the charging state of the battery at time t is given by [45]

$$SOC_{bat}(t+1) = SOC_{bat}(t)(1-\sigma) + \left[\left(P_{SPV\text{-}out}(t) + \frac{P_{DG}(t)}{\eta_{bdinv}}\right) - P_{EVCS\text{-}dem}(t)\right] \times \eta_{bat} \quad (7)$$

On the other hand, when RE sources are unable to create enough power to meet demand, the battery is employed as a backup to feed the load system and operate in discharging state. The discharging state of the battery at time t is given by

$$SOC_{bat}(t+1) = SOC_{bat}(t)(1-\sigma) - \frac{\left[P_{EVCS\text{-}dem}(t) - \left(P_{SPV\text{-}out}(t) + \frac{P_{DG}(t)}{\eta_{bdinv}}\right)\right]}{\eta_{bat}} \quad (8)$$

where $SOC_{bat}(t+1)$ and $SOC_{bat}(t)$ are the SOC of battery at an instant $(t+1)$ and (t), respectively, σ is the self-discharge rate of the battery, η_{bdinv} is the efficiency of bi-directional inverter used, $P_{SPV\text{-}out}(t)$ and $P_{DG}(t)$ are the power output of SPV and DG, respectively, and η_{bat} is the round trip efficiency of the battery. The round-trip efficiency of the battery can be expressed using Equation (9):

$$\eta_{bat} = \eta_{bat}^{charge} \times \eta_{bat}^{discharge} \quad (9)$$

where η_{bat}^{charge} represents the charging efficiency of the battery, whereas $\eta_{bat}^{discharge}$ denotes the discharging efficiency of the battery. The round-trip efficiency of the battery bank is reported to be 92.2 percent. Furthermore, charging/discharging efficiencies are assumed to be 85/100 percent, respectively. SOC_{max} round-trip efficiency of the battery is the maximum value of SOC and is equivalent to the total energy of the battery bank C_{total}(Ampere hour), as follows:

$$C_{total}(\text{Ampere hour}) = \frac{N_{bat}}{N_{bat}^{series}} C_{sbat}(\text{Ampere hour}) \quad (10)$$

where C_{sbat} is the energy of a single battery, N_{bat} represents the total number of batteries, and N_{bat}^{series} represents the number of batteries connected in series. The battery bank is only allowed to discharge to a certain minimal level, known as SOC_{min}. This restriction can be used as a system constraint depending on how the battery bank is utilized. Batteries are arranged in series to get the desired bus voltage. The number of batteries in a series can be computed using the formula

$$N_{bat}^{series} = \frac{V_{bus}}{V_{bat}} \quad (11)$$

where V_{bat} is the voltage level of a single battery.

The maximum charge/discharge power at any given time is another important consideration in battery modeling. It is determined by the maximum charging current and may be computed using the equation below:

$$P_{bat}^{max} = \frac{N_{bat} \times V_{bat} \times I_{max}}{1000} \quad (12)$$

where I_{max} is the maximum charging current drawn by the battery in amperes, and P_{bat}^{max} is the maximum input/output power of the battery.

2.4. Bi-Directional Inverter Modeling

The bi-directional inverter performs the function of conversion of DC to AC power and vice versa. Solar panels provide DC electricity to EVs at the charging station. Hence, the bi-directional inverter is required to transform the DC power of SPV. The output of the bi-directional inverter can be computed as follows:

$$P_{AC} = \eta_{bdinv} \times P_{DC} \quad (13)$$

The proposed system takes into account a bi-directional inverter efficiency of 97 percent.

2.5. Utility Grid

In the first phase, when the demand exceeds the generated power from the different energy sources and BESS capacity, the energy shortage can be fulfilled by borrowing the energy from the grid network and may be represented using Equation (14):

$$\begin{aligned} P_P^{grid}(t) = P_D^{EV}(t) \\ &- \left[P_{SPV\text{-}out}^{total}(t) + P_{DG}(t)\right. \\ &\left. + \left[\left(P_{SPV\text{-}out}^{total}(t) + \left(SOC_{bat}(t) - SOC_{bat}^{min}(t)\right)\right) \times \eta_{bdinv}\right]\right] \end{aligned} \tag{14}$$

where $P_P^{grid}(t)$ represents the energy to be borrowed from the grid in kWh, and $SOC_{bat}^{min}(t)$ denotes the minimum SOC of battery.

In the second phase, when the demand is less than the generated power from the different energy sources, and the battery is wholly charged, the extra energy is sold to the utility grid and may be represented using Equation (15):

$$\begin{aligned} P_S^{grid}(t) = &\left[P_{SPV\text{-}out}^{total}(t) + P_{DG}(t)\right. \\ &\left. + \left[\left(P_{SPV\text{-}out}^{total}(t) - (SOC_{bat}^{max}(t) - SOC_{bat}(t))\right) \times \eta_{bdinv}\right]\right] \\ &- P_D^{EV}(t) \end{aligned} \tag{15}$$

where $P_S^{grid}(t)$ represents the extra energy to be sold to the grid in kWh, and $SOC_{bat}^{max}(t)$ denotes the maximum SOC of the battery.

2.6. Modeling of EVCS Load

The main components of EVCS are dual converter, charging ports, and EVs. The charging station is coupled to a regulator, which facilitates controlling the direction of the power flow at a specific time. The charging of EV takes place based on its current SOC. SOC is defined as the ratio of available charging state to the maximum charging state, i.e., when the battery is fully charged. Hence, it describes how much battery needs to be charged.

The amount of power consumed by an EV is determined by the distance traveled, battery capacity, and driving mode. An EV uses power, which may be estimated using the equation

$$P_D^{EV} = \frac{D_{km} \times E_{req/km}}{T} \tag{16}$$

where D_{km} is the number of kilometers traveled, the power needed per kilometer is $E_{req/km}$, and T is the time to recharge the battery. T is the difference between the time of arrival and departure. T is determined by the SOC of the vehicle's battery. The power requirement of an EV can be illustrated using the capacity of battery, SOC, and time of charging:

$$P_D^{EV} = \frac{Q_{bat}^{EV} \times \left(SOC_{max}^{EV} - SOC_{min}^{EV}\right)}{T} \tag{17}$$

where Q_{bat}^{EV} represents the capacity of the EV battery, SOC_{max}^{EV} and SOC_{min}^{EV} are the maximum and minimum limit of SOC, and T denotes the duration of charging.

Therefore, the total power required for charging the N number of EVs is given as follows:

$$P_{total}^{EV} = \sum_{i=1}^{N_{EV}} P_D^i \tag{18}$$

The technical features of different components of a hybrid energy system are summarized in Table 2.

Table 2. Technical and economic parameters of different hybrid system components.

Components	Characteristics	Values
Solar PV panel (Canadian Solar Max Power CS6X-325P) [6]	Rated power	325 W
	Normal operating cell temperature	45 ± 2 °C
	Open circuit voltage (V_{OC})	45.5 Volts
	Short circuit current (I_{SC})	9.34 Amps
	Module efficiency:	16.94%
	Dimensions	$76.90 \times 38.70 \times 1.57$ inches
	Power tolerance	0/+5 W
	Initial cost	USD 950/kW
	Replacement cost	USD 900/kW
	Operation and maintenance cost	USD 10/kW
	Life span	25 years
	Derating factor	80%
Diesel generator (Generic 10 kW fixed capacity genset) [46]	Minimum load ratio	50%
	Heat recovery ratio	10%
	Operating time	15,000 h
	Fuel price	USD 1.21/L
	Density	820 Kg/m^3
	Initial capital cost	USD 500/kW
	Replacement cost	USD 500/kW
	Operation and maintenance cost	USD 10
Battery (Generic 1 kWh lead acid) [47]	Battery material	Lead acid
	Nominal voltage	12 V
	Nominal capacity	81–100 Ampere hour
	Lifetime	5–8 years
	Initial capital cost	USD 235/kW
	Replacement cost	USD 190/kW
	Operation and maintenance cost	USD 2/kW/year
Bi-directional converter [47]	Rated power	100 kW
	Initial capital cost	USD 171/kW
	Replacement cost	USD 171/kW
	Operation and maintenance cost	USD 4/kW/year
Utility grid [36]	Grid purchase price	USD 0.12/kWh
	Grid sell back price	USD 0.08/kWh
Others [47]	Nominal interest rate	3.75%
	Inflation rate	1.75%
	Lifetime of project	25 years
	SOC_{max}	0.95
	SOC_{min}	0.35
	Inverter efficiency (%)	0.95
	Rectifier efficiency (%)	0.95

3. Problem Formulation

In this paper, the problem formulation deals with the optimization of different configurations of energy system for satisfying the load demand of EVCS installed in the northwest region of Delhi, India, with real annual data of solar irradiance, temperature, and EV load. These configurations are stated as follows: (a) SPV/DG/battery-based EVCS, (b) SPV/battery-based EVCS, and (c) grid-and-SPV-based EVCS. The hybrid energy system is constructed with a maximum load of 25.45 kW, and the yearly average energy requirement is 257 kWh/day. The goal of optimization is to minimize the total net present cost (TNPC), keeping the reliability constraint in terms of loss of power supply probability within specified limits. Moreover, another parameter for analyzing the performance of the hybrid energy system is the levelized cost of energy (LCOE). The MSSA is implemented to improve its searching potential based on Levy flight distribution (LFD), enabling individ-

uals to move to new locations and individual spiraling around the optimum solution to improve the exploitation phase. In this work, the energy management of the hybrid system is based on load following strategy (LFS). The numbers of SPV panels and batteries are chosen as the optimization variables, which are determined optimally using the MSSA. The primary goal of this research work is to perform a cost-effective analysis for all selected configurations with lowest TNPC, LCOE, and best reliability. The MSSA performance in the optimization of hybrid energy system is compared to conventional SSA and GWO. The objective functions, i.e., TNPC and LCOE and constraints, are coded in MATLAB environment considering the energy management strategy reported in Figures 4 and 5 for SPV/DG/battery-based EVCS and grid-and-SPV-based EVCS, respectively.

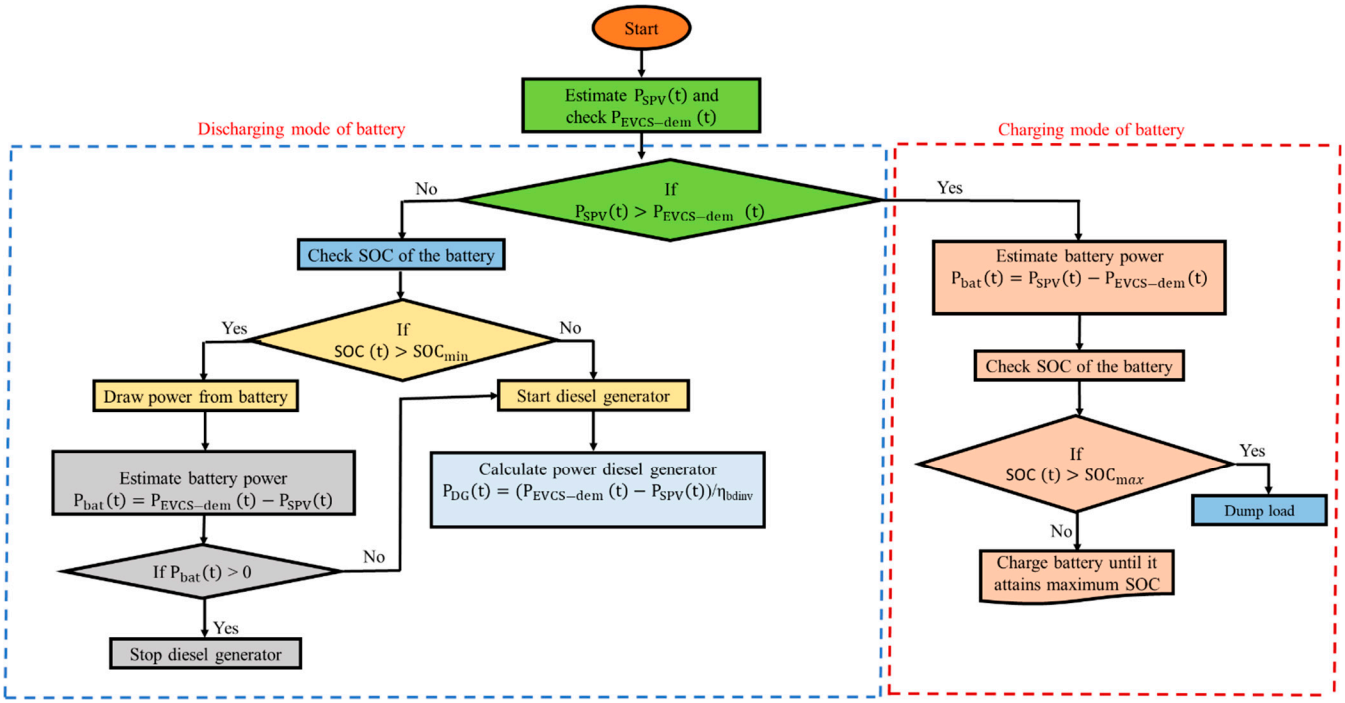


Figure 4. Flow chart of the energy management strategy of the SPV/DG/battery system.

3.1. Assumptions

- Charging stations operate for the whole day.
- Different number of EVs arrive at charging station on a daily basis.
- Only one charging station is considered.
- Limited output power of battery is taken into account.
- Load following strategy (LFS) is considered.
- Diesel generator is allowed to operate only when no power output from SPV is available.
- Cost of EV charger is around USD 50.

3.2. Objective Functions

The optimal number of solar panels, i.e., N_{SPV} and batteries, i.e., N_{bat}, must be determined to ensure that the system formulated in different configurations meets the charging demand of EVs. The objective function is formulated using the total net present cost (TNPC), levelized cost of energy (LCOE), and reliability in terms of loss of power supply probability (LPSP).

3.2.1. Total Net Present Cost (TNPC)

The total cost of a hybrid energy system is represented by the total net present cost (TNPC). It includes all costs and revenues incurred during the life cycle of the system, including system component capital costs, replacement costs incurred during the system's

operation, and maintenance expenses. The numbers of SPV panels and batteries are chosen as the two primary decision variables for the optimum configuration. Hence, mathematical equations for the NPC of various system components are derived as follows:

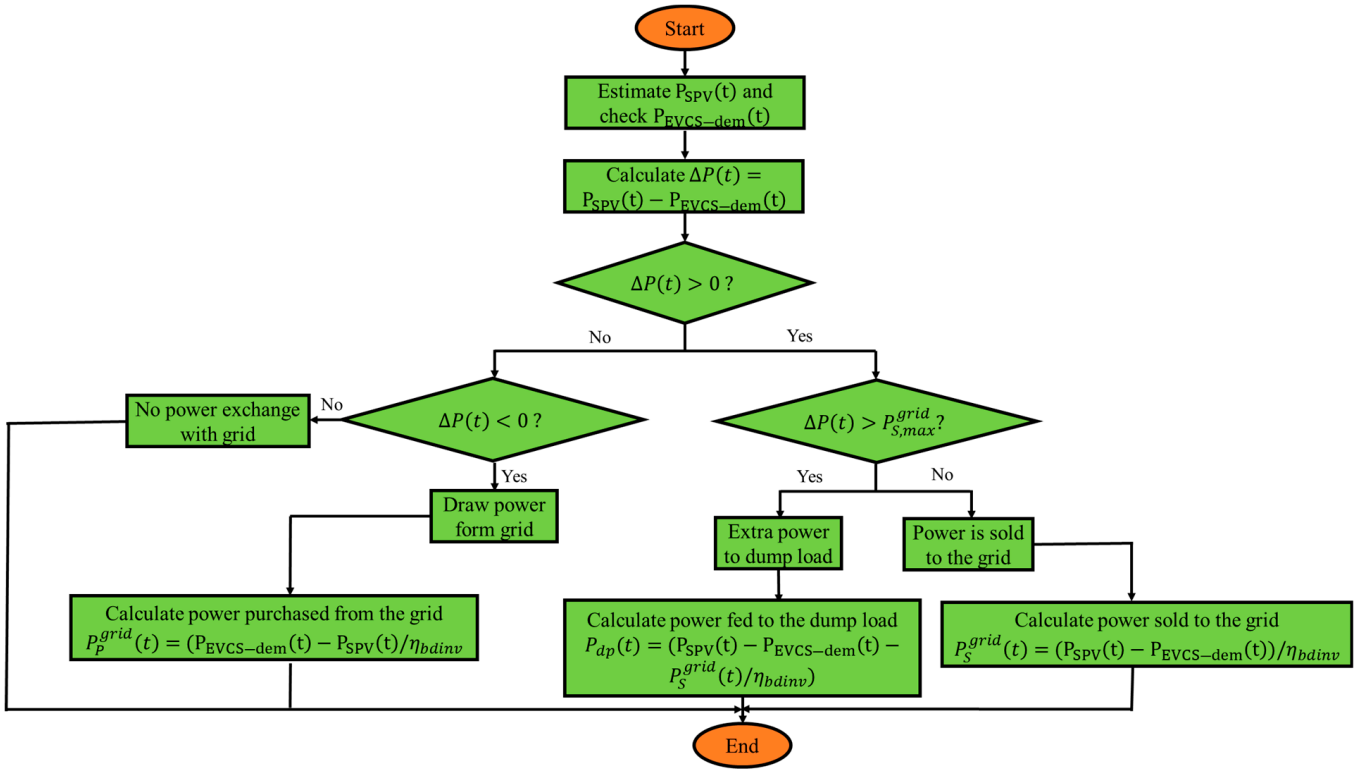


Figure 5. Flow chart of the energy management strategy in grid-and-solar PV-based EVCS.

Solar PV

The project lifetime is set to be equivalent to the lifetime of PV panels in this study. Furthermore, solar irradiance is used as a free fuel to generate power. Therefore, there will be no fuel costs. There, NPC of solar PV (NPC_{SPV}) consists of capital cost $\left(C_{SPV}^{CAP}\right)$, replacement cost $\left(C_{SPV}^{REP}\right)$, operation and maintenance cost $\left(C_{SPV}^{OM}\right)$, and salvage cost $\left(C_{SPV}^{SAL}\right)$ and can be given as follows

$$NPC_{SPV} = C_{SPV}^{CAP} + C_{SPV}^{REP} + C_{SPV}^{OM} - C_{SPV}^{SAL} \tag{19}$$

The capital cost of solar PV panels can be computed as

$$C_{SPV}^{CAP} = N_{SPV} \times \psi_{SPV} \times P_{SPV} \tag{20}$$

where ψ_{SPV} and P_{SPV} denote the initial cost and power rating of each SPV panel, respectively. Replacement cost, i.e., $\left(C_{SPV}^{REP}\right)$ of SPV panel, can be determined using the following equation:

$$C_{SPV}^{REP} = \zeta_{SPV}^{rep} \times N_{SPV} \times P_{SPV} \times \frac{R(1+R)^{\Omega}}{\left((1+R)^{\Omega} - 1\right)(1+R)^{\Omega_{SPV}}} \tag{21}$$

where ζ_{SPV}^{rep} denotes the cost of replacing the SPV panel, R denotes the rate of interest, Ω indicates the life span of the project in years, and y represents the lifetime of the SPV panel.

Further, operation and maintenance cost $\left(C_{SPV}^{OM}\right)$ can be computed as follows:

$$C_{SPV}^{OM} = N_{SPV} \times \omega_{SPV} \times P_{SPV} \times \sum_{i=1}^{\Omega} \left(\frac{1+\mu_{SPV}}{1+R}\right)^{i} \tag{22}$$

where ω_{SPV} is the OM cost of each panel, and μ_{SPV} is the escalation rate.

Salvage value is the cost of the component's remaining life after the project is completed. The salvage cost of the SPV panel can be defined as

$$C_{SPV}^{SAL} = N_{SPV} \times \varepsilon_{SPV} \times \left(\frac{1+\partial}{1+R}\right)^{\Omega} \times P_{SPV} \tag{23}$$

where ε_{SPV} denotes the resale price of the solar panel after completing its life, and ∂ is the inflation rate.

Battery

The NPC of battery storage system, i.e., NPC_{bat}, involves capital cost $\left(C_{bat}^{CAP}\right)$, replacement cost $\left(C_{bat}^{REP}\right)$, operation and maintenance cost $\left(C_{bat}^{OM}\right)$, and salvage cost $\left(C_{bat}^{SAL}\right)$ and can be given as follows

$$NPC_{bat} = C_{bat}^{CAP} + C_{bat}^{REP} + C_{bat}^{OM} - C_{bat}^{SAL} \tag{24}$$

$$C_{bat}^{CAP} = N_{bat} \times \psi_{bat} \tag{25}$$

$$C_{bat}^{OM} = N_{bat} \times \omega_{bat} \times \sum_{i=1}^{\Omega} \left(\frac{1+\mu_{bat}}{1+R}\right)^{i} \tag{26}$$

$$C_{bat}^{SAL} = N_{bat} \times \varepsilon_{bat} \times \left(\frac{1+\partial}{1+R}\right)^{\Omega} \tag{27}$$

where ψ_{bat} is the cost of one battery in USD. ω_{bat} and μ_{bat} indicate annual OM cost in USD/year and escalation rate (0.075) of batteries, respectively. ε_{bat} denotes the resale value of one battery in USD.

The battery life (Ω_{bat}) has been considered to be 5 years in this study, which is shorter than the project lifetime of 25 years. Hence, the battery must be replaced every 5 years. The number of battery replacements (NBR) is determined as

$$NBR = \frac{\Omega}{\Omega_{bat}} - 1 \tag{28}$$

The replacement cost of the battery can be determined as

$$C_{bat}^{REP} = N_{bat} \times \zeta_{bat}^{rep} \times \sum_{i=5,\ 10,\ 15,\ 20} \left(\frac{1+\mu_{bat}}{1+R}\right)^{i} \tag{29}$$

Bi-Directional Inverter

The NPC of the bi-directional inverter can be calculated as

$$NPC_{bdinv} = C_{bdinv}^{CAP} + C_{bdinv}^{REP} + C_{bdinv}^{OM} - C_{bdinv}^{SAL} \tag{30}$$

where C_{bdinv}^{CAP}, C_{bdinv}^{REP}, C_{bdinv}^{OM}, and C_{bdinv}^{SAL} represent the initial capital cost, replacement cost, operation and maintenance cost, and salvage cost of the bi-directional inverter, respectively.

$$C_{bdinv}^{CAP} = N_{bdinv} \times \psi_{bdinv} \tag{31}$$

$$C_{bdinv}^{OM} = N_{bdinv} \times \omega_{bdinv} \times \sum_{i=1}^{\Omega} \left(\frac{1 + \mu_{bdinv}}{1 + R} \right)^{i} \tag{32}$$

$$C_{bdinv}^{SAL} = N_{bdinv} \times \varepsilon_{bdinv} \times \left(\frac{1 + \partial}{1 + R} \right)^{\Omega} \tag{33}$$

The inverter's assumed lifetime is 10 years, which is shorter than the project's lifespan (25 years). As a result, the net present value of the replacement cost of the inverter is expressed as follows:

$$C_{bdinv}^{REP} = N_{bdinv} \times \zeta_{bdinv}^{rep} \times \sum_{i=10,20} \left(\frac{1 + \mu_{bdinv}}{1 + R} \right)^{i} \tag{34}$$

Sale and Purchase Capacity with Utility Grid

In an on-grid system, the NPC of selling electricity to the grid or purchasing electricity from the grid can be evaluated as follows:

$$C_{S}^{grid} = \xi_{S}^{grid} \times E_{S}^{grid} \times \left(\frac{1 + \partial_{grid}}{1 + R} \right)^{\Omega} \tag{35}$$

$$C_{P}^{grid} = \xi_{P}^{grid} \times E_{P}^{grid} \times \left(\frac{1 + \partial_{grid}}{1 + R} \right)^{\Omega} \tag{36}$$

where ξ_{S}^{grid} and ξ_{P}^{grid} indicate the cost for unit sale and of purchase of electricity to and from the utility grid, respectively. It is defined in USD per kWh.

Hence, one of the objectives of this research work is to minimize the TNPC of the system, which is described as follows:

$$TNPC(USD) = NPC_{SPV} + NPC_{DG} + NPC_{bat} + NPC_{bdinv} + C_{P}^{grid} - C_{S}^{grid} \tag{37}$$

3.2.2. Levelized Cost of Energy (LCOE)

Levelized cost of energy may be defined as the average cost incurred per kW of energy production by the system. LCOE is computed by dividing the annualized cost of energy generation by the total energy production per year. It can be computed with the help of capital recovery factor (CRF), using the following Equation (38) [10]:

$$LCOE = \frac{TNPC \times CRF}{\sum_{t=1}^{T} P_{gen}(t)} \text{ USD/kW} \tag{38}$$

CRF can be evaluated using Equation (39), as follows:

$$CRF = \frac{R(1 + R)^{\Omega}}{\left((1 + R)^{\Omega} - 1 \right)} \tag{39}$$

LCOE depends on numerous features, such as capital cost, solar radiation, lifetime, operation and maintenance cost, CRF and degradation of the SPV panels used, etc.

3.2.3. Renewable Fraction

The renewable fraction (RF) is an important criterion to consider when designing an energy system. The total energy produced via RE sources is divided by the total power drawn by the load to calculate the RF. RF can be calculated using Equation (40):

$$RF = \left(1 - \frac{\sum P_{DG}}{P_{SPV}} \right) \times 100\ \% \tag{40}$$

3.3. Decision Variables and Constraints

To optimize the hybrid system, a compromise must be made between the cost-based objectives and various technical constraints. However, the proposed optimization approach should be used to obtain the optimal decision variables. The decision variable of the suggested optimization procedure is subject to specific limits: the minimum limit put on the number of SPVs and battery, and the N_{SPV}^{max} = 1000, N_{bat}^{max} = 600.

- The following are the limits associated with the maximum and minimum sizes of decision variables:

$$N_{SPV}^{min} \leq N_{SPV} \leq N_{SPV}^{max} \tag{41}$$

$$N_{bat}^{min} \leq N_{bat} \leq N_{bat}^{max} \tag{42}$$

$$E_{bat}^{min} \leq N_{bat} \leq E_{bat}^{max} \tag{43}$$

- The charging state of the battery should be preserved using Equation (44). At any given time, the amount of energy stored in the battery $SOC_{bat}(t)$ is limited by the maximum and minimum quantities SOC_{bat}^{min} and SOC_{bat}^{max} as follows:

$$SOC_{bat}^{min} \leq SOC_{bat} \leq SOC_{bat}^{max} \tag{44}$$

SOC_{bat}^{max} takes the value of the theoretical capacity of the battery $C_{bus} \times V_{bus}$. The maximum depth of discharge, i.e., DOD and theoretical capacity $C_{bus} \times V_{bus}$ is used to calculate SOC_{bat}^{min}, as expressed in Equation (45):

$$SOC_{bat}^{min} = (1 - DOD) \times C_{bus} \times V_{bus} \tag{45}$$

DOD is assumed to be 50%.

- The maximum permissible lack of power supply should be considered for a reliable system:

$$LPSP \leq LPSP_{max} \tag{46}$$

where $LPSP_{max}$ represents the maximum allowable lack of power supply probability.

3.4. Operational Strategy

3.4.1. Solar PV and Diesel Generator-Based EVCS

To attain the reliability of the system in any hybrid energy system, optimal energy management is required. In this system, the DG is kept at the bottom of that list, which means it only needs to operate when the solar and battery systems are unable to supply the load requirement. The steps for the simplest implementation strategy for SPV/DG/battery system are as follows:

- If the total power generated by SPV panels is sufficient, demand can be met solely by solar power. After the load has been satisfied, excess electricity can be supplied to the BESS and is provided as follows:

$$P_{bat}(t) = P_{SPV\text{-}out}(t) - P_{EVCS\text{-}dem}(t) \tag{47}$$

- In the preceding case, if $P_{bat}(t)$ is greater than the maximum permissible capacity of the BESS, i.e., P_{bat}^{max}, then additional energy can be dumped or directed to deferrable loads. Excess or dump energy $(P_{dp}(t))$ can be given as

$$P_{dp}(t) = P_{bat}(t) - P_{bat}^{max}(t) \tag{48}$$

- If the SPV panels do not generate sufficient power, the battery can provide the remaining power, which can be calculated as

$$P_{bat}(t) = P_{EVCS\text{-}dem}(t) - P_{SPV\text{-}out}(t) \quad (49)$$

- If solar power is insufficient and batteries are unable to generate the required power to meet the power requirements, DG is used to energize the load. There are two ways to use DG.

(a) First, it employs a load-following strategy, which means that whenever it performs, it produces only the power necessary to satisfy the primary load requirements. The diesel generator's power output is calculated as

$$P_{DG}(t) = \frac{P_{EVCS\text{-}dem}(t) - P_{SPV\text{-}out}(t)}{\eta_{bdinv}} \quad (50)$$

(b) In the second strategy, it runs at maximum capacity or minimum load ratio. When the DG is fully operational, the surplus energy is used to charge batteries and is expressed as follows:

$$P_{bat}(t) = (P_{DG}(t) * \eta_{bdinv} - P_{EVCS\text{-}dem}(t)) + P_{SPV\text{-}out}(t) \quad (51)$$

The operational strategy of the SPV/DG/battery system can be understood with the help of the flow chart depicted in Figure 4.

3.4.2. Grid-and-Solar PV-Based EVCS

In this scenario, it is assumed that SPV is the only source to satisfy the load demand of EVs. The difference between power obtained by SPV and power demand by EVs can be determined as

$$\Delta P(t) = P_{SPV}(t) - P_{EVCS\text{-}dem}(t) \quad (52)$$

If solar panels are unable to fulfill the power requirements of EV load, the extra power is purchased from the grid network ($P_P^{grid}(t)$) to meet the requirements.

Furthermore, if more power is available from the SPV after meeting the requirements, it is sold to the grid ($P_S^{grid}(t)$). Nevertheless, there are some limitations on selling and purchasing power to and from the grid, which are defined as maximum purchase capacity of grid ($P_{P,\,max}^{grid}$) and maximum selling capacity of grid ($P_{S,\,max}^{grid}$). Beyond these limits, power cannot be purchased from or sold to the grid. The following cases are formed depending on the $\Delta P(t)$.

- When SPVs alone fulfill the power requirement of EV load, i.e., $\Delta P(t) > 0$, the extra power is sold to the grid network, which can be determined as

$$P_S^{grid}(t) = (P_{SPV}(t) - P_{EVCS\text{-}dem}(t))/\eta_{bdinv} \quad (53)$$

- When SPV power output is high enough, which fulfills the power requirement of EV load, as well as exceeds the maximum selling capacity of grid ($P_{S,\,max}^{grid}$), i.e., $\Delta P(t) > 0$ and $\Delta P(t) > P_{S,\,max}^{grid}$, the extra power is fed to the dump load, which can be determined as

$$P_{dp}(t) = (P_{SPV}(t) - P_{EVCS\text{-}dem}(t) - P_S^{grid}(t)/\eta_{bdinv}) \quad (54)$$

- When SPV power is unable to satisfy the charging demand of EVs, i.e., $\Delta P(t) < 0$, power is purchased from the grid which can be computed as

$$P_P^{grid}(t) = (P_{EVCS\text{-}dem}(t) - P_{SPV}(t)/\eta_{bdinv} \quad (55)$$

- If $\Delta P(t) = 0$, there is no exchange of power from grid, and SPV power fulfills the load requirement.

- When both SPV and grid are unable to satisfy the load demand, deficiency of power takes place, which can be given as

$$P_{def}(t) = P_P^{grid}(t) - P_{P,\,max}^{grid} \quad (56)$$

$P_{def}(t)$ must be zero to make sure the total power requirement of EV load demand is served in a reliable manner when minimizing the LPSP.

The LPSP can be kept within a specific tolerance band to solve the optimization problem. In this paper, the maximum limit of LPSP, i.e., $LPSP_{max}$ is assumed to be 1%. The energy management strategy for grid-and-solar PV-based EVCS can be demonstrated with the help of the flow chart shown in Figure 5.

4. Proposed Modified Salp Swarm Algorithm for Sizing of Hybrid Energy System

A brief discussion of the SSA and MSSA algorithms is presented in this section because they have yet to be examined for the application of optimal sizing of different energy sources for the designing of EVCS load in the northwest region of Delhi, India.

It is worth noting that the SSA is a robust heuristic algorithm that has been used to solve a variety of problems. However, it is susceptible to local optima in some instances and suffers from stagnation. To render and combat the shortcomings of the conventional SSA, an MSSA is presented. The suggested MSSA employs two different searching procedures to improve its searching potential. The first technique is to use Levy flight distribution (LFD) to improve the searching process, enabling individuals to move to new locations. The second technique relies on the individual spiraling around the optimum solution to improve the exploitation phase. Algorithm 1 [48] and Algorithm 2 [49] contain the pseudocode for the traditional SSA and the modified MSSA, respectively, whereas Figure 6 depicts the graphical representation of MSSA for SPV/DG/battery system optimization.

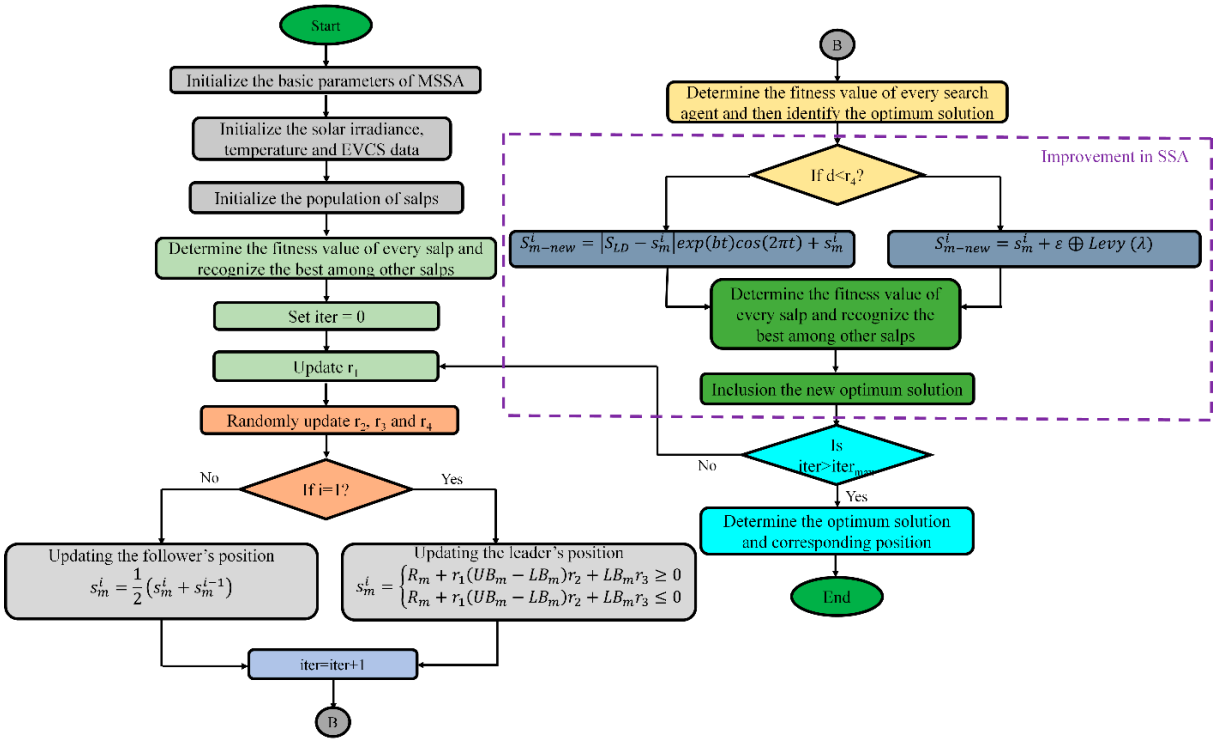


Figure 6. Graphical representation of MSSA for SPV/DG/battery system optimization.

4.1. Salp Swarm Algorithm (SSA)

In 2017, S. Mirjalili presented an individual-based algorithm known as salp swarm algorithm [48]. SSA models the mechanism of swarm salps, which will be scavenged in

oceans. The salp at the front of the chain will be the leader, while the other salps will be considered followers in the ocean.

Like that of other meta-heuristic algorithms, the position of salp can be defined in n dimensions, where n is the number of variables for a specific problem. Hence, the positions of all salps are maintained in a two-dimensional matrix known as s. For example, it is assumed that the food source (R) is available in search space and considered as the target of all salps in a swarm. So, the describing equations of SSA are given as follows:

$$s_m^i = \begin{cases} R_m + r_1(UB_m - LB_m)r_2 + LB_m r_3 \geq 0 \\ R_m + r_1(UB_m - LB_m)r_2 + LB_m r_3 \leq 0 \end{cases} \tag{57}$$

where s_m^i indicates the position of leader, while R_m represents the position of the food source. UB_m and LB_m denote the upper and lower bound in m-th dimension of search space.

The coefficient r_1 plays a vital role in SSA because it maintains a proper balance between exploration and exploitation capability and can be expressed as follows:

$$r_1 = 2 \times \exp\left(-\left(\frac{4p}{P}\right)^2\right) \tag{58}$$

where p denotes the current iteration, while P represents the maximum iterations.

The parameters r_2 and r_3 are uniformly produced random values, which lie in the range of 0 and 1. They specify whether the next point in the m-th dimension should be toward positive or negative infinity, as well as the step size.

The positions of the follower salps are updated in accordance with the newton second law of motion, which is stated as follows:

$$s_m^i = \frac{1}{2}at^2 + u_o t \tag{59}$$

where $i \geq 2$ and s_m^i depict the position of i-th follower salp in m-th dimension, u_o denotes the initial velocity, and $a = \frac{u_{final}}{u_o}$, where $u = \frac{x - x_o}{t}$.

The simulation time is determined by the number of iterations, and if the conflict between iterations is equal to 1, and assuming $u_o = 0$, this equation will be as follows:

$$s_m^i = \frac{1}{2}\left(s_m^i + s_m^{i-1}\right) \tag{60}$$

The salp chains can be modeled with the help of Equations (57)–(60).

Algorithm 1: Pseudocode of conventional SSA method

Start SSA

Step 1: Fix the basic parameters of SSA algorithm, such as maximum iterations (P), population number, the upper and lower limits of the decision variables
Step 2: Randomly initialize the population of salps individuals
Step 3: Calculate the fitness function for each salp
Step 4: Set iteration p = 1
Step 5: Update r_1 according to Equation (58)
Step 6: For every member of salps, if i == 1, update the position of leading salp by Equation (57)
Otherwise, update the position of follower salp by Equation (60)
Fitness calculation of every salp : Update (S^*) if there is a superior solution
Increment t to 1
Repeat steps 5 to 7 until $p = P_{max}$ is satisfied
Step 7: Return the best solution (S^*) and its fitness value F (S^*)

End SSA

4.2. Modified Salp Swarm Algorithm (MSSA)

The MSSA is intended to address the shortcomings of the traditional SSA, such as its proclivity to local optima and stagnation in some circumstances [49]. To improve the capability of the traditional SSA, two search techniques are described. The first adjustment is used to speed up the exploration process by changing the positions of salps using the Levy flight distribution (LFD) to help the population find new places and break the cycle of stagnation. The second change entails changing and updating the locations of salps in the spiral path around the optimum solution to improve its exploitation. The first change, which is based on the LFD, is written as follows:

$$S^{i}_{m-new} = s^{i}_{m} + \varepsilon \oplus \text{Levy } (\lambda) \tag{61}$$

where ε denotes a step size parameter that is set as follows:

$$\varepsilon \oplus \text{Levy } (\lambda) \sim 0.01 \frac{e}{|f|^{1/\lambda}} \left(S^{i}_{m} - S_{LD} \right) \tag{62}$$

where e and f can be obtained as follows:

$$e \sim N\left(0, \varnothing^{2}_{e}\right), \; f \sim N\left(0, \varnothing^{2}_{f}\right) \tag{63}$$

$$\varnothing_{e} = \left[\frac{\Gamma(1+\lambda) \times \sin(\pi \times \lambda/2)}{\Gamma\left[(1+\lambda)/2\right] \times \lambda} \right]^{1/\lambda}, \; \varnothing_{f} = 1 \tag{64}$$

where Γ indicates the standard gamma function. The second modification is based on modifying the locations of salps in the spiral path around the best solution, which can be expressed as follows:

$$\varepsilon\varepsilon \oplus \text{Levy } (\lambda) \sim 0.01 \frac{e}{|f|^{1/\lambda}} \left(S^{i}_{m} - S_{LD} \right) \tag{65}$$

The second adjustment is to change the salp locations in the spiral path around the optimum solution, which may be represented as follows:

$$S^{i}_{m-new} = \left| S_{LD} - s^{i}_{m} \right| \exp(bt) \cos(2\pi t) + s^{i}_{m} \tag{66}$$

where b is a number used to characterize the shape of a logarithmic spiral. The adaptive operator d is used to perform the job of balancing the exploration and exploitation processes:

$$d(t) = d_{min} + \left(\frac{d_{max} - d_{min}}{P_{max}} \right) \times p \tag{67}$$

where d_{max} and d_{min} are the maximum and minimum values of adaptive operator d. Lastly, the MSSA's exploring potential is enhanced by improving the exploration ability with the LFD by applying Equation (59) at the first iteration when the value of d is small, and the exploitation capability can be enhanced with the variable bandwidth transition by applying Equation (63) at the last iteration when the value of d is large.

Algorithm 2: Pseudocode of modified SSA method

Start MSSA

Step 1: Fix the basic parameters of SSA algorithm, such as maximum iterations (P), population number, the upper and lower limits of the decision variables
Step 2: Randomly initialize the population of salps individuals
Step 3: Calculate the fitness function for each salp
Step 4: Set iteration p = 1
Step 5: Update r_1 according to Equation (58)
Step 6: For every member of salps, if i == 1, update the position of leading salp by Equation (57)
Otherwise, update the position of follower salp by Equation (60)
Fitness calculation of every salp: Update (S^*), if there is a superior solution
Increment t to 1
Repeat steps 5 to 7 until $p = P_{max}$ is satisfied
Step 7: Arranging the salp according to their fitness values
Step 8: Update the d value according to Equation (67)
Step 9: Compare the value of d with a random value r_4
If the value of d is less than r_4
Update the salp positions based on the spiral orientation according to Equation (66)
Else
Update the salp positions based on Levy distribution according to Equation (61)
End
Step 10: Calculate the fitness function of the updated salps. Then, include the new solutions if these solutions are better than the solutions of the previous steps
Increment t to 1
Repeat steps 5 to 7 until ($p = P_{max}$) is satisfied
Step 11: Return the best solution (S^*) and its fitness value F (S^*)

End MSSA

5. Results and Discussion

5.1. Components and Methods

This section contains the important components, such as the place of the study area, resources available, and the load demand of the developed framework.

5.1.1. Area under Study

The proposed study is conducted in the northwest region of Delhi, India. The geographical coordinates of the studied area are 28.7408° N (latitude), 77.1126° E (longitude). Figure 7 depicts the study area from a geographical standpoint. The population of Rohini sector-17 is 21,460 individuals. Because the majority of citizens in this area are cultured and working, EVs are employed as a mode of transportation to get to their destinations on a day-to-day basis. This area has nearly 120–150 EVs, and there are 15 facilities in which these vehicles are kept and provided with electricity. As per the owner of different facilities, the amount required to charge the EVs is significantly high because these vehicles are charged using the grid.

5.1.2. Assessment of the Available Energy Sources

Solar radiation in Delhi is at its highest from March to October. It is available during the wintertime, but not at its high point. Additionally, it is regarded as an outstanding source of energy due to its practical features. Figure 8 shows the annual variation of irradiance at the chosen area. Figure 8 also depicts the clearness index, which is the percentage of solar radiation that is reflected by the atmosphere. It can also be defined as the ratio of surface irradiance and extraterrestrial irradiance. Figure 9 portrays the annual fluctuation of temperature for the study site. The DG is also incorporated as one of the energy providers in the proposed research work. Energy generation from DG is comparatively easy and cost effective as well. Thus, DG is used for charging EVs, especially during peak hours.



Figure 7. The geographical view of the study area.

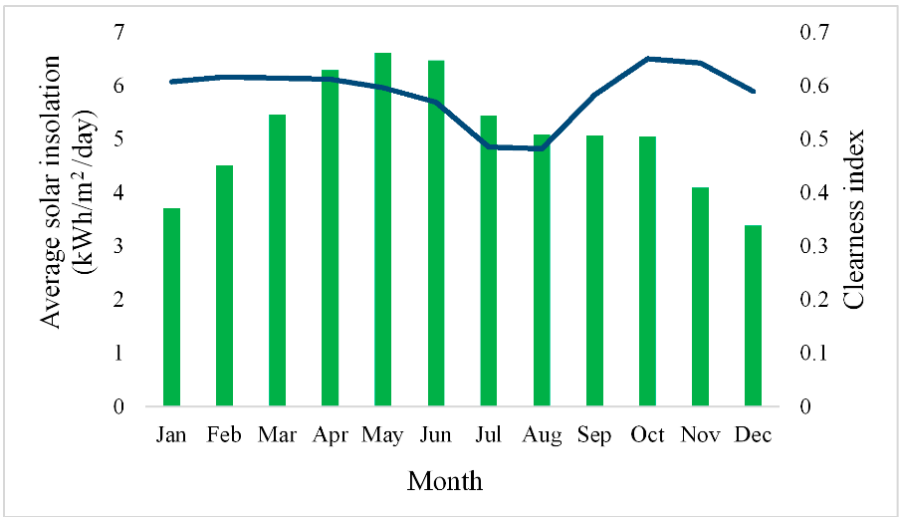


Figure 8. Annual variation of solar insolation and clearness index of the selected area.

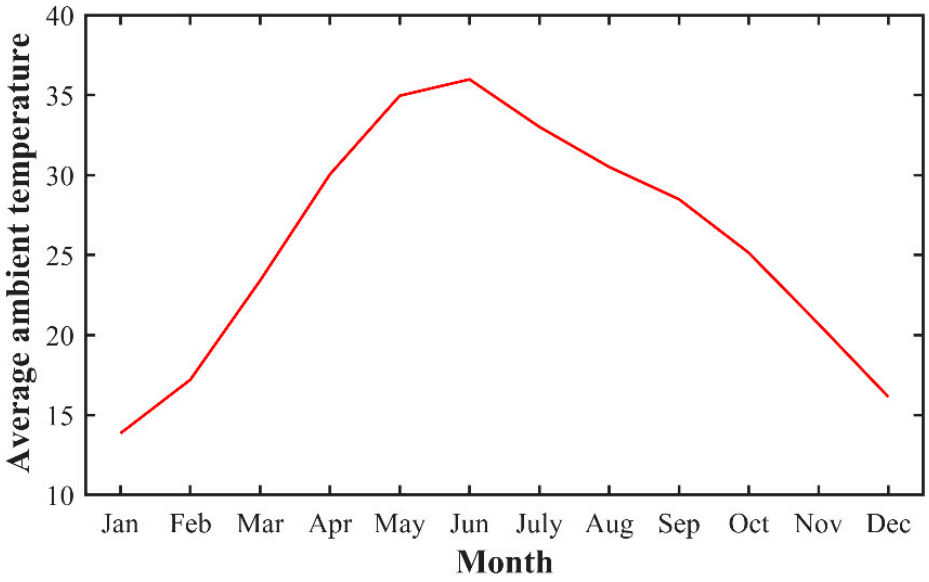


Figure 9. Annual variation of the mean temperature of the selected area.

5.1.3. Assessment of the EVCS Load in the Selected Area

The research study considers three-wheeler EVs, i.e., e-rickshaw, to be the system's load. Four 12-Volt lead-acid batteries with capacities of 160 to 180 Ampere hour are used in a three-wheeler. The theoretical battery energy is 2.2 kWh when the 12 V and 180 Ampere hour battery is considered. Generally, a 12 V battery is recharged to 90% of its design capacity. As a result, the battery's energy would be 1.94 kWh. The total energy consumption for four 12 V battery vehicles, such as e-rickshaws, would be 7.8 kWh. Taking the system's 15% system losses into account, the energy consumed is determined to be 9 kWh. Thus, the energy consumed by every e-rickshaw per day is 9 kWh. According to the research findings, the yearly average energy is 257 kWh/day, the average power is 10.71 kW, and the maximum load is 25.45 kW. A graphical representation of the EVCS load on an hourly and daily basis is shown in Figures 10 and 11, respectively.

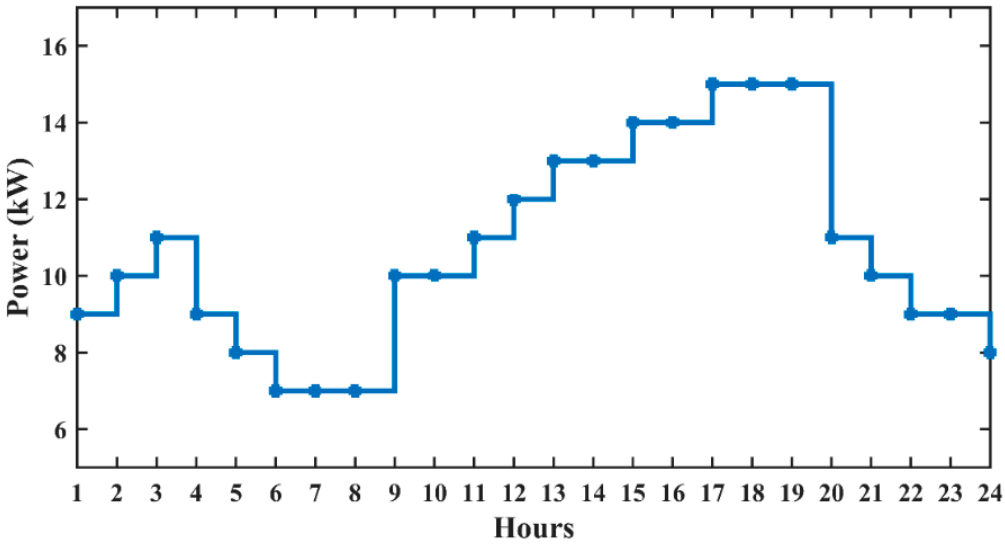


Figure 10. EVCS load profile on a daily basis.

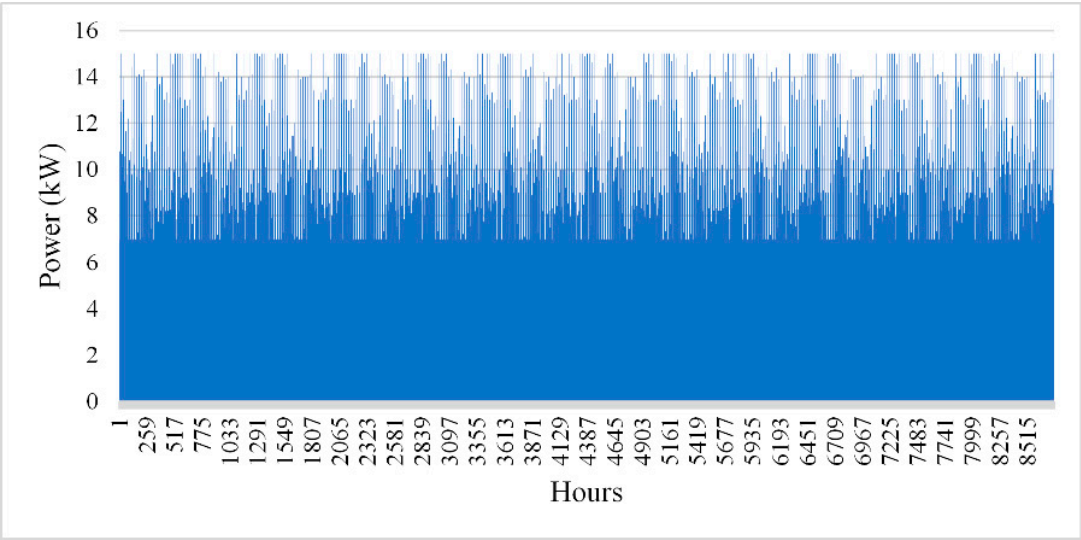


Figure 11. EVCS load profile on an hourly basis.

5.2. Implementation of Proposed MSSA Technique on Benchmark Functions

This section explores the application of the proposed MSSA technique on standard benchmark functions. Table 3 contains comprehensive descriptions of these functions. The proposed optimization strategy outperforms SSA and GWO in terms of maximum iterations and optimal solution of seven mathematical benchmark functions for 30 independent runs. MSSA adheres to a globally optimal solution without becoming stuck in a local optimal solution, resulting in faster convergence. Table 4 displays the numerical outcomes in terms of optimum solution and number of iterations.

Table 3. Standard benchmark functions.

S. No	Function Name	Mathematical Formulation	Dimension (D)	Search Range
1.	Sphere	$f(x) = \sum_{p=1}^{D} x_p^2$	30	$[-100, 100]$
2.	Rosenbrock	$f(x) = \sum_{p=1}^{D} \left[100 \times \left(x_p^2 - x_{p+1}\right)^2 + \left(1 - x_p\right)^2\right]$	30	$[-2.048, 2.048]$
3.	Rastragin	$f(x) = \sum_{p=1}^{D} \left[x_p^2 - 10\cos\left(2\pi x_p\right) + 10\right]$	30	$[-5.12, 5.12]$
4.	Greiwank	$f(x) = \frac{1}{4000}\sum_{p=1}^{D} x_p^2 - \prod_{p=1}^{D} \cos\left(\frac{x_p}{\sqrt{p}}\right) + 1$	30	$[-600, 600]$
5.	Schewefel	$f(x) = \sum_{p=1}^{D}\left(\sum_{q=1}^{p} x_q\right)^2$	30	$[-100, 100]$
6.	Ackley	$f(x) = -20e^{\left(-0.2\sqrt{\frac{1}{D}\sum_{p=1}^{D} x_p^2}\right)} - e^{\left(\frac{1}{D}\sum_{p=1}^{D}\cos\left(2\pi x_p\right)\right)} + 20 + e^1$	30	$[-32.76, 32.76]$
7.	Alpine	$f(x) = \sum_{p=1}^{D-1}\left\lvert x_p \sin x_p + 0.1x_p\right\rvert$	30	$[-10, 10]$

Table 4. Numerical results obtained using MSSA and their comparison with SSA and GWO.

Function Name	MSSA		SSA		GWO	
	Maximum Iterations	Optimal Solution	Maximum Iterations	Optimal Solution	Maximum Iterations	Optimal Solution
Sphere	1000	1.55×10^{-23}	897	5.89×10^{-19}	1000	6.25×10^{-19}
Rosenbrock	415	2.72×10^{-3}	537	5.81×10^{1}	548	4.57×10^{2}
Rastragin	338	7.57×10^{-14}	458	5.53×10^{-3}	495	2.53×10^{-2}
Greiwank	973	4.25×10^{-1}	1000	3.45×10^{-3}	1000	4.87×10^{-1}
Schewefel	1000	8592.6	984	7956.4	1000	6085.8
Ackley	1000	7.74×10^{-7}	785	9.14×10^{-6}	1000	8.54×10^{-3}
Alpine	1000	−192.73	1000	−192.45	1000	−192.50

5.3. Simulation Outcomes and Main Findings

This section evaluates the technical and economic parameters for three different scenarios using HOMER software and the proposed MSSA technique. These scenarios are SPV- and DG-based EVCS, only SPV-based EVCS, and grid-and-SPV-based EVCS. The objective is to achieve the best EVCS design by reducing the total net present cost, levelized cost of energy while keeping the lack of power supply probability within limits, accounting for environmental emissions, and different energy supply options. The outcomes are estimated in terms of TNPC, LCOE, SPV energy, the contribution of DG, battery and converter capacity, additional electricity generation, and the proportion of renewables utilized. The study then compares a hybrid energy system with a grid connection to a standalone system. The analysis of the obtained results for different scenarios is summarized in the latter part of this section. Furthermore, to demonstrate the efficacy of meta-heuristic techniques, different algorithms are used to analyze the results and are compared and contrasted with HOMER results. The three different scenarios considered are as follows:

- Scenario 1: Solar PV and diesel generator–based EVCS

This scenario addresses "range anxiety," a prevalent consideration among EV users about the vehicle's range. EV users can schedule long trips with greater certainty if an EVCS is as easily accessible as a petrol pump. Hence, the layout of an EVCS on roadways with SPV, DG, and a BESS with the different electricity providers is investigated and considered as Scenario 1. The goal is to reduce the capital, operation and maintenance, replacement, and fuel cost of every component associated with the system. The size and number of the SPV, DG, battery, and converter are considered to be the decision variable of the optimization problem.

- Scenario 2: Only Solar PV–based EVCS

In some countries, such as Saudi Arabia, the DG acts as a primary source of energy. In such cases, the diesel option is a good example. On the other hand, DG units are prohibitively expensive due to the high costs of maintenance, fuel supply, and fuel transport. Furthermore, DGs emit a lot of pollution. Hence, SPV- and battery-based EVCS are also being investigated and considered as Scenario 2. In this scenario, the size and number of solar PVs and batteries are the only decision variables.

- Scenario 3: Grid-and-solar PV–based EVCS

In this scenario, the EVCS is supposed to be grid-connected, and solar PV, battery, and grid are the various energy supply sources. Because EVCS is grid connected, it can purchase/sell power from/to the grid network, and the design is improved to account for the EVCS's net costs. The price of electricity provided in the Feed-in Tariff program of New Delhi, India, is taken into account for drawing/selling power from/to the utility grid. Feed-in Tariffs are the rates charged to RE providers for the energy provided by their power plants. The evaluating model is distinguished by project length and technology method and offers a profitable return on investment. The Feed-in-Tariff is a standard rate for the

buying and selling of power from and to the grid network throughout the agreement, whereas the Time-of-Use Tariff is only pertinent to buyers, in contrast to Feed-in Tariffs, and varies throughout the day over time. In this work, Feed-in Tariffs are considered.

5.3.1. Techno–Economic Analysis of Different Scenarios Using HOMER Software

HOMER is used to model the hybrid energy system to meet the load demand. HOMER defines weather information, different parametric values of components used, and EVCS load profile. Many different assumptions are made to attain the optimal plan that decreases the TNPC, LCOE, while maintaining restrictions, such as supply continuity. It is worth noting that the information related to cost is indicated in USD.

Table 5 shows the results obtained for Scenario 1, i.e., SPV/DG/battery system. SPV makes a significant contribution of 122,864 kWh per year and operates for 4369 h per year, while the DG constitutes 2767 kWh per year and functions for 364 h per year. To operate optimally, this system makes use of 590 batteries and a 12.1 kW bi-directional converter. The obtained system has a TNPC of USD 638,917.29, an initial capital cost of USD 375,855, and an operating cost of USD 16,144.69. The LCOE for this configuration is calculated to be USD 0.6328/kWh. Furthermore, the NPC of SPV and diesel generator is 42.8 percent and 11.1 percent of TNPC, respectively, whereas the NPC of the bi-directional converter is only 0.58 percent of TNPC.

Table 5. Results attained using HOMER software for Scenario 1.

SPV (kWh/Year)	DG (kWh/Year)	Batteries Used	Converter (kW)	Initial Capital Cost (USD)	Operating Cost (USD)	TNPC (USD)	LCOE (USD/kWh)	% of Renewables Utilized
122,864	2767	590	12.1	375,855	16,144.69	638,917.29	0.6328	96.5

Similarly, Tables 6 and 7 display the outcomes of Scenarios 2 and 3, respectively. When the only energy-producing source is an SPV and a battery, the LCOE is USD 0.7318/kWh, the TNPC is USD 875,481.24. In Scenario 3, when the grid is used to meet load demand, the LCOE is USD 0.587/kWh, and the TNPC is USD 275,349. In this scenario, the energy produced by SPV is 126,811 kWh per year, accounting for 71.3 percent of total energy production, while energy purchased from the grid is 51,155 kWh per year, accounting for 28.7 percent of total energy generation.

Table 6. Results attained using HOMER software for Scenario 2.

SPV (kWh/Year)	Batteries Used	Converter (kW)	Initial Capital Cost (USD)	Operating Cost (USD)	TNPC (USD)	LCOE (USD/kWh)	% of Renewables Utilized
298,911	590	-	647,261	14,079.41	875,481.24	0.7318	100

Table 7. Results attained using HOMER software for Scenario 3.

SPV (kWh/Year)	Grid Purchase (kWh/Year)	Batteries Used	Converter (kW)	Initial Capital Cost (USD)	Operating Cost (USD)	TNPC (USD)	LCOE (USD/kWh)	% of Renewables Utilized
126,811	51,155	590	49.4	228,402	2705	275,349	0.587	70.1

5.3.2. Techno-Economic Analysis of Different Scenarios Using Meta-Heuristic Algorithms

This section contains a summary of the optimization results obtained using various meta-heuristic techniques, as well as a comparison of those results for the various scenarios. The MSSA, SSA, and GWO are used in 30 separate runs with a population size of 50 and a maximum iteration of 100, with the best results cropped. The basic parameters used in the optimization problem are tabulated in Table 8. The three considered algorithms are executed for each scenario.

Table 8. Parameters used in MSSA.

Parameters	Value
Population size	50
Maximum iterations	100
d_{min}	0.34
d_{max}	0.79

- Techno–economic analysis for Scenario 1 (solar PV and diesel generator-based EVCS)

Table 9 shows a comparison of results for Scenario 1 using MSSA, SSA, and GWO in terms of LCOE, TNPC, and computational time. According to Table 9, MSSA has the lowest TNPC, LCOE, and computational time with USD 584,566.44, USD 0.4822/kWh, and 27,615.716 s, respectively, due to its lower capital, replacement, and operation and maintenance costs. MSSA takes less time than other compared techniques and gives minimum TNPC and LCOE as well. SSA comes in the second place, with high TNPC, LCOE, and taking more time to compute the results as compared to MSSA, whereas GWO gives a high value of TNPC, LCOE, and is slowest in comparison to the previous two meta-heuristic techniques. This system necessitates 232 SPV panels with 325 W energy, 3284 kWh of energy production per year from a diesel generator rated at 10 kW, 650 battery units, and 9 bi-directional converters rated at 100 kW. The maximum energy flow from the DC bus to the AC bus or vice versa determines the energy of a bi-directional converter.

Table 9. Comparative study of achieved outcomes using meta-heuristic techniques for Scenario 1.

Methods Used	Solar PV (325 W)	DG Power Production (kWh/Year)	Batteries	Converter (100 kW)	TNPC (USD)	LCOE (USD/kWh)	Computational Time (Sec)
MSSA	232	3284	650	9	5,84,566.44	0.4822	27,615.716
SSA	220	3056	580	9	6,04,482.82	0.5381	29,834.165
GWO	190	2859	610	9	6,77,615.38	0.6328	37,656.673

Figure 12 depicts hourly information about power generation for a three-day sample (1–3 October). It is observed that maximum EVs are charged during night hours, and there is no solar output available during that time. Hence, the energy requirement is fulfilled by DG and BESS using energy flow through the converter, as indicated in Figure 12. When PV power generates more than the load demand during the day, surplus electricity is used to charge the battery (Figure 12). Despite having a higher SPV energy, the converter energy is much lower in this context of system sizing. The amount of excess electricity produced in this scenario is 25,053 kWh per year, which is 19.2 % of total generated energy. It is worth noting that, despite the fact that the maximum energy shortage was set to zero, there was only a 23.7 kWh unmet load throughout the simulation model, accounting for 0.02 percent of the total load.

Figure 13 depicts a cost description of various system components for the SPV/DG/battery system. The SPV module and battery entail the most capital investment. Despite the fact that the BESS has a high cost of replacement over the project's lifespan, the resource cost is significantly more than the investment. As a result, this hybrid energy system needs constant economic infusion to keep the system running properly.

As shown in Figure 14, the DG contributes significantly when the load demand is high. The monthly average power production from the SPV module is lesser in summer weather in comparison to winter because of continuous rain and clouds in the summertime. Further, the effectiveness of the proposed algorithm, i.e., MSSA, can be proved using Figure 15. It has better convergence than other algorithms and results in a lower LCOE value.

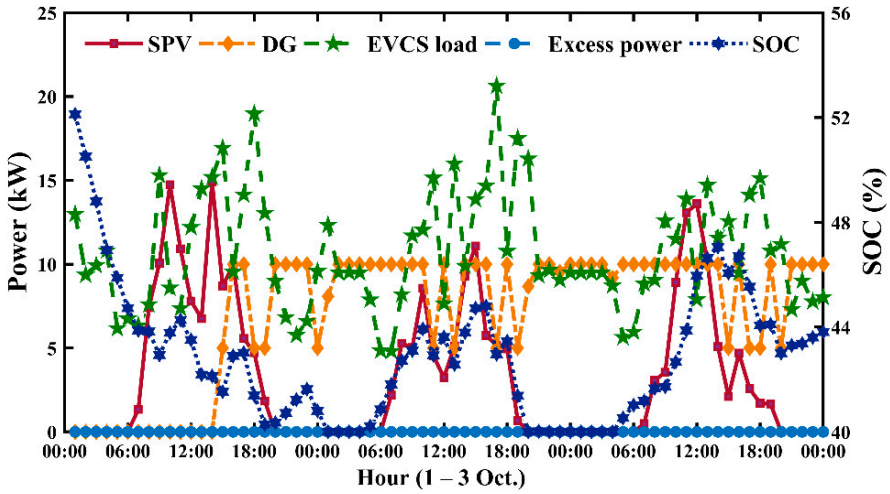


Figure 12. Power analysis on an hourly basis of SPV/DG/battery system.

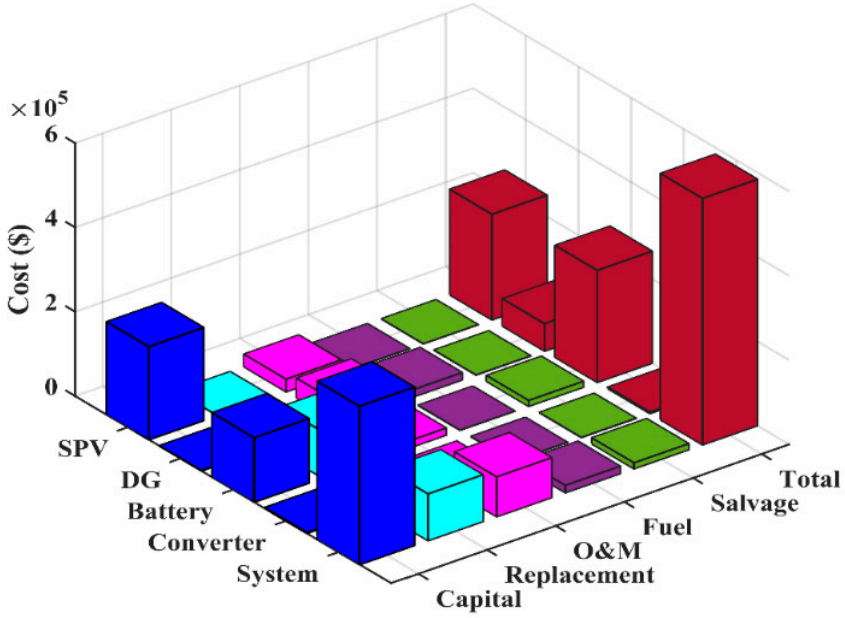


Figure 13. Cost description of the various system components for the SPV/DG/battery system.

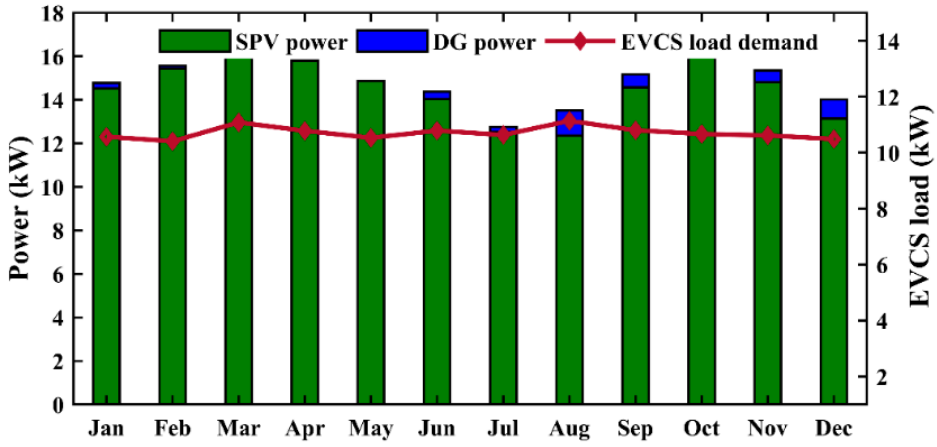


Figure 14. Monthly average power share for satisfying EVCS load demand in SPV/DG/battery system.

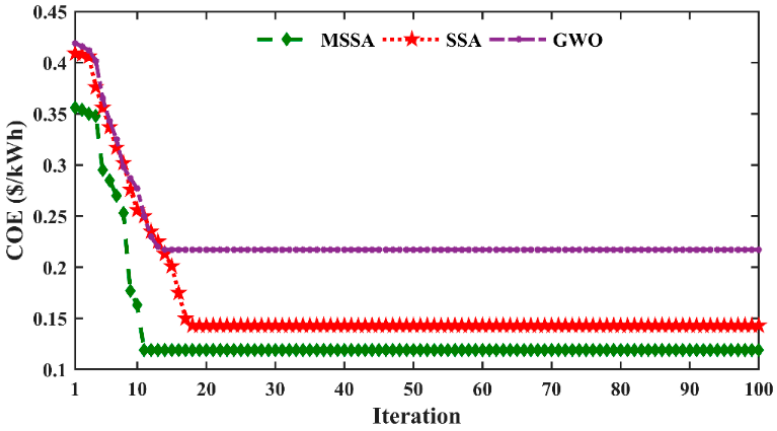


Figure 15. Convergence curves of different optimizers.

- Techno–economic analysis for Scenario 2 (only solar PV/battery-based EVCS)

In Scenario 2, diesel generator is not available, and solar PV is the only energy source available to meet the EVCS load requirement. This PV/battery system has the greatest LCOE and TNPC because it uses a larger PV module and battery storage to meet power demands, which increases the cost and produces more surplus power (194,140 kWh per year). Table 10 provides the different technical and economic parameters optimized by various algorithms for SPV/battery system.

Table 10. Comparative study of achieved outcomes using meta-heuristic techniques for Scenario 2.

Methods Used	Solar PV (325 W)	Batteries	Converter (100 kW)	TNPC (USD)	LCOE (USD/kWh)	Computational Time (Sec)
MSSA	545	548	12	843,461.28	0.6844	25,284.342
SSA	490	518	12	902,761.37	0.7045	29,451.984
GWO	415	476	12	981,736.38	0.7167	35,324.673

As in Scenario 1, computation time taken by MSSA is less as compared to other techniques. This system requires 545 solar PV panels with 325 W energy, 548 battery units, and 12 bi-directional converters rated at 100 kW. MSSA gives better results when equated with other techniques and, hence, proved its efficacy to solve various optimization problems. The hourly information about power generation for a three-day sample for SPV/battery system is depicted in Figure 16. The cost description of the various system components for the SPV/battery system is shown in Figure 17.

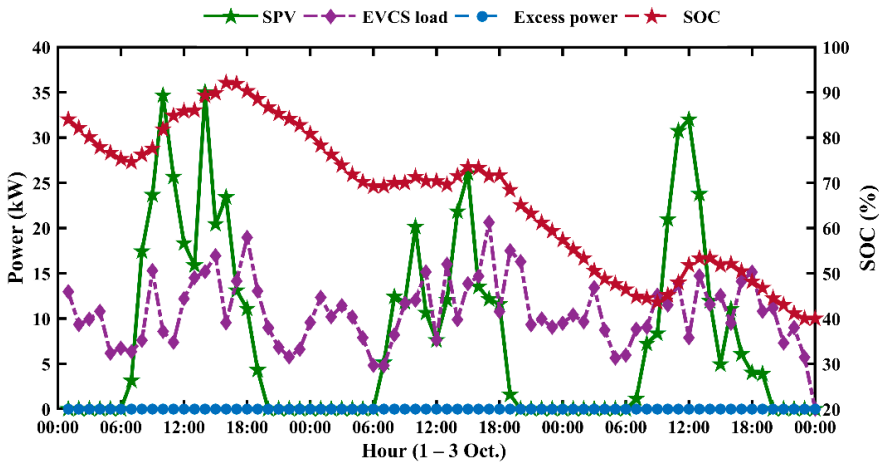


Figure 16. Power analysis on an hourly basis of SPV/battery system.

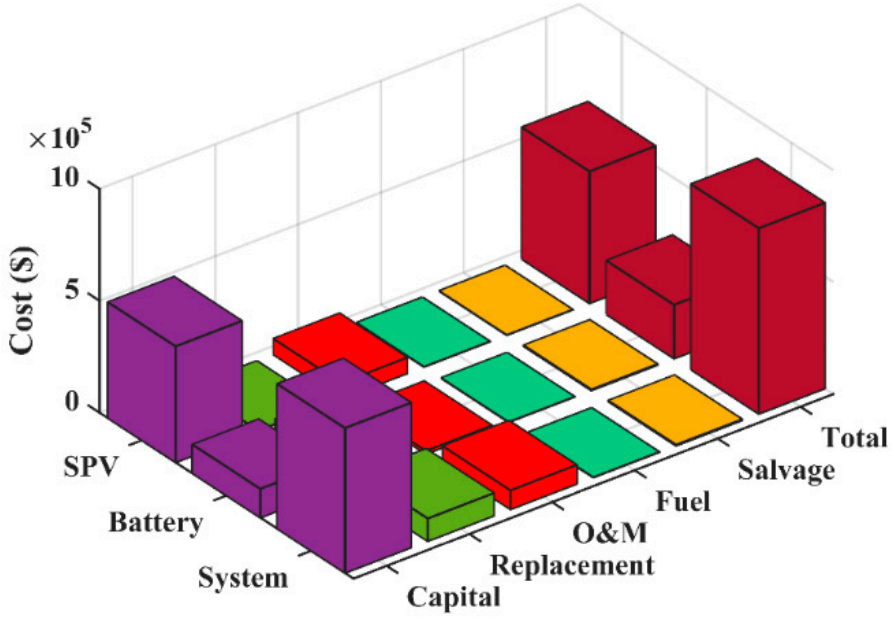


Figure 17. Cost description of the various system components for the SPV/battery system.

- Techno–economic analysis for Scenario 3 (grid-and-solar PV-based EVCS)

This scenario looks at the technological, economic, and environmental concerns of hybrid energy systems connected to the grid. In this strategy, electricity is purchased from the grid to meet power requirements when solar PV is unable to fulfill them. Excess power is sold back to the grid, requiring a small storage capacity while leveraging the monumental amount of additional energy dumped in hybrid energy options. In this respect, a fixed grid power price is chosen to be USD 0.12/kWh and a fixed grid sell back price is selected to be USD 0.08/kWh. According to the results in Table 11, the grid-connected solar PV-based EVCS has a significantly lower LCOE (0.119/kWh) than the solar PV and diesel generator-based EVCS (USD 0.482/kWh) and solar PV/battery-based EVCS (USD 0.684/kWh). The hybrid energy system in Scenarios 1 and 2 requires more resources and battery storage than the grid-connected system. whereas in the grid-connected system, net grid purchase is zero, i.e., the amount of energy purchased from the grid is equal to the amount of energy sold to the grid, and no battery storage is required. Despite having a 232 PV module of 325 W energy and a 49.4 kW bi-directional converter, the grid-connected hybrid system needs no battery storage and has insubstantial energy costs due to a significantly lower LCOE (USD 0.119/kWh) compared to a solar PV and diesel generator-based EVCS (USD 0.482/kWh) and solar PV/battery-based EVCS (USD 0.684/kWh). It is also worth noting that MSSA has a lower TNPC (USD 263377) and LCOE (USD 0.119/kWh) than SSA and GWO. More prominently, the extra electricity in the grid-connected system (408 kWh/year) is considerably lower than in the solar PV and diesel generator-based EVCS (25,053 kWh/year) and solar PV/battery-based EVCS (194,140 kWh/year). This is because a large portion of excess energy is sold back to the grid, and only a few storage devices are needed to meet the requirement when solar power is not available.

Table 11. Comparative study of achieved outcomes using meta-heuristic techniques for Scenario 3.

Methods Used	Solar PV (325 W)	Grid Purchase (kWh/Year)	Grid Sales (kWh/Year)	Converter (49.4 kW)	TNPC (USD)	LCOE (USD/kWh)	Computational Time (Sec)
MSSA	232	51,155	77,135	2	263,377	0.119	29,812.873
SSA	218	52,671	76,432	2	276,543	0.143	30,341.457
GWO	196	54,367	76,197	2	298,654	0.217	32,349.952

When solar power is not accessible, grid power meets the load requirements. According to Figure 18, the spring season necessitates a high amount of energy to be purchased from the grid due to higher power requirements. India, as a tropical country, has nearly equal solar exposure all year. Even though the summer season is supposed to generate more electricity from solar PV, rainy days lead to decreasing SPV production than the rest of the regular time of year.

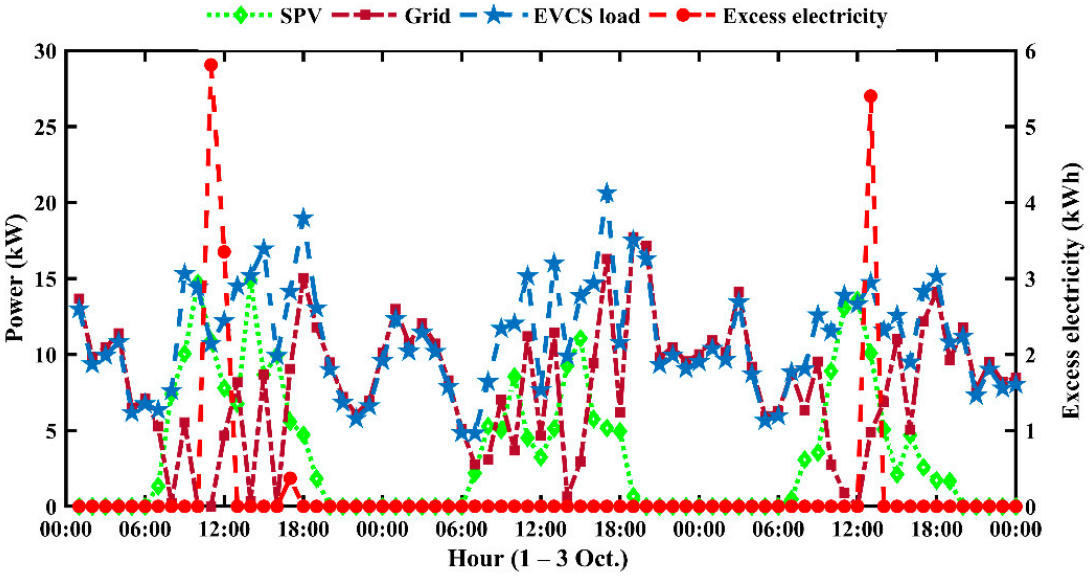


Figure 18. Power analysis on an hourly basis for grid-and-solar PV-based EVCS.

Solar energy accounts for approximately 71.3 percent of total energy demand, with grid power accounting for the remaining 28.7 percent. Figure 19 depicts the time series data for various components used to satisfy the load demand. The cost description of the various system components for the grid-and-solar PV-based EVCS is shown in Figure 20. The monthly average energy purchased/sold from/to the grid is portrayed in Figure 21.

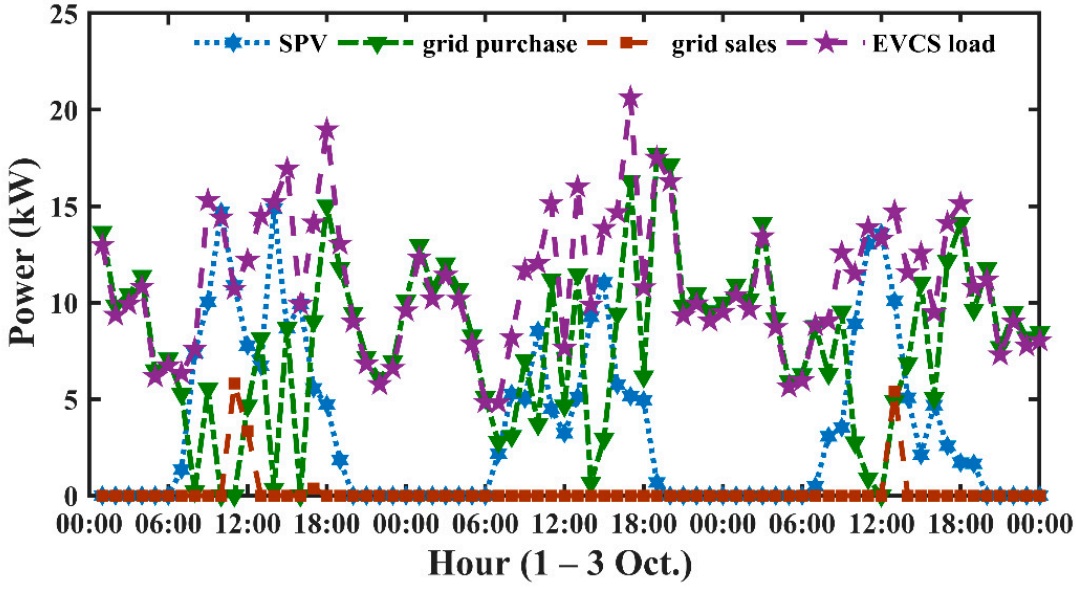


Figure 19. Time series data for various components used to satisfy the load demand.

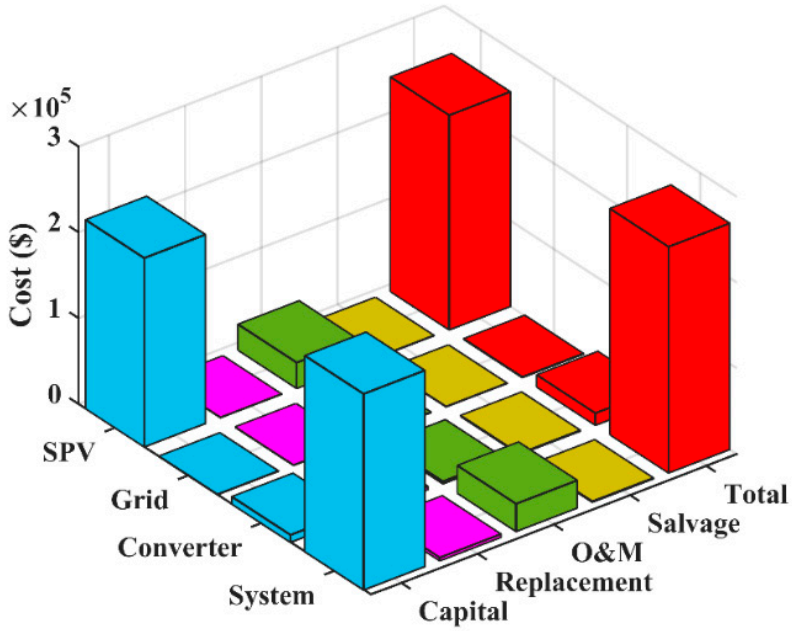


Figure 20. The cost-wise breakup of different components in grid-and-solar PV-based EVCS.

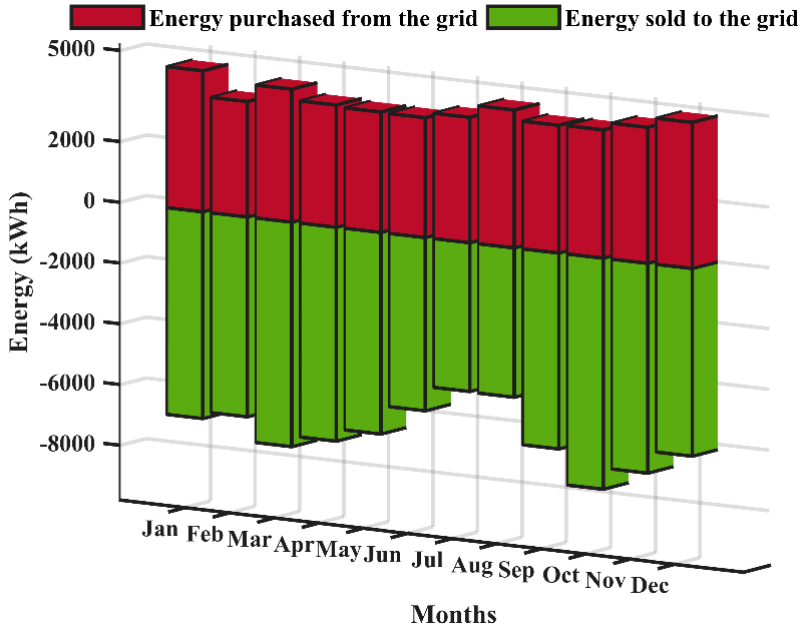


Figure 21. Monthly average energy purchased from the grid and sold to the grid.

- Impact of grid purchase and grid sales on LCOE

The implications of energy buying from grid infrastructure and the quantity of energy being sold to the grid on the LCOE are outlined and discussed in this section. A net grid procurement of 0% indicates that the quantity of energy buying from and sold back to the grid is the same. The value of LCOE increases more with increasing purchases from the grid than with increasing grid sell-back, as shown in Figure 22. According to Figure 22, an increase in grid energy to fulfill load demand leads to an increase in LCOE. For example, a 50% increase in grid energy purchase versus selling back to the grid results in a 36% increase in LCOE.

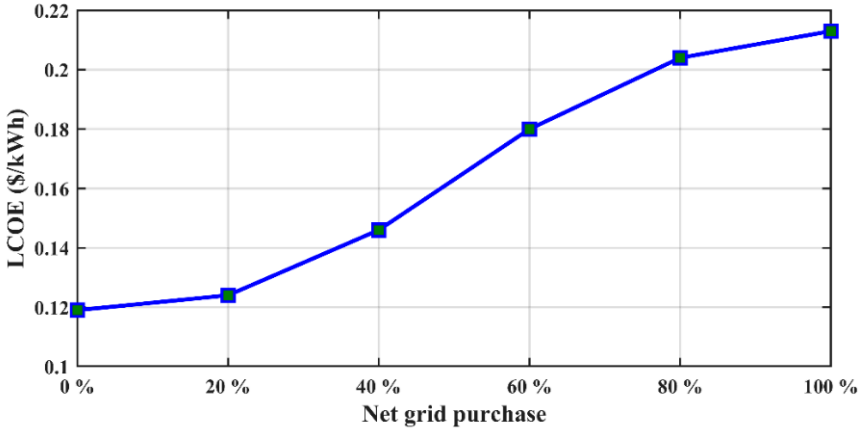


Figure 22. Impact of net grid purchase on LCOE.

5.3.3. Analysis of the Environmental Emissions in Different Scenarios

As previously stated, one of the primary goals of this work is to lower emissions by utilizing renewable technologies. The results displayed in Table 12 show that SPV/battery-based EVCS (Scenario 2) reduces the overall emissions by a significant amount when equated to all other cases. However, while the SPV/DG/battery-based EVCS (Scenario 1) produces higher emissions than SPV/battery-based EVCS, it will be far more ecologically friendly than the grid-and-SPV-based EVCS.

Table 12. Scenario-wise comparison of environmental emissions.

Pollutants	Scenario 1	Scenario 2	Scenario 3
Carbon dioxide (CO_2) (Kg/year)	2910	2417	32330
Carbon monoxide (CO) (Kg/year)	22	0	0
Unburned hydrocarbons (UHC) (Kg/year)	0.802	0	0
Particulate matter (PM) (Kg/year)	1.33	0	0
Sulphur dioxide (SO_2) (Kg/year)	7.14	4.94	140
Nitrogen dioxide (NO_x) (Kg/year)	25	0	68.5

6. Future Challenges and Opportunities

Along with industrial and organizational advancement to promote RE set up in India, identifying the potential obstacles to project implementation is important. The hybrid energy system's infrastructure and maintenance costs have hampered significant advances in RE technologies by increasing the COE. As a developing nation, most of the public are not able to avail high LCOE in comparison to traditional grid purchases. Financial constraints, such as ineffective economic pathways, intense competition with fossil fuels, a scarcity of shareholders, and very few public subsidies, all had a substantial influence on the growth of hybrid energy systems [50]. Furthermore, less economic investment and funding constraints as a result of the low lending charges and longer repayment time make RE investment dangerous. Although economic inflation and growing population necessarily dictate electricity requirements, substituting the current infrastructure with 100 percent renewable use is neither technically nor economically feasible due to ambiguity in energy price in the future [51]. There are still technical intricacies, such as specifications, indicators, processes, instructions, and power management systems for hybrid energy systems. In the meantime, research efforts to overcome such technological challenges are critical and must be taken into consideration when planning a hybrid energy system [52].

Despite their multiple issues, the hybrid energy systems provide lots of opportunities in the area of natural, cultural, financial, and energy security concerns. The hybrid energy system will reduce greenhouse gas emissions and protect the natural environment from deterioration. The steep decline in PV and wind prices will hasten the implementation

of a hybrid energy system with a massive decrease in original investment. The International Renewable Energy Agency (IRENA) forecasted a 74% and a 24% decline in SPV and WT prices by 2020, respectively. Furthermore, continuing research and rapid technological advancements encourage RE systems to start competing with traditional energy sources. Effective hybrid technology selection, taking into account geographical region and environmental consequences, will broaden the existing energy resources and minimize reliance on non-renewable resources. Furthermore, these schemes will have an impact on employment options, the standard of living, business growth, health problems, economic status, women's rights, and economic development. RE is among the most basic needs for human progress, and it can help to generate earnings, improve healthcare and education performance, and poverty eradication. India is geologically favored for the spread of solar energy. Despite the numerous advantages of RE sources, the bulk of people are unwilling to accept the advanced technologies. The lack of knowledge shared among stakeholders is needed for the implementation of the RE-based hybridized micro-grid. In this respect, the government of India can introduce such projects through a public–private partnership framework, in which a group of developers, with support from the government, will participate in putting such projects in place.

7. Conclusions

Hybrid energy systems can deliver energy to the grid and off-grid sites cost effectively and reliably. The efficiency of the system improves when RE components are implemented with the grid. In this article, the complete mathematical modeling is presented for the diverse configurations of the hybrid energy system to meet the load requirements of EVCS situated in the northwest region of Delhi, India. The financial, technological, and ecologic implications of different configurations are also carefully investigated. The use of meta-heuristic methods, which include MSSA, SSA, and GWO, is carried out to optimize the objective functions. The obtained results using MSSA are compared with HOMER software and other techniques. According to comparisons, MSSA gives a more efficient solution than HOMER. Furthermore, the MSSA offers a robust framework, which aids model development. In all three configurations, the MSSA results in reduced LCOE values of USD 0.482/kWh, USD 0.684/kWh, and USD 0.119/kWh, respectively. According to the analysis, the grid-connected solar PV-based EVCS offers significant cost savings over the other two configurations. In addition, a hybrid energy system with equal grid purchase and sell-back can be a profitable choice. More significantly, the LCOE associated with the grid selling back schemes is lower than the grid purchase price. The results show that a 60% increase in grid sell-back price leads to a 46% decrease in LCOE.

Permitting tax-free imports can reduce the cost of hardware devices, lowering the COE. The government of India must enable full cooperation in research and technology by generating opportunities for cooperation and entrepreneurship communication between societies. Even though the hybrid micro-grid alternative is a lucrative, consistent, and long-term remedy, people in rural areas cannot accommodate it due to the increased investment. As a result, more alternative financial assistance from government and non-governmental organizations can address the underlying intricacy of the issue.

The feasibility analysis discussed in this work can be used to guide the development and operation of hybrid energy system applications in distant regions where grid utility is not accessible, as well as hybrid grid-connected systems in India. Future research is needed to investigate the impacts of charging/discharging cycles of a battery on its lifespan and cost of energy.

Author Contributions: Conceptualization, M.B. and I.A.; formal analysis, M.B., I.A., M.A. and F.M.A.; funding acquisition, I.A., M.A. and F.M.A.; investigation, M.B. and M.R.; methodology, M.B., I.A., M.A. and F.M.A.; project administration, M.R. and I.A.; supervision, M.R.; writing—original draft, M.B.; writing—review and editing, M.B., M.R., I.A., M.A. and F.M.A. All authors have read and agreed to the published version of the manuscript.

Funding: The research work is funded by the Chair of Prince Faisal for Artificial intelligence research (CPFAI) through project number QU-CPFAI-2-9-5.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the Chair of Prince Faisal for Artificial intelligence research (CPFAI) for funding this research work through project number QU-CPFAI-2-9-5. Additionally, the authors would like to extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education, and the Deanship of Scientific Research, Qassim University for their support for this research.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

EV	Electric Vehicle
EVCS	Electric Vehicle Charging Station
SPV	Solar Photovoltaic
DG	Diesel Generator
HOMER	Hybrid Optimization of Multiple Energy Resources
GWO	Gray Wolf Optimization
SSA	Salp Swarm Algorithm
MSSA	Modified Salp Swarm Algorithm
RE	Renewable Energy
GHG	Green House Gases
WT	Wind Turbine
NBR	Number of Battery Replacements
RF	Renewable Fraction
GA	Genetic Algorithm
PSO	Particle Swarm Optimization
BESS	Battery Energy Storage System
NPC	Net Present Cost
TNPC	Total Net Present Cost
LCOE	Levelized Cost of Energy
LPSP	Lack of Power Supply Probability
MAS	Multi Agent System
MFO	Moth Flame Optimization
SOC	State of Charge
NOCT	Nominal Operating Cell Temperature
CRF	Capital Recovery Factor
DOD	Depth of Discharge

List of symbols

$P_{SPV\text{-}out}(t)$	Solar photovoltaic power output
η_{SPV}	Efficiency of solar panel
A_{SPV}	Surface area of the solar panel
$G_h(t)$	Hourly solar irradiance falling on the surface of solar panel
T_{cell}	Temperature of cell
T_a	Ambient temperature
P_{max}	Maximum power output of solar panel
V_{max}	Maximum value of voltage
I_{max}	Maximum value of current
N_{SPV}	Number of installed solar PV panel
$u(t)$	Fuel consumption in litres per hour
$P_{DG}(t)$	Power output of diesel generator in kW
$P_{rated\text{-}DG}$	Rated power of diesel generator
a	Constant value (0.246)

b	Constant value (0.08415)
$\eta_{overall}$	Overall efficiency of DG
$\eta_{brake\ thermal}$	Brake thermal efficiency of DG
$SOC_{bat}(t+1)$	State of charge of battery at time instant $(t+1)$
$SOC_{bat}(t)$	State of charge of battery at time instant (t)
σ	Self-discharge rate of battery
η_{bdinv}	Efficiency of bi-directional inverter
η_{bat}	Round trip efficiency of the battery
η_{bat}^{charge}	Charging efficiency of battery
$\eta_{bat}^{discharge}$	Discharging efficiency of battery
C_{total}	Total capacity of battery bank
N_{bat}	Total number of batteries
N_{bat}^{series}	Number of batteries connected in series
C_{sbat}	Capacity of a single battery
NPC_{bat}	Net present cost of battery
C_{bat}^{CAP}	Capital cost of battery
C_{bat}^{REP}	Replacement cost of battery
C_{bat}^{OM}	Operation and maintenance cost of battery
C_{bat}^{SAL}	Salvage cost of battery
ψ_{bat}	Cost of one battery in USD
ω_{bat}	OM cost of battery
μ_{bat}	Escalation rate of battery
ε_{bat}	Resale value of one battery in USD
ζ_{bat}^{rep}	Replacement cost of battery
NPC_{bdinv}	Net present cost of battery
C_{bdinv}^{CAP}	Capital cost of bi-directional inverter
$P_S^{grid}(t)$	Extra energy to be sold to the grid in kWh
$SOC_{bat}^{min}(t)$	Minimum SOC of battery
$SOC_{bat}^{max}(t)$	Maximum SOC of battery
D_{Km}	Number of kilometers traveled
$E_{req/km}$	Energy needed per kilometer
T	Duration of charging
Q_{bat}^{EV}	Capacity of EV battery
N_{SPV}	Number of solar panels
V_{bus}	Voltage level of bus
V_{bat}	Voltage level of single battery
P_{bat}^{max}	Maximum input/output power of battery.
$P_P^{grid}(t)$	Energy to be borrowed from the grid in kWh
$P_D^{EV}(t)$	Energy demand of EV
$P_{EVCS\text{-}dem}(t)$	Power demand of EVCS
NPC_{SPV}	Net present cost of solar panel
C_{SPV}^{CAP}	Capital cost of solar panel
C_{SPV}^{REP}	Replacement cost of solar panel
C_{SPV}^{OM}	Operation and maintenance cost of solar panel
C_{SPV}^{SAL}	Salvage cost of solar panel
ψ_{SPV}	Initial cost and of each SPV panel
P_{SPV}	Power rating of each SPV panel
ζ_{SPV}^{rep}	Cost of replacing the SPV panel
R	Rate of interest
Ω	Life span of project in years
ω_{SPV}	OM cost of each panel
μ_{SPV}	Escalation rate
ε_{SPV}	Resale price solar panel after completing their life
∂	Inflation rate
ε_{bdinv}	Resale value of bi-directional inverter in USD

ζ_{bdinv}^{rep}	Replacement cost of bi-directional inverter
ξ_{S}^{grid}	Cost for unit sale of electricity to the utility grid
ξ_{P}^{grid}	Cost for unit purchase of electricity from the utility grid
C_{P}^{grid}	Cost of purchasing electricity from the grid
C_{S}^{grid}	Cost of selling electricity to the grid
$P_{gen}(t)$	Total power generated by energy components
E_{bat}^{min}	Minimum energy of battery bank
E_{bat}^{max}	Maximum energy of battery bank
P_{def}	Deferrable load

References

1. Available online: https://mnre.gov.in/ (accessed on 2 January 2022).
2. Ashtari, A.; Bibeau, E.; Shahidinejad, S.; Molinski, T. PEV Charging Profile Prediction and Analysis Based on Vehicle Usage Data. *IEEE Trans. Smart Grid* **2012**, *3*, 341–350. [CrossRef]
3. Martinez, C.M.; Hu, X.; Cao, D.; Velenis, E.; Gao, B.; Wellers, M. Energy Management in Plug-in Hybrid Electric Vehicles: Recent Progress and a Connected Vehicles Perspective. *IEEE Trans. Veh. Technol.* **2017**, *66*, 4534–4549. [CrossRef]
4. Wu, X.; Hu, X.; Teng, Y.; Qian, S.; Cheng, R. Optimal integration of a hybrid solar-battery power source into smart home nanogrid with plug-in electric vehicle. *J. Power Sources* **2017**, *363*, 277–283. [CrossRef]
5. Karmaker, A.K.; Ahmed, M.R.; Hossain, M.A.; Sikder, M.M. Feasibility assessment & design of hybrid renewable energy based electric vehicle charging station in Bangladesh. *Sustain. Cities Soc.* **2018**, *39*, 189–202. [CrossRef]
6. Podder, A.K.; Supti, S.A.; Islam, S.; Malvoni, M.; Jayakumar, A.; Deb, S.; Kumar, N.M. Feasibility Assessment of Hybrid Solar Photovoltaic-Biogas Generator Based Charging Station: A Case of Easy Bike and Auto Rickshaw Scenario in a Developing Nation. *Sustainability* **2021**, *14*, 166. [CrossRef]
7. Fares, D.; Fathi, M.; Mekhilef, S. Performance evaluation of metaheuristic techniques for optimal sizing of a stand-alone hybrid PV/wind/battery system. *Appl. Energy* **2022**, *305*, 117823. [CrossRef]
8. Bernal-Agustín, J.L.; Dufo-López, R. Techno-economical optimization of the production of hydrogen from PV-Wind systems connected to the electrical grid. *Renew. Energy* **2010**, *35*, 747–758. [CrossRef]
9. Giallanza, A.; Porretto, M.; Puma, G.L.; Marannano, G. A sizing approach for stand-alone hybrid photovoltaic-wind-battery systems: A Sicilian case study. *J. Clean. Prod.* **2018**, *199*, 817–830. [CrossRef]
10. Halabi, L.M.; Mekhilef, S.; Olatomiwa, L.; Hazelton, J. Performance analysis of hybrid PV/diesel/battery system using HOMER: A case study Sabah, Malaysia. *Energy Convers. Manag.* **2017**, *144*, 322–339. [CrossRef]
11. Lian, J.; Zhang, Y.; Ma, C.; Yang, Y.; Chaima, E. A review on recent sizing methodologies of hybrid renewable energy systems. *Energy Convers. Manag.* **2019**, *199*, 112027. [CrossRef]
12. Al Busaidi, A.S.; Kazem, H.A.; Al-Badi, A.H.; Farooq Khan, M. A review of optimum sizing of hybrid PV–Wind renewable energy systems in oman. *Renew. Sustain. Energy Rev.* **2016**, *53*, 185–193. [CrossRef]
13. Zheng, X.; Lin, H.; Liu, Z.; Li, D.; Llopis-Albert, C.; Zeng, S. Manufacturing Decisions and Government Subsidies for Electric Vehicles in China: A Maximal Social Welfare Perspective. *Sustainability* **2018**, *10*, 672. [CrossRef]
14. Ye, B.; Jiang, J.; Miao, L.; Yang, P.; Li, J.; Shen, B. Feasibility Study of a Solar-Powered Electric Vehicle Charging Station Model. *Energies* **2015**, *8*, 13265–13283. [CrossRef]
15. Ilieva, L.M.; Iliev, S.P. Feasibility assessment of a solar-powered charging station for electric vehicles in the North Central region of Bulgaria. *Renew. Energy Environ. Sustain.* **2016**, *1*, 12. [CrossRef]
16. Filote, C.; Felseghi, R.; Raboaca, M.S.; Aşchilean, I. Environmental impact assessment of green energy systems for power supply of electric vehicle charging station. *Int. J. Energy Res.* **2020**, *44*, 10471–10494. [CrossRef]
17. Niccolai, A.; Bettini, L.; Zich, R. Optimization of electric vehicles charging station deployment by means of evolutionary algorithms. *Int. J. Intell. Syst.* **2021**, *36*, 5359–5383. [CrossRef]
18. Mohamed, A.; Salehi, V.; Ma, T.; Mohammed, O. Real-Time Energy Management Algorithm for Plug-In Hybrid Electric Vehicle Charging Parks Involving Sustainable Energy. *IEEE Trans. Sustain. Energy* **2014**, *5*, 577–586. [CrossRef]
19. Mozafar, M.R.; Moradi, M.H.; Amini, M.H. A simultaneous approach for optimal allocation of renewable energy sources and electric vehicle charging stations in smart grids based on improved GA-PSO algorithm. *Sustain. Cities Soc.* **2017**, *32*, 627–637. [CrossRef]
20. Xu, M.; Meng, Q. Optimal deployment of charging stations considering path deviation and nonlinear elastic demand. *Transp. Res. Part B Methodol.* **2020**, *135*, 120–142. [CrossRef]
21. Bhatti, A.R.; Salam, Z.; Sultana, B.; Rasheed, N.; Awan, A.B.; Sultana, U.; Younas, M. Optimized sizing of photovoltaic grid-connected electric vehicle charging system using particle swarm optimization. *Int. J. Energy Res.* **2019**, *43*, 500–522. [CrossRef]
22. Li, H.; Liu, H.; Ji, A.; Li, F.; Jia, Y. Design of a hybrid solar-wind powered charging station for electric vehicles. In Proceedings of the 2013 International Conference on Materials for Renewable Energy and Environment, Chengdu, China, 19–21 August 2013; IEEE: New York, NY, USA, 2013; pp. 977–981.

23. Hafez, O.; Bhattacharya, K. Optimal design of electric vehicle charging stations considering various energy resources. *Renew. Energy* **2017**, *107*, 576–589. [CrossRef]
24. Bouguerra, S.; Bhar Layeb, S. Determining optimal deployment of electric vehicles charging stations: Case of Tunis City, Tunisia. *Case Stud. Transp. Policy* **2019**, *7*, 628–642. [CrossRef]
25. Chandra Mouli, G.R.; Bauer, P.; Zeman, M. System design for a solar powered electric vehicle charging station for workplaces. *Appl. Energy* **2016**, *168*, 434–443. [CrossRef]
26. Sadeghian, O.; Nazari-Heris, M.; Abapour, M.; Taheri, S.S.; Zare, K. Improving reliability of distribution networks using plug-in electric vehicles and demand response. *J. Mod. Power Syst. Clean Energy* **2019**, *7*, 1189–1199. [CrossRef]
27. Zhang, Y.; Zhang, Q.; Farnoosh, A.; Chen, S.; Li, Y. GIS-Based Multi-Objective Particle Swarm Optimization of charging stations for electric vehicles. *Energy* **2019**, *169*, 844–853. [CrossRef]
28. Ray, P.; Bhattacharjee, C.; Dhenuvakonda, K.R. Swarm intelligence-based energy management of electric vehicle charging station integrated with renewable energy sources. *Int. J. Energy Res.* **2021**. [CrossRef]
29. Miceli, R.; Viola, F. Designing a Sustainable University Recharge Area for Electric Vehicles: Technical and Economic Analysis. *Energies* **2017**, *10*, 1604. [CrossRef]
30. Akbari, M.; Brenna, M.; Longo, M. Optimal Locating of Electric Vehicle Charging Stations by Application of Genetic Algorithm. *Sustainability* **2018**, *10*, 1076. [CrossRef]
31. Dai, Q.; Liu, J.; Wei, Q. Optimal Photovoltaic/Battery Energy Storage/Electric Vehicle Charging Station Design Based on Multi-Agent Particle Swarm Optimization Algorithm. *Sustainability* **2019**, *11*, 1973. [CrossRef]
32. Pan, A.; Zhao, T.; Yu, H.; Zhang, Y. Deploying Public Charging Stations for Electric Taxis: A Charging Demand Simulation Embedded Approach. *IEEE Access* **2019**, *7*, 17412–17424. [CrossRef]
33. Ekren, O.; Hakan Canbaz, C.; Güvel, Ç.B. Sizing of a solar-wind hybrid electric vehicle charging station by using HOMER software. *J. Clean. Prod.* **2021**, *279*, 123615. [CrossRef]
34. Ahmad, F.; Khalid, M.; Panigrahi, B.K. An enhanced approach to optimally place the solar powered electric vehicle charging station in distribution network. *J. Energy Storage* **2021**, *42*, 103090. [CrossRef]
35. Singh, S.; Chauhan, P.; Jap Singh, N. Feasibility of Grid-connected Solar-wind Hybrid System with Electric Vehicle Charging Station. *J. Mod. Power Syst. Clean Energy* **2021**, *9*, 295–306. [CrossRef]
36. Aldhanhani, T.; Al-Durra, A.; El-Saadany, E.F. Optimal design of electric vehicle charging stations integrated with renewable DG. In Proceedings of the 2017 IEEE Innovative Smart Grid Technologies-Asia (ISGT-Asia), Auckland, New Zealand, 4–7 December 2017; IEEE: New York, NY, USA, 2017; pp. 1–6.
37. Al Wahedi, A.; Bicer, Y. Techno-economic assessment of a renewable energy-based electric vehicle fast-charging station in Qatar. *Comput. Aided Chem. Eng.* **2021**, *50*, 1629–1634.
38. Abdelkader, A.; Rabeh, A.; Mohamed Ali, D.; Mohamed, J. Multi-objective genetic algorithm based sizing optimization of a stand-alone wind/PV power supply system with enhanced battery/supercapacitor hybrid energy storage. *Energy* **2018**, *163*, 351–363. [CrossRef]
39. Xu, X.; Hu, W.; Cao, D.; Huang, Q.; Chen, C.; Chen, Z. Optimized sizing of a standalone PV-wind-hydropower station with pumped-storage installation hybrid energy system. *Renew. Energy* **2020**, *147*, 1418–1431. [CrossRef]
40. Yan, D.; Ma, C. Stochastic planning of electric vehicle charging station integrated with photovoltaic and battery systems. *IET Gener. Transm. Distrib.* **2020**, *14*, 4217–4224. [CrossRef]
41. Bandopadhyay, J.; Roy, P.K. Application of hybrid multi-objective moth flame optimization technique for optimal performance of hybrid micro-grid system. *Appl. Soft Comput.* **2020**, *95*, 106487. [CrossRef]
42. Das, B.K.; Hassan, R.; Tushar, M.S.H.K.; Zaman, F.; Hasan, M.; Das, P. Techno-economic and environmental assessment of a hybrid renewable energy system using multi-objective genetic algorithm: A case study for remote Island in Bangladesh. *Energy Convers. Manag.* **2021**, *230*, 113823. [CrossRef]
43. Nishanthy, J.; Charles Raja, S.; Praveen, T.; Jeslin Drusila Nesamalar, J.; Venkatesh, P. Techno-economic analysis of a hybrid solar wind electric vehicle charging station in highway roads. *Int. J. Energy Res.* **2022**. [CrossRef]
44. Emad, D.; El-Hameed, M.A.; El-Fergany, A.A. Optimal techno-economic design of hybrid PV/wind system comprising battery energy storage: Case study for a remote area. *Energy Convers. Manag.* **2021**, *249*, 114847. [CrossRef]
45. Geleta, D.K.; Manshahia, M.S.; Vasant, P.; Banik, A. Grey wolf optimizer for optimal sizing of hybrid wind and solar renewable energy system. *Comput. Intell.* **2020**, *158*, 941–951. [CrossRef]
46. Borhanazad, H.; Mekhilef, S.; Gounder Ganapathy, V.; Modiri-Delshad, M.; Mirtaheri, A. Optimization of micro-grid system using MOPSO. *Renew. Energy* **2014**, *71*, 295–306. [CrossRef]
47. Kumar, A.; Rizwan, M.; Nangia, U. A new approach to design and optimize sizing of hybrid microgrid in deregulated electricity environment. *CSEE J. Power Energy Syst.* **2020**, 1–11. [CrossRef]
48. Mirjalili, S.; Gandomi, A.H.; Mirjalili, S.Z.; Saremi, S.; Faris, H.; Mirjalili, S.M. Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems. *Adv. Eng. Softw.* **2017**, *114*, 163–191. [CrossRef]
49. Khan, N.H.; Wang, Y.; Tian, D.; Jamal, R.; Kamel, S.; Ebeed, M. Optimal Siting and Sizing of SSSC Using Modified Salp Swarm Algorithm Considering Optimal Reactive Power Dispatch Problem. *IEEE Access* **2021**, *9*, 49249–49266. [CrossRef]
50. Nanaki, E.A.; Xydis, G.A. Deployment of Renewable Energy Systems: Barriers, Challenges, and Opportunities. In *Advances in Renewable Energies and Power Technologies*; Elsevier: Amsterdam, The Netherlands, 2018; pp. 207–229.

51. Sen, S.; Ganguly, S. Opportunities, barriers and issues with renewable energy development—A discussion. *Renew. Sustain. Energy Rev.* **2017**, *69*, 1170–1181. [CrossRef]
52. Dominković, D.F.; Bačeković, I.; Pedersen, A.S.; Krajačić, G. The future of transportation in sustainable energy systems: Opportunities and barriers in a clean energy transition. *Renew. Sustain. Energy Rev.* **2018**, *82*, 1823–1838. [CrossRef]

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 307-321
Published Online November 2021 in IJEAST (http://www.ijeast.com)

TRAFFIC VOLUME STUDY AND CONGESTION SOLUTIONS

Nishant Singh,
Department of Civil Engineering
Delhi Technological University

Nitin Chauhan
Department of Civil Engineering
Delhi Technological University of Civil Engineering,

Nishant Pande
Department of Civil Engineering
Delhi Technological University

Narad Muni Prasad
Associate Professor, Department
Delhi Technological University of Civil Engineering,

***Abstract*-Traffic Engineering is used to accomplish secure and time bound commutation of persons and freights on road network which in turn relies on traffic flow. Due to an increase in the standard of living of middle-class families in the last decade, traffic volume has considerably increased. Thus, lack of proper management of roads can result in non-sustainable development. This work analyses traffic properties of Shivaji Marg, Janakpuri district centre, Delhi, India by performing direct manual counting method of survey. Objective of this study is to record traffic volume of the site for the weekdays from 8AM to 2PM, determine the morning peak hour and the morning peak hour factor using the standard passenger car unit (PCU) values. Also, this study attempts to determine the practical PCU values for different automobiles running on the site and calculate the corresponding morning peak hour factor. Lastly, this work comments on the congestion level of the site by comparing the recorded hourly traffic volume with the permissible volume per hour. The study found out that the practical PCU values differ from their corresponding standard values as it depends on the speed and space occupied by the automobiles which can vary from site to site. Also, it was found that the hourly traffic volume is within the permissible limit implying no congestion in the current scenario. There are, however, certain limitations of this work like absence of advanced technology, variation in the PCU values, time constraints. Nevertheless, this study will help the stakeholders in the construction sector to prioritise their resources towards factors which need to be improved to enhance the serviceability of the highway for present as well as for the future generation.**

Keywords: **Traffic Volume, Standard Passenger Car Unit, Practical Passenger Car Unit, Peak hour factor (PHF).**

I. INTRODUCTION

Since time immemorial, transport system has been sustaining several societies. However, it has also been one of the most burning problems in every corner of the world. Each and every nation has been responding to the situation as per their resources and capabilities. Whenever we design a structure, we calculate the load acting on it to find the amount of reinforcements which is to be provided for its safety. Just like that, volume of traffic is measured to find the capacity of a road network. Volume of traffic is necessary for the planning, styling, secure functioning, and its constant evolution. Traffic engineering is that part of engineering that employs engineering techniques to realize the secure and structured commutation of individuals and freights on road network.

Transportation volume research are specifically performed to work out the count, movement, and classification of road automobiles at a specific point to become conscious of peak time flow period, impact of heavy vehicles or people who walk on foot on vehicle traffic flow. In case of a non-homogeneous traffic flow, it has been a drag to work out the traffic volume as vehicles have different velocities and body measurements. The above problem has been rectified by the use of a quantity known as passenger car unit (PCU). Since volume of traffic isn't steady, endless assessment of traffic volume is of utmost importance otherwise, the transport system may fail, and therefore the financial condition of state may have certain problems.

Some studies have already been conducted on traffic volume analysis of different types of roads. Halder and Alam [2] have conducted the traffic volume analysis at AUST-Flyover junction to Shatrasta junction in Dhaka, Bangladesh. They conducted direct manual counting method and drew the flow fluctuation curve showing number of vehicles at different times in a day. They also suggested few measures to reduce the congestion and improve the efficiency of the highway. Venkatcharyulu and Mallikarjunareddy [1] conducted the traffic volume survey from Miyapur road to Nizampet crossroad, a semi urban road in Hyderabad. They determined the vehicular composition and directional distribution of different vehicles at different times using manual counting method. Yaseen and Swamy [5] conducted traffic volume survey using manual method at Mothi circle, Davangere,

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 307-321
Published Online November 2021 in IJEAST (http://www.ijeast.com)

Karnataka where they proposed for a signal time optimization for better and efficient flow of the traffic. Singh and Goyal [7] conducted the manual counting and determined the level of serviceability of the road within Punjab University campus. They determined the traffic flow pattern on weekends and weekdays for hourly and daily variations. Objectives of this study are:

- To conduct a survey of volume of traffic of Shivaji Marg, Janakpuri district centre, Delhi using direct manual counting method.
- To determine the practical passenger car unit (PCU) value for different class of automobiles running on the site and then determine the morning peak hour factor (PHF).
- To compare the practical PCU (for the particular site selected) and corresponding PHF values with the standard PCU and corresponding PHF values respectively.
- To comment on the congestion level of the site and provide some measures, if necessary.

This paper is organised into 5 sections. Section 2 deals with the different methods of counting the number of vehicles, different parameters used to determine the average annual daily traffic and peak hour factor. Section 3 presents the methodology adopted in the study. Section 4 gives the result and discussions. Lastly, section 5 concludes the study.

II. TRAFFIC SURVEY

Information on traffic is a prerequisite for the highway engineers in order to design the road network and also to manage the traffic system. In order to design various traffic facilities, select the standards of geometry, and assess from the economic angle, the knowledge about traffic is required. They make use of such information to install control devices like traffic signs, traffic signals, marking of pavements, and also to assess the effectiveness of various programs and schemes. Such an information about the volume of traffic must be regularly updated as it is an ever-changing system. There are certain ways to perform the survey of volume of traffic which have been discussed in the following section.

2.1 Methods of Volume Survey

a) Manual Counting Method: - in the considered category, counting of automobiles is performed using one's hands. It is further categorised into 2 methods: -

- **Direct Method: -** A hand tally and devices like manual counters and enumerators are used in such a method. The reason why such a method is considered is that such a data can be used immediately after the conduction of the survey and also vehicle classification can be done easily. But such a method cannot be practicably used for a survey of long duration and also when the flow of traffic is quite high as there is a fear of error like missing the vehicles or forgetting the count or double counting the same automobile. Also, such a method cannot be used in oppressive weather conditions. [2]
- **Indirect Method: -** A video camera is used for recording the footage in such a method. Here, data is collected later by

replaying the footage of the recorded video. Apart from volume of traffic, we can also find various other parameters. Whenever the flow of automobiles is large and the operation is non-lane based, this method can be handy. One of the constraints is that for better recording of the footage, a raised place or platform is needed. Other thing is that the data cannot be used immediately after conducting the survey. It has to be transcripted from the video first. Moreover, the capacity of the film of the recorder is finite, so it is not applicable for a long time period of counting of vehicles. Furthermore, the sky has to be clear, and sun needs to be shining brightly for better quality of the video recording. [2]

b) Automatic counting method: - In such a method, counting of automobiles is performed using devices which do not require any interference by human beings. Such a method is further categorised into two broad classifications, first being the one in which contact takes place between the device and vehicles which includes pneumatic tubes and the other being the one in which no contact is required which includes ultrasound or infrared radar. Such a method can easily be used in cases of long time period count. Moreover, it does not need many personnel and is free from any error and also it does not get affected by oppressive weather conditions. Some of the constraints are that it cannot classify the automobiles into various categories and sometimes it may happen that two vehicles cross the point at same time, but the machine counts them as one. Thus, it can be deduced that the accuracy of this method is not as high as the previous methods. Moreover, the cost of installation of the machines is quite high. [2]

a) 2.2 Peak hour factor (PHF)

It is defined as the average volume during the peak 60 minutes period divided by 4 times the average volume during the peak 15 minutes period. It is descriptive of trip generation patterns and may apply to an area or portion of a street and highway system. One can also use 5,10- or 20-minutes interval for the calculation of PHF. However, in that case, the multiplying factor in the denominator would change. [10]

$$PHF = V_{av}^{60} / (4* V_{av}^{15}) \quad \text{--------(1)}$$

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 307-321
Published Online November 2021 in IJEAST (http://www.ijeast.com)

III. METHODOLOGY

Direct manual counting method has been selected to conduct the survey. Traffic volume survey was conducted at the mid-section of the road for the time interval 8AM-2PM, for five days in a week starting from Monday (25/10/2021) to Friday (29/10/2021). A map of the site has been depicted below this section. Data was collected for every 15 minutes in order to calculate the morning peak hour factor for every day. The count of different class of vehicles was then multiplied by standard passenger car unit values to get the total number of vehicles crossing the section. The corresponding standard morning peak hour factor was determined for every day using equation (1) and then averaged to get the mean value.

Now, since PCU depends upon the speed and space occupied by automobiles, the practical PCU value may differ from road to road. Thus, in order to calculate the practical PCU value for different vehicles, a formula suggested in Chandra and Kumar (2003) was used.

$$PCU = (\frac{Vc}{Vi})/(\frac{Ac}{Ai}) \qquad \text{----------(2)}$$

where, Vc - speed of car
V_i - speed of i^{th} vehicle
A_c - projected rectangular area of a car
A_i - projected rectangular area of i^{th} vehicle
V_c/V_i - speed ratio of the car to the i^{th} vehicle
A_c/A_i - space ratio of the car to the i^{th} vehicle

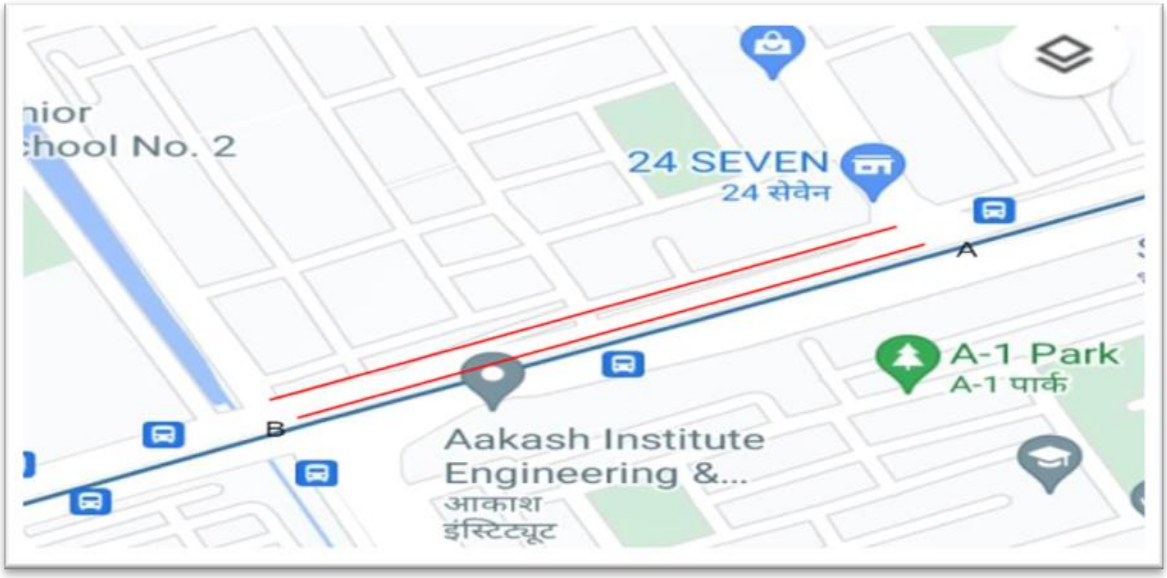
Projected rectangular area was calculated by measuring the length and breadth of the vehicles and then multiplying them together. Speed of the vehicle was determined by dividing distance by time. Distance of the stretch was found to be 500 metres. Time to traverse this distance was found using stopwatch and video recording. Speed of a number of automobiles were calculated and then averaged to get the mean speed of a particular class of automobile. Once the practical PCU value was calculated, it was used to calculate the corresponding practical morning peak hour factor for every day of the survey.



Figure 1: Different views of site taken for Traffic study

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 307-321
Published Online November 2021 in IJEAST (http://www.ijeast.com)

Figure 2: Map View of Site taken for Traffic Study



IV. OBSERVATIONS AND CALCULATIONS

Table 1(a). Speed of vehicles on Monday

Monday

Vehicle Type	Number of vehicles considered	Average Speed in kmph
2-Wheelers	60	33.7
3-Wheelers	25	28.5
Car	60	37.78
Bus/Truck	10	30.25
LCV	20	29.45

Table 1(b). Speed of vehicles on Tuesday

Tuesday

Vehicle Type	Number of vehicles considered	Average Speed in kmph
2-Wheelers	60	32.9
3-Wheelers	25	29.15
Car	60	37.1
Bus/Truck	10	29.55
LCV	20	30.5

Table 1(c). Speed of vehicles on Wednesday

Wednesday

Vehicle Type	Number of vehicles considered	Average Speed in kmph
2-Wheelers	60	34.4

Table 1(d). Speed of vehicles on Thursday

Thursday

Vehicle Type	Number of vehicles considered	Average Speed in kmph
2-Wheelers	60	33.55

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 307-321
Published Online November 2021 in IJEAST (http://www.ijeast.com)

3-Wheelers	25	29.05
Car	60	37.45
Bus/Truck	10	31.95
LCV	20	30

3-Wheelers	25	29.6
Car	60	38.95
Bus/Truck	10	31.5
LCV	20	32.05

Table 1(e). Speed of vehicles on Friday **Friday**

Vehicle Type	**Number of vehicles considered**	**Average Speed in kmph**
2-Wheelers	60	34.05
3-Wheelers	25	28
Car	60	36.55
Bus/Truck	10	32.5
LCV	20	30.92

Average Speed of every class of vehicle

2-Wheeler = (33.7+ 32.9 +34.4 + 33.55+ 34.05)/5
= 33.72

Similarly average speed for each class is calculated

Table 2. Average Speed of vehicles considering all the five days of survey

Vehicle Type	**Mean Average Speed in kmph**
2-Wheelers	33.72
3-Wheelers	28.86
Car	37.57
Bus/Truck	31.15
LCV	30.58

Table 3. Projected area for each class of vehicles

Vehicle Type	**Dimensions (meter)**		**Projected Area (meter squares)**
	Length	**Width**	
2-Wheelers	2.28	0.98	2.2344
3-Wheelers	2.58	1.32	3.4056

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 307-321
Published Online November 2021 in IJEAST (http://www.ijeast.com)

Car	4.5	1.8	8.1
Bus/Truck	9.5	2.45	23.275
LCV	4.85	1.9	9.215

PCU value for each class of vehicle for the given segment (Practical PCU)

PCU= (Speed of Car/Speed Of vehicle class considered) / Area of Car/Area of vehicle class considered)

2-wheelers PCU = (37.57/33.72) / (8.1/2.23)
= 0.31

Table 4. Practical PCU

Vehicle Type	**PCU Value**
2-Wheelers	0.31
3-Wheelers	0.55
Car	1.00
Bus/Truck	3.47
LCV	1.40

Table 5. Standard PCU as per IRC

Vehicle	**PCU Value**
2-Wheeler	0.4
3-Wheeler	0.5
Car	1
Bus/Truck	2.2
LCV	1

Standard PCU values considered for study of any Road are:-

Table6(a). Vehicular composition for Day 1 –Monday (25/10/2021)

Time	**2-Wheelers**	**3-Wheelers**	**Car**	**Bus/Truck**	**Pick-up (LCV)**
8- 8:15	102	12	49	5	8
8:15- 8:30	93	11	52	5	10
8:30- 8:45	98	13	68	7	6

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 307-321
Published Online November 2021 in IJEAST (http://www.ijeast.com)

8:45- 9:00	109	16	55	9	3
9:00- 9:15	**115**	**19**	**71**	**13**	**4**
9:15- 9:30	**111**	**18**	**87**	**17**	**6**
9:30- 9:45	**133**	**21**	**109**	**14**	**8**
9:45- 10:00	**121**	**24**	**96**	**21**	**10**
10:00- 10:15	107	22	89	19	12
10:15- 10:30	115	26	78	10	15
10:30- 10:45	120	19	87	11	7
10:45- 11:00	118	23	80	8	8
11:00- 11:15	95	22	75	12	15
11:15- 11:30	103	19	79	14	3
11:30- 11:45	111	15	81	11	4
11:45- 12:00	88	11	75	10	6
12:00- 12:15	92	13	80	17	8
12:15- 12:30	100	17	81	13	2
12:30- 12:45	99	21	78	8	4
12:45- 1:00	94	13	79	11	5
1:00- 1:15	90	9	60	11	16
1:15- 1:30	86	13	58	10	8
1:30- 1:45	82	10	52	17	6
1:45- 2:00	85	15	55	13	9

Table 6(b). Hourly Traffic as per Practical PCU

Monday

Vehicle Type	8- 9 AM	9- 10 AM	10- 11 AM	11- 12 PM	12- 1 PM	1- 2 PM
2-Wheelers	402	480	460	397	385	343
3-Wheelers	52	82	90	67	64	47
Car	224	363	334	310	318	225
Bus/Truck	26	65	48	47	49	51
Pick-up	27	28	25	28	19	39
Total	**731**	**1018**	**957**	**849**	**835**	**705**
*Total Volume (total*PCU)*	**505.24**	**821.65**	**727.66**	**672.21**	**669.18**	**588.75**

Table 6(c). Hourly Traffic as per standard PCU

Vehicle Type	8- 9 AM	9- 10 AM	10- 11 AM	11- 12 PM	12- 1 PM	1- 2 PM
2-Wheelers	402	480	460	397	385	343
3-Wheelers	52	82	90	67	64	47
Car	224	363	334	310	318	225
Bus/Truck	26	65	48	47	49	51
Pick-up	27	28	25	28	19	39
Total	**731**	**1018**	**957**	**849**	**835**	**705**
*Total Volume (total*PCU)*	**495**	**767**	**693.6**	**633.7**	**630.8**	**536.9**

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 307-321
Published Online November 2021 in IJEAST (http://www.ijeast.com)

Table 6(d). Peak Hour traffic

Vehicle Type	9:00-9:15 AM	9:15-9:30 AM	9:30-9:45 AM	9:45-10:00 AM
2-Wheelers	115	111	133	121
3-Wheelers	19	18	21	24
Car	71	87	109	96
Bus/Truck	13	17	14	21
Pick-up	4	6	8	10

Figure 3: Pie chart representing vehicular composition percentage for Day 1(Monday)

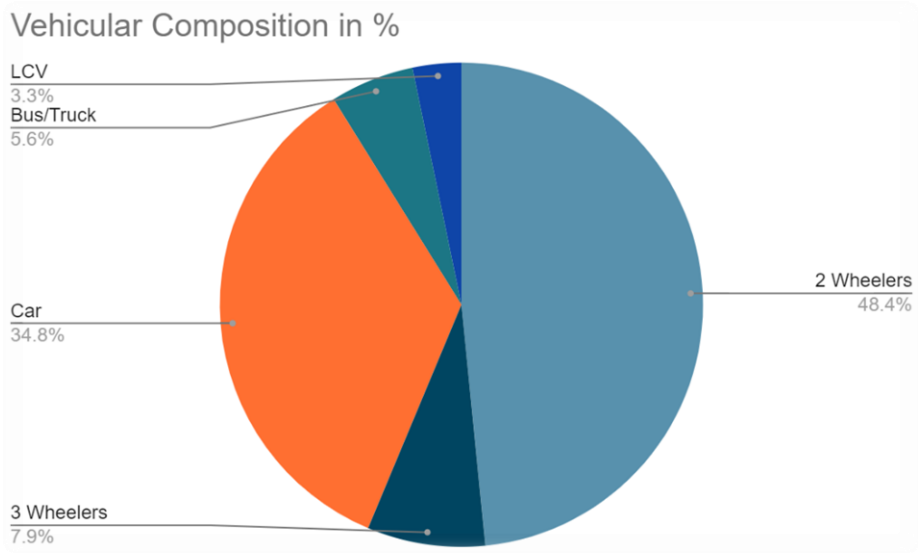


Table 7(a). Day 2- Tuesday (26/10/2021)

Vehicle Type	8- 9 AM	9- 10 AM	10- 11 AM	11- 12 PM	12- 1 PM	1- 2 PM
2-Wheelers	412	492	466	451	395	350
3-Wheelers	45	57	57	63	62	50
Car	212	363	334	310	318	225
Bus/Truck	17	31	50	34	39	34
Pick-up	27	29	25	34	22	27
Total	**713**	**972**	**932**	**892**	**836**	**686**
Total Volume (Practical PCU)	**461.26**	**695.04**	**718.31**	**650.04**	**640.68**	**516.78**
Total Volume (Standard PCU)	**463.7**	**685.5**	**683.9**	**630.7**	**614.8**	**491.8**

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 307-321
Published Online November 2021 in IJEAST (http://www.ijeast.com)

Table 7(b). Peak Hour traffic as per standard PCU

Vehicle Type	9:00 – 9:15 AM	9:15 – 9:30 AM	9:30 – 9:45 AM	9:45 – 10:00 AM
2-Wheelers	120	123	117	132
3-Wheelers	14	14	14	15
Car	91	89	86	97
Bus/Truck	8	8	7	8
Pick-up	7	7	7	8

Table 7(c). Peak Hour Traffic as per practical PCU

Vehicle Type	10:00 – 10:15 AM	10:15 – 10:30 AM	10:30 – 10:45 AM	10:45 – 11:00 AM
2-Wheelers	112	106	118	130
3-Wheelers	14	17	15	11
Car	81	86	84	83
Bus/Truck	12	16	11	11
Pick-up	5	6	6	8

Figure 4: Pie chart representing vehicular composition percentage for Day 2(Tuesday)

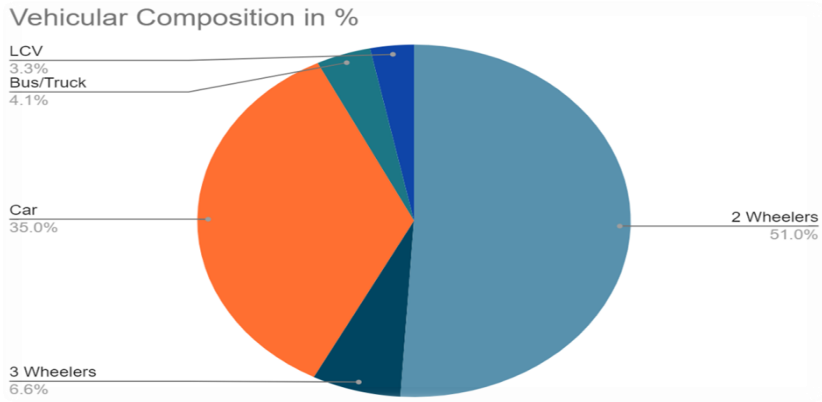


Table 8(a). Day 3 – Wednesday (27/10/2021)

Vehicle Type	8- 9 AM	9- 10 AM	10- 11 AM	11- 12 PM	12- 1 PM	1- 2 PM
2-Wheelers	320	356	345	325	300	305
3-Wheelers	31	57	85	75	65	52
Car	200	232	190	201	170	155
Bus/Truck	18	39	55	26	30	41
Pick-up	21	24	37	42	28	19

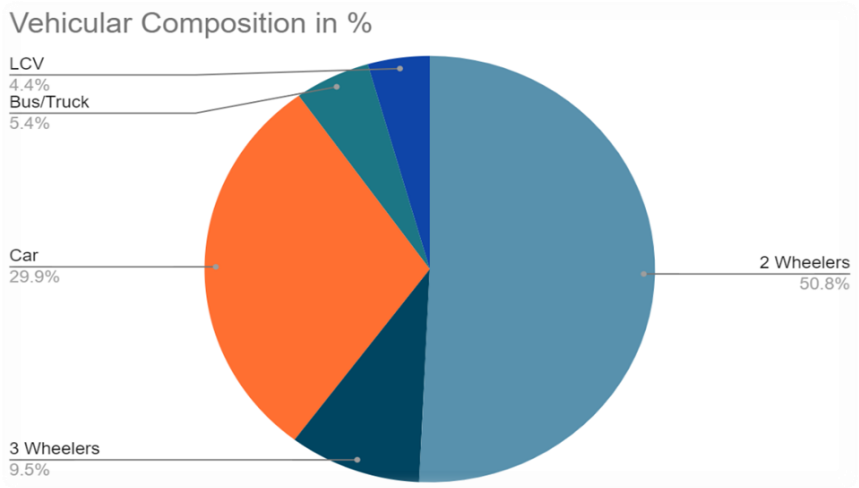
International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 307-321
Published Online November 2021 in IJEAST (http://www.ijeast.com)

Total	**590**	**708**	**712**	**669**	**593**	**572**
Total Volume (Practical PCU)	**408.11**	**542.64**	**586.35**	**492.02**	**442.05**	**447.02**
Total Volume (Standard PCU)	**404.1**	**512.7**	**514.9**	**467.7**	**416.5**	**412.2**

Table 8(b). Peak hour traffic

Vehicle Type	**10:00 – 10:15 AM**	**10:15 – 10:30 AM**	**10:30 – 10:45 AM**	**10:45 – 11:00 AM**
2-Wheelers	85	86	82	92
3-Wheelers	21	19	23	22
Car	68	69	65	72
Bus/Truck	14	16	13	12
Pick-up	6	6	6	6

Figure 5: Pie chart representing vehicular composition percentage for Day 3(Wednesday)



International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 307-321
Published Online November 2021 in IJEAST (http://www.ijeast.com)

Table 9(a). Day 4 – Thursday (28/10/2021)

Vehicle Type	**8- 9 AM**	**9- 10 AM**	**10- 11 AM**	**11- 12 PM**	**12- 1 PM**	**1- 2 PM**
2-Wheelers	322	391	360	312	308	297
3-Wheelers	36	50	65	71	62	47
Car	165	208	185	185	195	132
Bus/Truck	12	27	22	17	23	27
Pick-up	18	26	40	48	20	11
Total	**553**	**702**	**672**	**633**	**608**	**514**
Total Volume (Practical PCU)	**351.46**	**486.8**	**464.69**	**446.96**	**432.39**	**359.01**
Total Volume (Standard PCU)	**356.2**	**474.8**	**449.9**	**430.7**	**419.8**	**344.7**

Table 9(b). Peak hour Traffic

Vehicle Type	**9:00 – 9:15 AM**	**9:15 – 9:30 AM**	**9:30 – 9:45 AM**	**9:45 – 10:00 AM**
2-Wheelers	95	98	93	105
3-Wheelers	13	12	12	13
Car	52	51	50	56
Bus/Truck	5	9	6	7
Pick-up	7	6	6	7

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 307-321
Published Online November 2021 in IJEAST (http://www.ijeast.com)

Figure 6: Pie chart representing vehicular composition percentage for Day 4(Thursday)

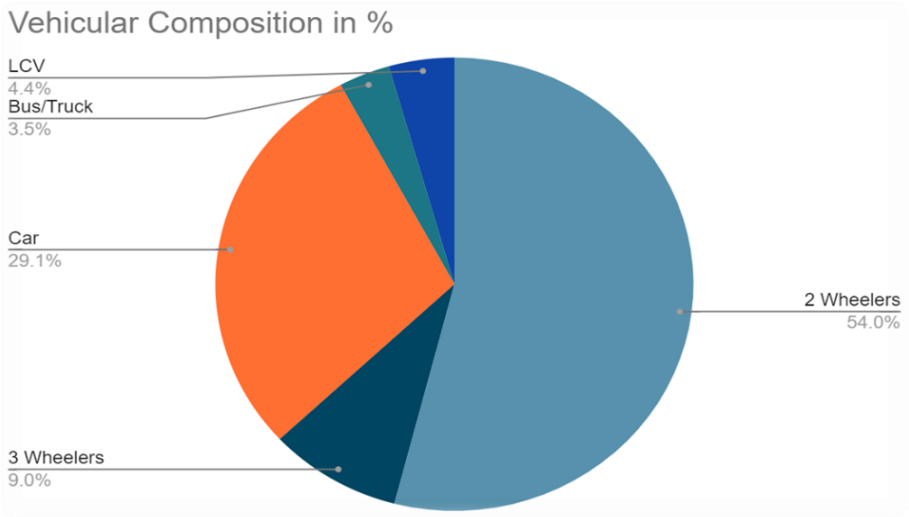


Table 10(a). Day 5 – Friday (29/10/2021)

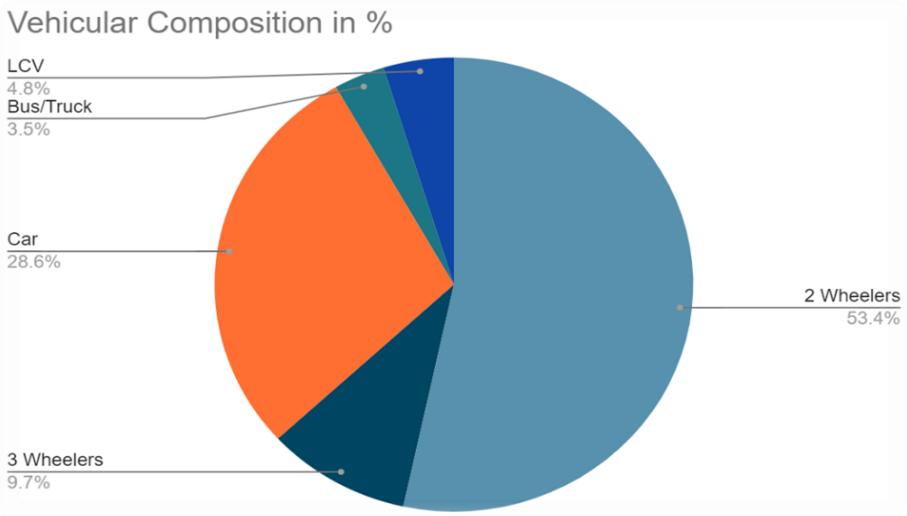
Vehicle Type	**8- 9 AM**	**9- 10 AM**	**10- 11 AM**	**11- 12 PM**	**12- 1 PM**	**1- 2 PM**
2-Wheelers	304	365	340	308	295	306
3-Wheelers	21	50	81	68	75	55
Car	170	185	165	186	171	152
Bus/Truck	12	30	18	19	25	20
Pick-up	18	31	55	33	19	15
Total	**525**	**661**	**659**	**614**	**585**	**548**
Total Volume (Practical PCU)	**342.63**	**473.15**	**454.41**	**431.01**	**417.05**	**367.51**
Total Volume (Standard PCU)	**346.5**	**453**	**436.1**	**418**	**400.5**	**360.9**

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 307-321
Published Online November 2021 in IJEAST (http://www.ijeast.com)

Table 10(b). Peak hour Traffic

Vehicle Type	**9:00 – 9:15 AM**	**9:15 – 9:30 AM**	**9:30 – 9:45 AM**	**9:45 – 10:00 AM**
2-Wheelers	89	91	87	98
3-Wheelers	13	12	12	13
Car	46	45	44	50
Bus/Truck	9	6	5	10
Pick-up	8	8	7	8

Figure 7: Pie chart representing vehicular composition percentage for Day 5(Friday)



Calculations (Shown for Monday)

Standard PCU calculations (9-10 AM)

2 wheelers = 480*0.4 = 192
3 wheelers = 82*0.5 = 41
Car = 363*1 = 363
Bus/Truck = 65*2.2 = 143
Pick-up= 28*1 = 28
Total Volume = 192+41+363+143+28
= 767 per hour

PHF Calculation

PHF for Monday (9-10am) = 767/ (4*212) [using equation 1]
= 0.902

Practical PCU Calculation (9-10 AM)

2 wheelers= 480*0.31= 148.8
3 wheelers= 82*0.55= 45.1
Cars= 363*1= 363

International Journal of Engineering Applied Sciences and Teclogy, 2021
Vol. 6, Issue 7, ISSN No. 2455-2143, Pages 307-321
Published Online November 2021 in IJEAST (http://www.ijeast.com)

Bus/Truck= 65*3.47= 225.55
pick-up= 28*1.4= 39.2

Total Volume = 148.8+45.1+363+225.55+39.2=821.65

PHF Calculation
PHF for Monday (9-10am) = 821.65/(4*233.58)
0.879

Similarly:-

Table 11(a). Standard morning PHF

Day	**PHF (Standard)**
Monday	0.902
Tuesday	0.937
Wednesday	0.858
Thursday	0.936
Friday	0.935
Average	**0.913**

Table 11(b). Practical morning PHF

Day	**PHF (Practical)**
Monday	0.879
Tuesday	0.928
Wednesday	0.862
Thursday	0.937
Friday	0.886
Average	**0.898**

Table 12. Permissible Capacity in PCU per hour for different traffic conditions

No. of Traffic Lanes and Width	**Traffic Flow**	**Capacity in PCU per hour for traffic condition**	
		Roads with no frontage access, no standing vehicles, little cross traffic	**Roads with frontage access, no standing vehicles and high-capacity intersection**
Two lane (7.0- 7.5m)	One way	2400	**1500**
Two lane (7.0- 7.5m)	Two way	1500	1200
Three lane (10.5m)	One way	3600	2500
Four lane (14.0m)	One way	4800	3000
Four lane (14.0m)	Two way	4000	2500
Six lane (21.0m)	Two way	6000	4200

(Referred from "Highway Engineering" book by Justo and Khanna)

Table 1(a) to 1(e) depict the calculation of speed of different automobiles on different weekdays. In table 2, speed in different weekdays has been averaged to get a mean speed. Table 3 shows the dimensions and the corresponding projected area of the automobiles. Table 4 presents the practical PCU values calculated using Chandra and Kumar method. Table 5 represents the standard PCU value of an automobile.

Table 6(a) gives a detailed description of the first day of the survey. Data for an interval of every 15 minutes has been shown. In table 6(b), the data has been multiplied by the practical PCU values to get the total PCU per hour. Similarly, in table 6(c), the data has been multiplied by standard PCU values. Table 6(d) depicts the peak hour traffic for Monday. Similarly, tables from 7 to 10 represent the total PCU per hour for the remaining days of the survey.

Table 11(a) and 11(b) present the standard and practical morning PHF calculated using (1).
Table 12 has been referred from Justo and Khanna (2011) which depicts the maximum PCU per hour for different traffic conditions. The condition under which our site falls has been kept in bold.
The pie charts depict the proportion of different classes of vehicles on each day of the survey. It represents the dependency of people on different classes of vehicles.

V. CONCLUSION

In this study, traffic volume survey was conducted for Shivaji Marg, Janakpuri west, Delhi for five continuous days of a week from 25/10/2021 to 29/10/2021 from 8AM to 2PM. The morning peak hour factor was also calculated for every day of the survey and then averaged to get the mean value.
First of all, coming to the PCU values, it can be observed that the practical PCU value of 3-wheelers, bus/truck and LCV have come out to be greater than the standard value. For 2-wheelers, the practical value is less than the standard. Thus, we can say that the PCU value cannot be generalised (standardised) as it depends on a number of factors like condition of road, type of road, nature of traffic and even on the time chosen for conducting the survey.
Moving further, as per the practical PCU, the morning peak hour for three days of the survey was 9AM-10AM which is correct as per IRC as people tend to move out to go to their offices or schools or any other place of work. However, for two days, i.e., Tuesday and Wednesday, morning peak hour was 10AM-11AM . This might be because of the varying timetable of the various coaching institutes set up in the area. As per the standard PCU, morning peak hour was 9-10AM for all days except Wednesday when it was 10-11AM. As time passes, volume of traffic tends to decrease gently which is also correct as per IRC. Slight variation in the result is expected due to some difference in the PCU values.
The site selected is a one-way two-lane urban road and its permissible volume as per table 12 is 1500 PCU per hour. As per the practical PCU, the maximum PCU per hour is on Monday from 9-10AM (821.65 PCU/hour) and as per the standard PCU also, it is on Monday from 9-10AM (767 PCU/hour). Thus, it can be inferred that the site is well below the level of congestion.
Average morning PHF was found to be 0.898 and 0.913 corresponding to practical and standard PCU respectively . A higher PHF value indicates that the traffic volume in every 15-minute interval is almost the same and thus consistent throughout the hour. This results in less variable traffic flows and free flows of traffic.
From the pie-charts, it can be concluded that people are heavily dependent on two-wheelers, followed by four-wheelers. Public transport system like bus is less favoured by the commoners especially because of the pandemic situation. As a result, people move out in their private vehicles which could result in heavy traffic in coming few years because with the increase in standard of living, people tend to buy private vehicles to show their stature in the society. As a result, the road transport authorities should be vigilante to regularly assess the traffic volume of the site and undertake necessary improvements for the betterment of the efficiency of the road and its level of service.

VI. REFERENCES

[1]. Venkatcharyulu S., Mallikarjunreddy V., "Traffic volume analysis of newly developing semi-urban road", E3S Web of Conferences 184 01116 (2020), pp. 01-08. DOI: https://doi.org/10.1051/e3sconf/202018401116
[2]. Halder S., Bin Alam M.R., Islam M.A., Hassan M., "Traffic volume study", ResearchGate, https://www.researchgate.net/publication/271833441 (accessed June 20, 2021).
[3]. Kumar Rai R., Rinku, Prasad N., "Traffic analysis and capacity evaluation", International Journal of Engineering Applied Sciences and Technology, vol. 5, no. 1, pp. 359-362, 2020. URL: https://www.ijeast.com
[4]. Haq Bhat A.U., Gupta D.R., "A review paper on study of traffic volume and its safety measures on national highways", International Research Journal of Engineering and Technology, vol. 5, no. 7, pp. 1346-1348, 2018. URL: https://www.academia.edu
[5]. Yaseen M., Swamy L.V., Kirankumar H.S., "Traffic volume count and signal optimisation at Mothi Circle Davanagere", International Journal of Engineering Research and Technology, vol. 4, no. 9, pp. 533-538, 2015. DOI: http://dx.doi.org/10.17577/IJERTV4IS090695
[6]. Kumar G., Priyarakshitha K., "Traffic congestion study with a reality approach- A review", International Journal of Pure and Applied Mathematics, vol. 119, no. 17, pp. 1061-1069, 2018. URL: http://www.acadpubl.eu/hub/
[7]. Singh B., Goyal D., "Study of traffic volume and level of service of Punjab University, Chandigarh", Int. Journal of Engineering Research and Applications, vol. 5, no. 7, pp. 09-14, 2015. URL: https://www.ijera.com
[8]. Singh R., Tekwani H., Joshi B., Sudhakaran P., Singh J., "Study of traffic volume and its safety measurement at Dadabari Circle, Career Point University, Kota", International Journal of Current Engineering and Technology, vol. 8, no. 1, pp. 72-74, 2018. URL: http://inpressco.com/category/ijcet
[9]. Khanna S., Justo C., "Traffic Engineering", in Highway Engineering, Nem Chand and Bros, 2011, pp. 159-267.
[10]. Mathew D., "Measurement at a Point", NPTEL, https://nptel.ac.in/content/storage2/courses/105101008/downloads/cete_05.pdf (accessed Oct. 23,2021).
[11]. Shokeen N., Gopal V., Bhave V., Prasad N., "Design of traffic signal using webster method", International



Journal of Scientific Engineering and Applied Sciences, vol. 7, no. 5, pp. 177-196, 2021. URL: https://www.ijseas.com

Transductive Instance Transfer Learning for Cross-Language Defect Prediction

Rahul Chopra
Department of Software Engineering, Delhi Technological University, New Delhi, India
rahul.dilli02@gmail.com

Shreoshi Roy
Department of Software Engineering, Delhi Technological University, New Delhi, India
shreoshiroy99@gmail.com

Ruchika Malhotra
Department of Software Engineering, Delhi Technological University, New Delhi, India
ruchikamalhotra@dtu.ac.in

ABSTRACT

Predicting defects (bugs) is critical to increasing software quality. Many software defect prediction algorithms have been presented, and many of them have shown to be effective in practice. However, because existing works are largely limited to a single project, their effectiveness in predicting cross-project defects is usually poor. This is primarily due to the issue of class imbalance and discrepancies in feature distribution between the source and destination projects. However, because of the disparities in distribution amongst datasets from different studies, developing high-quality Cross Project Defect Prediction (CPDP) models remains a difficulty. In our study, instead of collecting data from a single project, we have collected source code from multiple code submissions on a programming contest website and employed Natural Language Processing (NLP) models to detect software defects in them.

CCS CONCEPTS

• Software and its engineering; **• Software creation and management**; **• Software verification and validation**; **• Software defect analysis**; **• Software testing and debugging**;

KEYWORDS

Defect Prediction, Transfer Learning, Natural Language Processing, Artificial Neural Network, Long Short Term Memory, Doc2Vec Embedding

ACM Reference Format:
Rahul Chopra, Shreoshi Roy, and Ruchika Malhotra. 2022. Transductive Instance Transfer Learning for Cross-Language Defect Prediction. In *2022 4th Asia Pacific Information Technology Conference (APIT 2022), January 14–16, 2022, Virtual Event, Thailand.* ACM, New York, NY, USA, **??** pages. https://doi.org/10.1145/3512353.3512379

1 INTRODUCTION

Defect prediction is an essential tool for assessing software quality and lowering development, testing and maintenance expenses. A defect is a coding error or bug that results in inaccurate or unexpected outputs from a program that is not in accordance with what it is supposed to perform. If a defect is identified early in a software development lifecycle, it saves us monumental costs than when it is identified at a later stage.

There have already been many studies for minimizing software defects in the past, but most of them are based on software metrics derived from the source code. This method works well for a specific project, but does not generalize for cross project defect prediction. This is because of change in the feature space and distribution or lack of data for a new project. These variations from one project to another hinder the performance of transfer learning using these techniques. Most statistical models must be rebuilt from scratch when the distribution changes, utilizing newly gathered training data. Retrieving the necessary training data and rebuilding the models is expensive or unfeasible in many real-world situations. Reducing the need for the effort required to obtain training data would be ideal. Knowledge transfer or transfer learning between task domains would be advantageous in such instances.

The goal of this research is to build a cross programming language defect prediction tool which is being trained on one programming language and can be used to find defects or errors in other languages. The primary development of this research will be based on using NLP models such as Long Short Term Memory (LSTM), Artificial Neuron Network (ANN) etc. along with the combination of transfer learning techniques. These models are used to study the structure of code to identify the rules followed by the Java programming language. This learning can be transferred to code written in another language (C++) using transfer learning because most programming languages follow a similar structure.

Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task. There are many different types of transfer learning but we will focus on taking help from transductive instance transfer learning for this research as it suits well to our research problem. We train our labeled training data to predict the labels of the testing dataset whose labels are not observed during training. In general, the most commonly utilized method is model-based deep transfer learning. However, there has been very little study focusing on increasing deep transfer learning by focusing on the impact of data. We offer an instance-based strategy for improving deep transfer learning in a target domain in this study. In particular, we select a pre-trained model from a source domain (defect prediction in Java) and use it to predict the impact of training samples in a target domain (defect prediction in C++).

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.
APIT 2022, January 14–16, 2022, Virtual Event, Thailand
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9557-1/22/01…$15.00
https://doi.org/10.1145/3512353.3512379

2 BACKGROUND AND RELATED WORK

2.1 Defect prediction

CPDP is important for projects with little or no data to create predictors, according to Zimmermann et al. (2009). Only 3.4 percent of the 622 CPDPs they attempted were successful. The features of the data as well as the procedure are critical to the CPDP's performance. CPDP has been regarded as a difficult problem that more researchers should look into [1].

Turhan et al. (2009) proposed employing nearest-neighbor filtering to implement a CPDP method. They claimed that a defect classifier created with Within Project (WP) data outperformed those created with Cross Project (CP) data. They suggested a two-phase method based on the findings.

Companies are encouraged to use CPDP and begin collecting WP data in phase one. Companies should stop using CP data once enough WP data has been collected and instead use predictors learned from WP data as phase two. They emphasized the importance of WP data in boosting prediction performance.

2.2 Transfer learning

The feature distribution between training and test data is assumed to be the same in machine learning algorithms [2]. The performance of prediction models will be poor if this assumption fails. Transfer learning approaches solve this challenge by transferring knowledge gained from a source project into models for the target project.

Given adequate labeled source data and insufficient labeled target data, some ways have been offered to address the problem setting.

TrAdaBoost was proposed by Dai et al. (2007), which uses boosting to improve knowledge transfer. The target and source data are looked at separately. For the target data, AdaBoost (Freund and Schapire 1997) is used to raise the weight of misclassified target cases. To reduce the weight of misclassified source samples, the weighted majority approach is used.

TrAdaBoost, according to Shi et al. (2008), performs poorly when the source data is unimportant. Yao and Doretto (2010) presented the Multi STrAda Boost algorithm to handle this problem, which transfers information from numerous sources to locate a single source that is highly connected to the goal.

Methods that increase prediction performance using WP data, such as those used in these studies, could be integrated into CPDP.

3 MOTIVATION

3.1 Why explore defect prediction?

Many lessons about personnel practices, testing methodologies, language choices, and so on might be transferred between projects. While all of these topics are significant and deserving of investigation, this section explains why we concentrate on fault prediction.

Human programmers are intelligent, but they have flaws. Coding adds features, but it also introduces some bugs. As a result, software may occasionally crash (perhaps at the most inconvenient or dangerous time) or provide incorrect functionality.

Programming naturally produces faults, it's critical to test programs before deploying them. But testing is costly. Software evaluation resources are limited, but evaluation effectiveness rises in lockstep with assessment effort. For example, a linear increase in the confidence C of identifying problems in black-box testing methods can require exponentially more effort. Exponential costs quickly deplete finite resources, hence it is usual practice to use the best available algorithms on the most crucial code parts. However, any strategy that concentrates on specific parts of the code can make us oblivious to flaws in other areas. To investigate the rest of the system, a simple sample policy should be utilized. This sampling strategy will always be insufficient. Nonetheless, when resources preclude a comprehensive review of everything, it is the only alternative. Defect predictors learned from software measurements such as static code attributes are one example of a lightweight sampling policy. Data miners, for example, can determine where the probability of software problems is largest given static code descriptors for each module and a count of the number of issues identified during inspect (or at runtime). The remainder of this section argues that defect predictors of this type are simple to use, commonly used, and useful.

Easy to use: Even for very large systems, many software metrics such as static code attributes and process metrics can be routinely collected from software repositories. Manual code reviews, for example, are significantly slower and more labor-intensive. For example, depending on the review procedures used, 8 to 20 LOC can be inspected per minute, and this effort is repeated for all members of the review team, which can be as large as four or six individuals.

Widely used: Software metrics are used by researchers and industry practitioners to inform software quality forecasts. Large industrial businesses such as Google, Microsoft, AT&T, and Samsung have acknowledged using defect prediction models.

3.2 Why explore transfer learning?

The study of the nature of generality in SE is one motivation to investigate transfer learning. When professional organizations publish lists of ostensibly generic "best practices," they believe that such generalities exist. For example, the IEEE 1012 software verification standard proposes a number of ways for evaluating software quality [3].

Endres and Rombach list hundreds of software engineering lessons, including McCabe's Law (functions with a "cyclomatic complexity" larger than 10 are more mistake prone) [4] [5]. Many other well-known researchers, such as Jones and Glass, do the same, citing Brooks' Law as an example (adding programmers to a late project makes it later).

Budgen and Kitchenham, in general, want to reorganize SE research by drawing general findings from a larger number of investigations [6].

4 EXPERIMENTAL DESIGN AND SETUP

For the purpose of our study, we intend to employ NLP for finding software defects. NLP is commonly used for information extraction from naturally occurring languages like textual material or transcribed speech. It aims to study the rules and structures present in natural languages to find patterns such as which words are nouns, verbs, adjectives etc. It does so by converting words into numerical data which can be understood by the computer. Similarly, NLP can also be used for finding patterns in programming languages, because similar to natural languages, programming languages are

Table 1: Characteristics of data

Dataset	No. of Instances	Defective	Non Defective
Java	17600	8720	8880
C++	4440	2220	2220

also written in a structured format and are bound to rules, in fact more than what natural languages are. When this structure is not followed, an error or defect occurs which is what we intend to detect.

4.1 Data collection

The first step of our study is data collection. Since we are using NLP models, we need data in text format. For data collection, we scrape data from a programming contest site - Atcoder. We are using Beautiful Soup for web scraping the submissions on this website. For Java as well as CPP, we scrape accepted submissions to represent defect-free code and submissions with compilation error as defective code that contains some syntactical error. We are using the HTML parser for parsing the data and saving the code as text files. Details about the class distribution of the data scraped are mentioned in below table 1.

4.2 Data preprocessing

After data extraction, we pre-process our data and remove all the single line comments and multi-line comments as they are not part of the code structure and should be omitted. We do not need to perform stop-word removal or stemming as all keywords in the code are equally important and no changes to the base form are required. We combine all the lines of code of one file into one array as every file has one label of 0 or 1 representing defective or non-defective code respectively.

5 PROPOSED METHODOLOGY

In our proposed methodology, we take the code file as an input, pre-process the data inside it, convert it into multi-dimensional real-valued vectors in a predefined vector space using Doc2Vec embedding, Train the data on 3 deep learning models - ANN, basic LSTM and stacked LSTM using codes written in Java and finally test these models on C++ code using transfer learning. In all our machine learning models, we have split 80% of the data in Java as a training set and 20% as a testing set.

Figure 1 depicts a step-by-step process of our methodology. The following sections provide a full description of each stage.

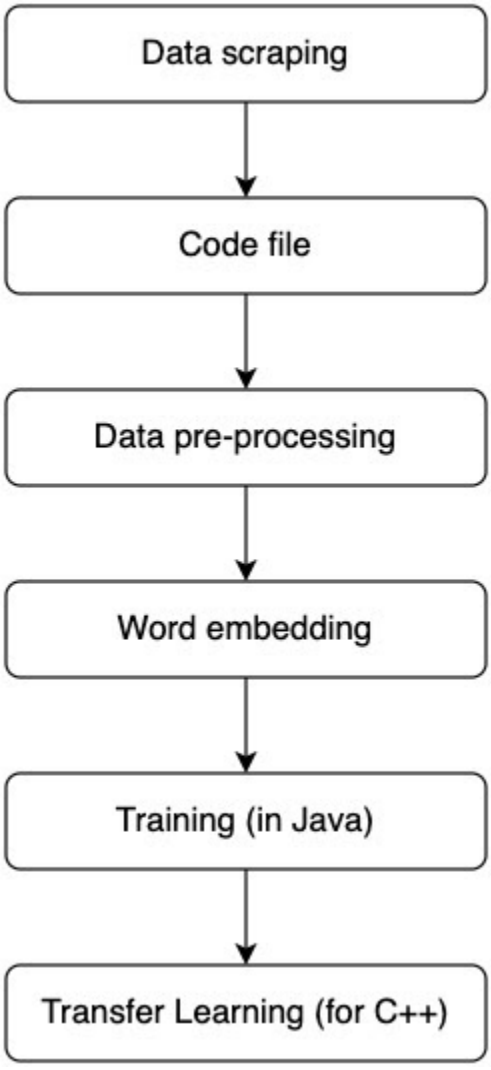


Figure 1: Proposed methodology

5.1 Doc2Vec embedding

Le and Mikolov's [7] distributed representation of sentences and documents, commonly known as Doc2Vec embedding, is an extension of the word2vec embedding. Doc2Vec converts the corpus into paragraph vectors instead of word vectors.

If the corpus has N paragraphs and the vocabulary contains M words, and we wish to learn paragraph vectors in which each paragraph is mapped to p dimensions and each word is mapped to q dimensions, the model will have $N \times p + M \times q$ parameters (excluding the softmax parameters). When N is big, the number of parameters can be huge, but the updates during training are usually sparse and hence efficient.

In the Doc2Vec model, each paragraph is used to make features and hence the order of occurrence of words is considered. Every paragraph is converted into a vector containing user-defined dimensions and these dimensions form the features that can be used to represent that paragraph. These features can directly be used as inputs in any conventional machine learning model.

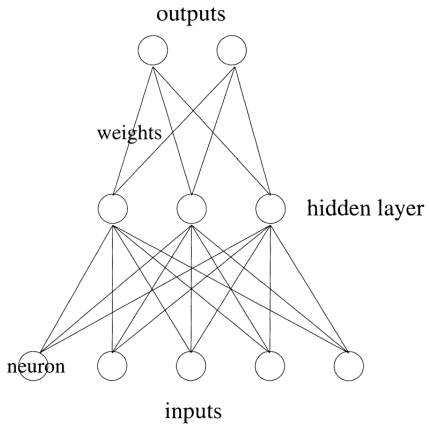


Figure 2: 1907 ANN Architecture. Picture by Wang, Sun-Chong, as cited in . (https://doi.org/10.1007/978-1-4615-0377-4_5.

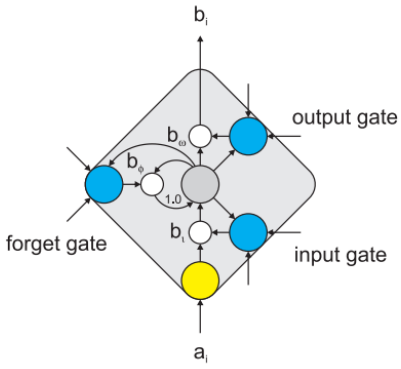


Figure 3: 1907 Architecture of an LSTM memory cell. Picture by Sundermeyer, Martin, Ralf Schlüter, and Hermann Ney, as cited in . (http://doi.org/10.21437/Interspeech.2012-65.

In our research, we have used the Doc2Vec model from the Gensim library and trained it on the Java dataset. We have defined the vector_size as 100 which means that each paragraph will be represented as a combination of 100 features. We have ignored all paragraphs with total frequency lower than 2 and trained the model for 30 epochs.

5.2 ANN with dropout

Artificial Neural Network (ANN) as described in [8] is a collection of neurons or nodes stacked in layers. The nodes accept data as input and pass it through a function whose output is passed on to the next layer of neurons. Each node's output is referred to as its activation or node value. Figure 2 shows the architecture of the ANN model [8]. Each link in the network has a weight, which is a numerical value. This weight is usually adjusted as learning progresses and the goal of ANN model is to determine these weights as accurately as possible.

The first layer of ANN is called the visible layer. Subsequently there could be many layers that are called hidden layers and finally there is the output layer. The input data is passed through a function (usually ReLU or Tanh) and is fed into the input layer of ANN. The data is then multiplied with the weights of each node. Initially the weights are generated randomly. The weighted sum of the input is converted into an output based on the activation function and then sent to the next layer. This process continues through all layers of ANN. This is called forward propagation. The predicted output is compared with the actual output and based on the bias, back propagation is carried out. This process continues till the bias is minimized.

The hidden layer's output, h_i is denoted by:

$$h_i = \sigma\left(\sum_{j=1}^{N} V_{ij}x_j + T_i^{hid}\right)$$

where $\sigma()$ is the activation function, N is the number of input neurons, V_{ij} is the weights, x_j inputs to the input neurons, and T_i^{hid} is the hidden neurons' threshold terms.

ANN is a very powerful model and requires a huge dataset. Since our dataset is in the range of thousands, we have followed the technique mentioned in [9] to avoid overfitting of our model. According to this paper, adding a dropout layer not only prevents overfitting but also helps in efficiently combining different architectures of ANN. 'Dropout' refers to dropping out nodes from an ANN temporarily, along with its incoming and outgoing edges.

We have designed our model to contain 4 ANN layers and 3 dropout layers (one between every two ANN layers). The first layer contains 50 neurons to accept 100 dimensions and keeps reducing to half in the consequent layers. These layers have their activation function as ReLU. The last output layer contains 1 node and has the sigmoid activation function as we need our output as either 0 or 1. We have used the Adam optimizer for optimization and binary-cross entropy for the loss function. We trained this model on a batch size of 32 for 3 epochs.

5.3 LSTM

Long Short Term Memory networks or "LSTMs" are a type of RNN that can learn long-term dependencies. In the recurrent hidden layer of the LSTM, there are memory blocks, which are special units. Memory cells with self-connections store the network's temporal state, while special units called gates manage the flow of information in the memory blocks. Each memory block has an input gate, output gate and forget gate.

Apart from the input activation and output activation present in standard neural networks, LSTM also adds several other functions in its memory block. Like a simple RNN, it contains a hidden state, H(t-1) representing the previous timestamp's hidden state and Ht representing the current timestamp's hidden state. LSTMs also have a cell state, which is represented by C(t-1) and C(t), respectively, for the previous and current timestamp. The architecture of a single LSTM memory cell is shown in Figure 3 [10].

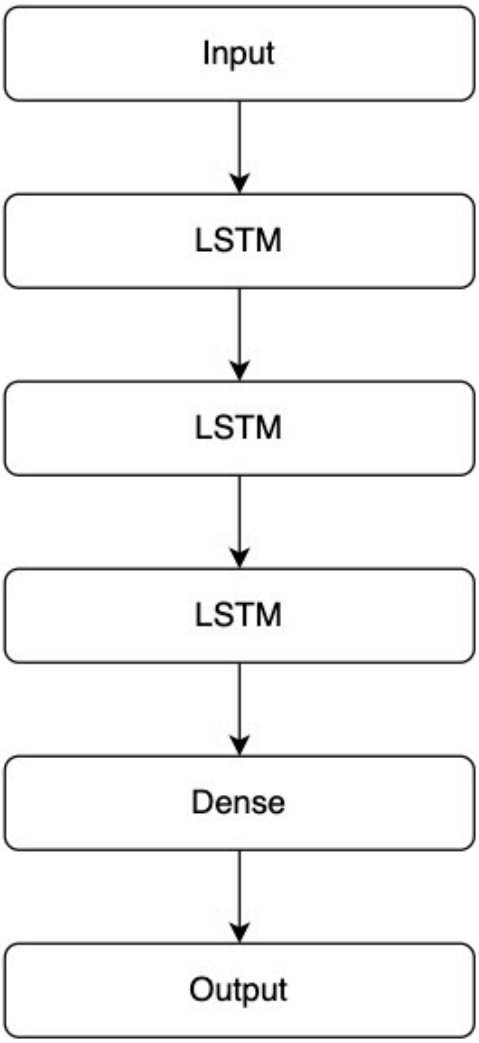


Figure 4: Stacked LSTM

For our research, we have reshaped the data first and added timestamp. Then we used the LSTM model from Keras library using similar model configurations as ANN and trained the data on a batch size of 32 for 10 epochs.

5.4 Stacked LSTM

An LSTM model with numerous LSTM layers is known as a stacked LSTM architecture. An LSTM layer sends a sequence of values to the LSTM layer after it, rather than a single value. Specifically, instead of one output time step for all input time steps, one output time step per input time step is used. The architecture of the stacked LSTM we have used is shown in Figure 4

We have stacked 3 layers of LSTM with the ReLU activation function and added an ANN layer with the sigmoid function to produce outputs as either 0 or 1. In this model also, we have applied the Adam optimizer and binary cross entropy for loss function.

5.5 Transfer learning

Transfer learning is defined as the ability to transfer the information of a pre-trained model to a new situation. In NLP, transfer learning can be divided into three categories:

- The nature of the source and target domains
- The order in which the tasks are learned; and
- If the source and target settings deal with the same task.

We employ cross-lingual learning with Transfer learning in our scenario because the source and target settings deal with the same task but in different languages. We are training our data in the Java programming language but are testing it in C++. So the same model which is used for defect prediction in Java is directly applied for C++ code without any additional training. Thus, transfer learning proves to be very beneficial because it doesn't require a huge training dataset for C++ and also saves training time.

6 EXPERIMENTAL RESULTS

The results achieved using the models mentioned above and then applying transfer learning on them, have been illustrated in table 2 and 3:

From table 2 and 3, it can be observed that ANN performs better than LSTM and stacked LSTM in terms of all performance metrics and is able to predict defects in Java code with an accuracy of 86.73% and C++ codes with an accuracy of 65.17%. This is a fairly good score considering our dataset size, and without training our model on C++ code.

The precision value denotes the percentage of positive predictions that were actually true i.e. how accurately the model can detect the number of software defects that will be present in a dataset and recall value denotes percentage of true positives that were correctly detected. The Area Under the Curve (AUC) is a summary of the ROC curve that measures a binary classifier's ability to distinguish between the two classes.

ANN with dropout performs better than LSTM because LSTM is a very powerful model that requires a huge dataset. Since our dataset is not large enough, there may be a problem of overfitting in case of LSTM and the model getting trained very specific to the training data and not being able to classify new data points properly. This is why it is not able to generalize well and be generic enough to accurately predict defects in C++ code through transfer learning.

7 THREATS TO VALIDITY

7.1 Internal Validity

The type of errors we considered as software defects in our study is currently limited to compilation errors i.e. errors in the syntax and structure of the code that does not allow a piece of code to compile. But there could be other types of errors such as run-time errors that may occur in certain situations depending on invalid input by the user, or logical errors that the program commits. These error types can be expanded to include these errors in future work. Secondly, we have used NLP models that are known to perform well on natural languages. There are several NLP models that have been trained on data of programming languages such as Code BERT which can be used. However, since BERT is a powerful deep learning model,

Table 2: Performance metrics for Java

Performance	Model		
Metrics	ANN with dropout	LSTM	Stacked LSTM
Accuracy	86.73%	62.81%	58.8%
Precision	71.72%	57.29%	55.52%
Recall	88.52%	98.96%	86.95%
AUC Score	77.01%	63.02%	58.95%

Table 3: Performance metrics for C++

Performance	Model		
Metrics	ANN with dropout	LSTM	Stacked LSTM
Accuracy	65.17%	60.31%	59.8%
Precision	63.92%	71.19%	66.52%
Recall	85.64%	47.97%	55.64%
AUC Score	62.61%	61.85%	60.32%

and our data may not be large enough to train on it properly, we have avoided its use.

7.2 External Validity

The data for our study has been mined from a coding contest website where codes are written to solve a particular problem statement. Although, this helps in making our data more generalized and not limited to a specific project, due to the variations in different systems designed for different domains with varying degrees of complexity, one could think more about the ability of our method to be applied in industrial applications. Similar research testing our method on other software systems will undoubtedly be valuable in generalizing our findings and improving our method.

The results produced in our research are based on data in the range of thousands of files. The results cover only a small number of datasets and may not necessarily generalize to other datasets. It may improve or worsen based on the quality of datasets. This threat can be eliminated by using a very large dataset containing data from multiple sources.

8 CONCLUSION AND FUTURE WORK

Defect Prediction is becoming a crucial part of software quality assessment as the quantity and complexity of software keeps growing day by day. Defect prediction can help by automatically identifying the areas of the software that contain bugs or faults. Generally, software metrics are used as features in traditional classifiers to identify problematic software modules. However, these metrics often do not consider the structure and syntax of the code which in many cases, could be the cause of error. Through this study, we have demonstrated the use of NLP techniques for defect prediction by using them for defect prediction and then used the same model for defect prediction in C++ code using Transfer learning. The results of this study show that embedding the code using Doc2Vec and then using ANN with dropout layers for training produces the best results.

Our research can be expanded in the future to include:

- **A larger dataset** - Currently our dataset consists of 17600 Java files for training which is not very large. This dataset can be enlarged to include more files from various programming websites to make it more comprehensive. Also, since NLP models are deep learning models that are very powerful, they work best with large datasets. So, the problem of overfitting can be avoided by using a larger dataset and obtaining a more generalized result that will be easier to transfer to more languages.
- **More NLP models** - We can explore more and newer NLP models such as BERT, GPT-2, GPT-3, CodeBert etc. Currently, our dataset is too small to be trained on such powerful models. But once we have a larger dataset, we can train it on these models and test their performance.
- **Transfer learning on more programming languages** - Currently we have used transfer learning from Java code to C++ code. This can be extended to more programming languages. It is expected to work better with object oriented languages such as Ruby and Scala. It can also be expanded for not strictly object oriented languages like Python, JavaScript etc.

REFERENCES

[1] T. Menzies, A. Butcher, D. Cok, A. Marcus, L. Lay- man, F. Shull, B. Turhan, and T. Zimmermann. 2013. Local versus global lessons for defect prediction and effort estimation. In Proceedings of the IEEE Trans. Softw. Eng., 39(6):822–834. http://doi.org/10.1109/TSE.2012.83
[2] J. Nam, S. J. Pan, and S. Kim. 2013. Transfer defect learning. In Proceedings of 35th international conference on software engineering (ICSE). IEEE, 2013. http://doi.org/10.1109/ICSE.2013.6606584
[3] IEEE-1012, "IEEE standard 1012-2004 for software verification and validation". 2004.
[4] Albert Endres, H. Dieter Rombach. 2003. A handbook of software and systems engineering: Empirical observations, laws, and theories. Pearson Education.
[5] Glass, Robert L. 2002. Facts and Fallacies of Software Engineering. Addison-Wesley Professional. Boston, USA.
[6] D. Budgen. 2006. Keynote address, CSEET'06.
[7] Le, Quoc, and Tomas Mikolov. 2014. "Distributed representations of sentences and documents." In Proceedings of the International conference on machine learning, PMLR. https://arxiv.org/abs/1405.4053
[8] Wang, Sun-Chong. 2003. "Artificial neural network." Interdisciplinary computing in java programming – Springer, Boston, MA. 81-100. https://doi.org/10.1007/978-

1-4615-0377-4_5

[9] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. 2014. "Dropout: a simple way to prevent neural networks from overfitting." Journal of machine learning research 15.1: 1929-1958. https://dl.acm.org/doi/epdf/10.5555/2627435.2670313

[10] Sundermeyer, Martin, Ralf Schlüter, and Hermann Ney. 2012. "LSTM neural networks for language modeling." In Proceedings of the Thirteenth annual conference of the international speech communication association. http://doi.org/10.21437/Interspeech.2012-65

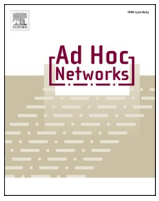
Ad Hoc Networks 130 (2022) 102823



Contents lists available at ScienceDirect

Ad Hoc Networks

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



UAVs assisted Network Partition Detection and Connectivity Restoration in Wireless Sensor and Actor Networks

Aditi Zear [a,*], Virender Ranga [b]

[a] Department of Computer Engineering, NIT Kurukshetra, India
[b] Department of Information Technology, Delhi Technological University, Delhi, India

ARTICLE INFO

Keywords:
Unmanned Aerial Vehicle
k-means++
partition detection
partition recovery
relay node
UAV-NetRest

ABSTRACT

Wireless sensor networks have been extensively employed in different data collection and monitoring applications because of their serviceability in harsh environments. The interconnected network is a fundamental requirement for these applications. The simultaneous failure of multiple nodes causes connectivity loss and can partition the network into various segments. It is crucial to identify the partitions in the partitioned network and reconnect them to achieve application-related goals. This paper provides a new Unmanned Aerial Vehicles (UAVs) assisted Network Partition Detection and Connectivity Restoration (UAV-NetRest) solution encompassing all phases from failure detection to network partition recovery. The proposed algorithm is tested according to detection and recovery time, distance traveled by UAVs in detection and recovery phases, the number of messages transmitted, and the number of deployed relay nodes. The performance of the UAV-NetRest's recovery algorithm is compared with the existing recovery techniques proposed in the literature. The simulation results demonstrate that the UAV-NetRest recovers the network with significantly less recovery time and minimum distance traveled by UAVs. The algorithm deploys minimum relay nodes as compared to other techniques. Network partitioning is an unpredictable incident, and simulation results entirely depend on the number of failed nodes, proximity and topology of network partitions, and the number of UAVs employed. The paper provides a new framework that covers all the aspects required for network partition recovery. This proposed work can be a seed source for future researchers to further optimize the detection and recovery phases.

1. Introduction

The recent industrial developments have increased the momentum for automation and accelerated the usage of Wireless Sensor and Actor Networks (WSANs) in different applications. The low cost, swift deployment, and effortlessly operable advantages favor the employment of WSANs as a de-facto mechanism for surveillance and monitoring in sterile environments such as combat region surveillance, space exploration, and environment monitoring [1,2]. WSANs can operate in such harsh environments without causing any harmful impact on humans and wildlife. Sensor nodes in various applications are set up in an ad-hoc way to implement application specialized tasks. Each sensor node coordinates with nearby sensor nodes to establish an interconnected network [3]. The essential requirements for achieving application-related targets are seamless data sharing and interconnected network. The sensor network may function within an unassisted environment. Therefore, the network must endure node or connection failures to keep the network connected and coordinate recovery without depending on external sources such as remote command centers or manual intervention [4,5]. However, working in a rough application area with immanent vulnerable situations escalate the chances of node failures. These failures can lead to connectivity loss and possibly partition the whole network into isolated segments. The extent of failure depends on the number of affected nodes and their positions in the network [6].

Network partitioned because of single cut-vertex node failure is easy to detect and recover by repositioning available healthy nodes or replacing failed nodes with another active node. However, it becomes more challenging to detect and recover network fragmented because of multiple dispersed and collocated node failures [7]. Primarily, it is challenging to identify the span of damage when numerous nodes fail since a healthy node cannot determine that it has missed contact with its neighbor nodes because of cut-vertex or multiple node failures. In [8], it is described that recovering a fragmented network is challenging

* Corresponding author at: Aditi Zear, Department of Computer Engineering, NIT Kurukshetra, Kurukshetra, 136119, Haryana, India.
E-mail addresses: aditizear93@gmail.com (A. Zear), virender.ranga@dtu.ac.in (V. Ranga).

https://doi.org/10.1016/j.adhoc.2022.102823
Received 20 August 2021; Received in revised form 21 February 2022; Accepted 24 February 2022
Available online 6 March 2022
1570-8705/© 2022 Elsevier B.V. All rights reserved.

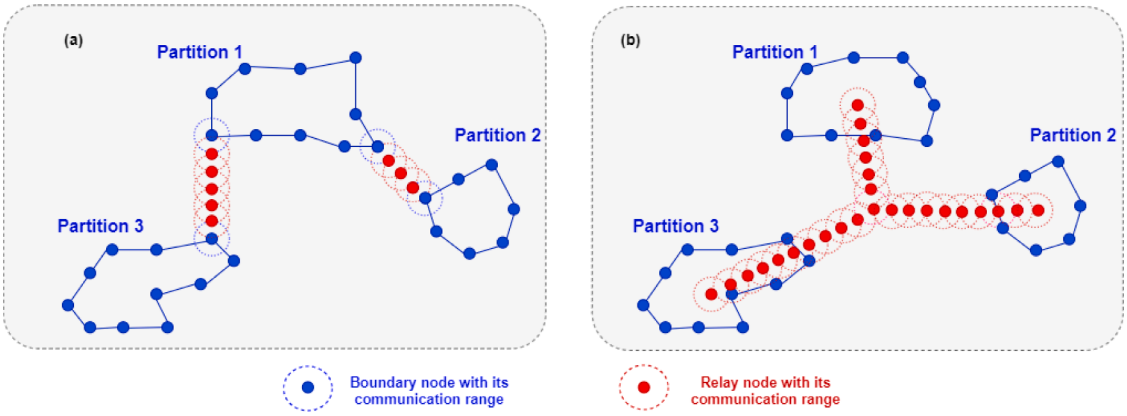


Fig. 1. (a) Relay nodes connecting boundary nodes of network partitions (b) Relay nodes connecting centroids of network partitions [15].

because of the lack of information available about the damaged area. It is impractical to know the details of failed and healthy nodes, the count of disconnected network segments, and the exact number of nodes in an individual partition. Most past research studies have not emphasized network partition detection and extraction of complete information about all the nodes in a partitioned network. Some studies [9–12] have attempted to detect the network partition. These studies could not distinguish whether there are fundamental node failures or actual network partitioning. The vast majority of published schemes to recover partitioned networks assume that information about the number of partitions and location of nodes in each partition is obtained through UAVs, satellites, mobile nodes, etc. Based on this assumption, these schemes proceed to apply recovery solutions [13–15]. However, it is challenging to get complete information about the number of partitions and nodes in each partition.

The methods to recover multiple node failures (collocated or dispersed) employ additional nodes (relay nodes) to operate as relay between separate partitions. The primary purpose of these approaches is to minimize the number of deployed relay nodes. Many studies [16–19] formulate that the relay node placement problem can be solved by constructing Steiner Minimum Tree (SMT). For SMT construction, a representative node symbolizes each segment and acts as an interface point. However, implementing SMT is an NP-hard problem [20] and pursues heuristic solutions. Usually, the standard heuristic approach to setup SMT includes three terminals in the sets of connected components and gradually interconnects them [17,18] towards the damaged area. This technique is inadequate because the shape and size of the individual segment are not considered. Instead of factoring in a representative node from each segment, employing boundary nodes of segment pairs [15] is more advantageous, as shown in Fig. 1. In this figure, a similar case is illustrated. Eight relay nodes are required to connect the network through boundary nodes of network partitions. However, recovering network partitions by connecting centroids of partitions requires twenty-six relay nodes. Therefore, the recovery solutions that employ centroids or representative nodes present inside the network partitions for SMT construction are not always efficient.

Conventionally, the strategies used for network partitions recovery are proactive and reactive. During network deployment, proactive strategies entail initial node placement to introduce redundancy by employing backup nodes or establishing k-connected network topology in advance. The strategy, as mentioned earlier, can lessen the chances of network partition problems. However, inadequate safeguards against massive scale damages can lead to node and their associated backup failures. On the other hand, reactive strategies restructure the network after the failure and form an intersegment topology to reinstate connectivity between disjoint partitions. The recovery procedure can involve relocating available healthy nodes or deploying additional RNs to create a connected topology.

Reactive strategies are implemented using centralized and distributed approaches [2,3]. Centralized approaches assume the information about the whole network state and then use this available information to propose an optimum recovery procedure. These approaches can yield an optimum solution only if they can acquire a complete network state through aircraft, mobile robots, satellites, etc. After collecting the information, it is crucial to determine whether the network is partitioned and details of nodes present in each partition. Therefore, many published techniques consider distributed approaches more applicable [6, 15].

In distributed approaches, each active node assembles information about the partitioned network locally by exploiting either one-hop or two-hop neighbor details, sending representative nodes or Mobile Data Collectors (MDCs) from each segment to gather at the commonplace. Collecting information through these methods and applying recovery solutions seems more practical. Furthermore, these approaches presume MDCs to have adequate storage space, which is feasible because of devices' high communication and computing capabilities these days [1,21, 22]. However, another assumption based on which some approaches work is that a sufficient number of RNs are present in the network and can relocate without any negative impact on network coverage. This assumption is not valid in practical situations because damages can occur randomly [15].

Although the techniques mentioned above have distinct limitations under different scenarios, a significant drawback of all these approaches is that none is applicable under extreme disaster scenarios and obstacle-prone areas. One solution for solving this problem is an Unmanned and Autonomous Aerial Sensing (UAS) solution that can assess the damaged network in such areas. There can be better recovery solutions if the information about the damaged network is more accurate and available within the time limit. As an emerging technology, Unmanned Aerial Vehicles (UAVs) are approaching fast to occupy skies, and industries are investing heavily in their development for many applications. These reliable vehicles with reasonable cost are in demand because of their numerous advantages such as adaptive altitude, controlled mobility, avoiding obstacles, line of sight (LOS) communication links, and flexibility [23–25]. The UAVs can strengthen the connectivity in wireless networks, coordinate with ground sensor nodes and recover the connectivity in the damaged network. Therefore, UAVs can assist in network partition detection and recovery. They can act as either RNs or MDCs. However, UAVs as RNs can increase the recovery cost when the network is damaged significantly, and UAVs as MDCs can only provide a temporary recovery solution [26–28].

This paper proposes a more efficient and cost-effective technique called UAV-NetRest that employs UAVs to deploy sensor nodes in the damaged network. Furthermore, a centralized mechanism is proposed that engages UAVs to get complete information about the damaged network. Network nodes detect the failed nodes in the network and send

UPDATE messages to the sink node. Sink node on getting the UPDATE messages demands new one-hop neighbor lists from the network nodes. On receiving updated lists from the network nodes, the sink constructs a new adjacency list and failed nodes list and shares them with the UAV control center. UAV control center employs UAVs to get the complete information about the network, updates the adjacency list from the received information from UAVs, and detects whether the network is partitioned by applying Depth First Search (DFS) [29] on the updated list. The UAV control center shares the updated adjacency list and failed node list to the sink node. The sink node first finds the convex hull of partitions and applies Minimum Spanning Tree (SMT) to find the least distant boundary pairs. Sink also applies some optimization steps to reduce the number of relay nodes and prepare the algorithm for small and large-scale failures. However, these optimizations depend entirely on the failed nodes, the topology of the partitions, and the proximity between boundary nodes of different partitions. After finding the total number of nodes required to reconnect the partitions, the sink forwards their positional coordinates to the UAV control center. UAV control center employs UAVs to deploy nodes at the given coordinates. The k-means++ [30] clustering algorithm divides the failed and relay nodes into clusters in the detection and relay node placement phases. In both detection and deployment processes, the proposed algorithm tries to reduce the UAV's travel distance by computing every time the minimum distant node from UAVs in their clusters. Furthermore, UAVs reduce their travel distance by not visiting the nodes whose ids are present in one-hop neighbor lists (OHNL) received from active network nodes. These active nodes are assumed as failed nodes in the failure detection phase by the sink node.

2. Related work

Recently, enduring node failures in WSANs operating in hostile terrains have raised substantial attention. According to the various published schemes [1,21,22], there are four failure recovery approaches. The first approach restores the network by relocating some active nodes [2,5,11]. However, these approaches assume that surviving nodes can move to restore communication links and require complex hardware. Furthermore, the relocation process also leads to high energy consumption [2,4,31]. The approaches in the second category rely on provisioning additional nodes as either backup or for establishing k-connectivity with the aim that the network operates even after some node damages [32]. However, this technique results in massive communication overhead and significant deployment expenditure. The third network recovery approach is relay node deployment (RND). It is the most popular and efficient mechanism of deploying additional nodes to restore intersegment topology. However, it depends on the assumption that many relay nodes are present in the network for deploying [6, 10,33,34]. The fourth approach is a temporary recovery mechanism that employs Mobile Data Collectors (MDCs) [1] to provide intermittent connections between the partitioned network segments [21,22]. This solution is convenient whenever the availability of relay nodes is inadequate. However, it is not an efficient solution because of the increased overhead and data latency.

An additional way to classify the network recovery mechanisms relies on the number of failed nodes. Surviving a single cut-vertex failure is done by restructuring the network topology and replacing the failed node with either a surviving node or an additional relay node. Partition Detection and Recovery Algorithm (PADRA) in [35] detects the partitioned network by identifying the failure of cut-vertex and reconnects the partitions through the controlled movement of active sensor nodes. Abbasi et al [36]. proposed the least disruptive topology repair algorithm (LeDiR), which shifts the block with fewer sensor nodes among the network segments for recovery. Distributed Node Recovery Algorithm (DARA) in [37] establishes connectivity in a partitioned network with few actor networks. A hybrid approach Distributed Prioritized Connectivity Restoration Algorithm (DPCRA) [38] assigns failure handlers (FH) to all cut-vertices in the network. When the network gets partitioned, the FH nodes handle the recovery operation. Chouikhi et al [5]. proposed two approaches, Preventive Failure Recovery (PFR) and Rotating Nodes based Failure Recovery (RFNR), for restoring disjoint network partitions. The former approach ensures the network should be robust from network partitions. The latter approach consists of algorithms to recover the partitioned network. The main objective of these techniques is to minimize the number of nodes involved and travel distance. Nevertheless, these techniques cannot handle multiple node failures since the surviving nodes cannot estimate the damage in network topology.

Multiple node failures can be either collocated or dispersed. Centralized and distributed techniques handle the case of numerous dispersed node failures. Distributed techniques are challenging to apply because of resource conflicts that arise during recovery. Generally, a node may need to relocate simultaneously to serve two single node failures at different locations. The usage of synchronization mechanisms can solve this dispute [32,39]. The centralized techniques handle this case by resolving multiple single-node failures independently [40–43]. Loss of numerous collocated nodes is crucial to handle since the scope of failure is difficult to identify by the surviving nodes. The recovery solutions are node relocation, employing MDCs, and relay node placement. Recovery approaches that rely on node relocation first fill the network hole created from simultaneous collocated node failures with relay nodes and then apply optimization techniques to minimize relay node count. Senturk et al [14]. proposed a heuristic and distributed approach to heal partitioned sensor networks. The partitions implement the game theory model based on Nash equilibrium. The nodes in every partition compute the equilibrium value of neighboring partitions, which determines the probabilistic function of these partitions separately. The nodes in other partitions move towards the partition with dominant probabilistic function value for establishing connectivity until the network becomes connected. The simulation results demonstrate that the proposed algorithm provides good coverage even with increased network size and reduces the movement cost of the nodes. However, the recovery time performance is a little less than the centralized heuristic-based approaches. Senturk et al [44]. deploy RNs in the center of the region using game theory. The relay nodes then spread towards the boundary of partitions. The virtual force-based model guides the movement of these RNs.

DaarDs [45] and DORMS [18] employ mobile relay nodes to establish connectivity. When the network gets fragmented, the nodes from each partition move towards the pre-decided common point till they come into each other's communication range and re-establish connectivity. After restoring the network, the optimization phase reduces the count of deployed RNs. AuR [46] reconnects the partitioned network by moving the nodes present in the boundaries of segments as blocks towards the center of the area. Approach in [47] factors obstacles and decides the candidate nodes and their travel path to restore the network. In [11], Round Table Negotiation (RTN) approach selects a negotiator from every partition. These negotiators meet at the round table to share information about their partition. Then, based on combined data from all the partitions, some selected candidate nodes move towards the region between the segment pairs to re-establish connectivity.

Many studies focus on relay node positioning intending to form connected intersegment topology and minimize the relay count. Cheng et al [48]. employ a three-step method that connects nodes with a distance less than R first. The second step consists of calculating a Steiner point that can connect a subset of three nodes. The final stage is creating a minimum spanning tree for the remaining nodes and populating it with relay nodes. The approach in [49] employs a Geometric Disk Cover Algorithm for SMT construction that minimizes relay count. Notably, at least one relay node connects each node of the network and forms an inter-relay topology. ORC [19] identifies a convex hull and sets up relay nodes inwards from its boundary. In [17] FeSTA utilizes Discrete Fermat Point optimization that selects triangles to connect the partitions. Senel et al [13]. proposed an algorithm to attain better coverage and node

degree by providing two-vertex connectivity. 1C-spider web provides on-vertex connectivity by forming a ring. Further improvement in this ring offers two-vertex connectivity by connecting partitions with different nodes to create another individual path. Similarly, CORP [50] focuses on creating a highly connected topology with a node degree greater than two. Delauney triangulation in [16] optimizes the creation of SMT and utilizes the edges of the triangle to select Steiner points. These solutions represent every partition with some terminal and do not consider their shape and size.

Lalouani et al [15]. proposed Boundary-aware optimized Interconnection of Disjoint segments (BIND to solve the problem of recovering connectivity in immensely separated network segments by setting up the fewest relay nodes (RN). Rather than representing every segment using a single node, this approach considers its shape and size. BIND prefers to maintain the path between the disconnected segments through additional Steiner points and reconnect them by structuring shortened length topology with the help of boundary nodes.

Mobile data collectors (MDCs), also known as data mules, collect and disseminate data in fragmented and sparse networks [51]. Many studies employ MDCs to restore partitioned networks as a good solution when the availability of RNs is limited. The main focus of these studies is to reduce tour length to save energy and minimize data latency. In MINDS [7] and IDM-kMDC [52], MDCs tour the group of disjoint network segments. It utilizes a minimum spanning tree (MST) as a backbone to divide the set of disconnected segments into groups. MINDS [7] separate MST from midpoint and keep subdividing the larger half until all MDCs have their groups. IDM-kMDC [52] applies the greedy method and assigns an MST edge to each MDC. After that, it merges the MDC tours with the least total path length until all MDCs have their segment groups. Thus, the number of MDC tours should be equal to the available MDCs count. Lalouani et al [22]. employ MDCs to visit segments and transport their data. The authors focused on curtailing the tour length to minimize the data latency and traveling overhead. Their approach recovers the network with a fixed MDCs count and considers the segments' size and shape. The proposed Connecting Isolated Segments through Intermittent Links (CISIL) approach works in two steps. In the first step, Delaunay triangulation (DT) determines the effective tours. The second step selects the subset of these computed tours with the minimum length that forms a fully connected topology. This optimization problem maps into K-edge MST in a hyper-graph of the estimated MDCs tours. Similarly, in [21], DT determines the MDCs tours. This approach also considers the presence the obstacles while planning the path for MDC tours. However, these approaches do not work when some MDCs cannot move and stay stationary in the recovered network. MiMSI [53] considers this case and utilizes k-means++ to divide the disconnected segments into groups. The MDCs tour these groups of segments. Furthermore, it uses stationary RNs to shorten the MDC travel path that has the most extended length.

Overall, the reported past research studies emphasize network recovery solutions compared to partition detection, and the studies focused on the detection process could not differentiate between network partitioning and fundamental node failures. Moreover, the proposed recovery solutions based on network node relocation and relay node placement do not work in disaster and obstacle-prone areas. These two techniques work on the assumption that sufficient relay nodes are present for recovery in all failure scenarios. MDC-based approaches provide temporary recovery solutions with increased data latency. Most of these recovery techniques do not consider the shape and size of the partitions and represent them with some terminal nodes. Furthermore, these approaches do not focus on network recovery confirmation after applying recovery solutions.

The proposed UAV-NetRest algorithm acknowledges these missing elements. The algorithm distinctively detects the number of partitions and nodes in those partitions. UAV-NetRest considers the shape and size of the partitions during the recovery process. After the relay node deployment process, the sink node confirms network recovery when it receives OHNL from all the surviving nodes.

3. Contributions

As per our best knowledge, any work applying the combination of RND and MDC-based techniques for network partition detection and recovery is not reported yet in the literature. MDC-based approaches provide temporary recovery solutions. Relay-based solutions are more focused on mapping the partition recovery problem to an optimization problem. Hence, these approaches focus less on determining the number of partitions and complete information about the nodes in these partitions and how the network recovery is confirmed after relay node placement. In some works, it is assumed that aircrafts such as drones, robots equipped with a camera, mini-robots, etc., [14,15] collect all the information about the network. It is nowhere mentioned in these works that how such devices can detect network partitioning and provide complete information about the partitions. After detecting some node failures in distributed approaches, nodes start collecting information about the reachable nodes by sharing and receiving their one-hop neighbor list in various iterations [10,11]. After some iterations, the nodes have information about their clusters. Still, this distributed solution can not differentiate between partition detection and multiple node failures. The proposed algorithm UAV-NetRest in this paper covers all these left-out aspects and works in different phases. The main contributions of the proposed method are summarized below:

UAV-NetRest engages network nodes for sending update messages such as failure and recovery reports to the sink node. The energy consumption is reduced by not involving deployed nodes in relocation and detection processes during recovery.

The network partition detection phase is proposed, which employs a UAV control center to accumulate complete network information and confirm whether the network is partitioned.

UAV-NetRest dynamically adjusts the UAVs count according to the number of failed nodes. The UAV control center commands UAVs to get complete details about the network. The k-means++ clustering algorithm divides the failed nodes into k clusters and assigns these clusters to UAVs.

UAV-NetRest reduces the distance traveled by UAVs. UAVs move towards the minimum distant node in their respective clusters and identify whether the node has failed. If the UAV receives a one-hop neighbor list (OHNL) from this node, it further attempts to reduce its travel distance by comparing the received OHNL with the failed nodes in its cluster.

UAV control center receives details about the network from UAVs and determines the number of partitions by applying the DFS algorithm.

The recovery mechanism at the sink node identifies minimum distant boundary nodes to reduce the relay node count.

UAV-NetRest further attempts to reduce the relay node count by checking the new relay nodes' coverage. In addition, the algorithm works for small and large-scale failures and compares the relay node count with failed nodes before deployment.

UAV control center manages the relay node placement process. Like the partition detection phase, k-means++ clustering divides the relay node locations and assigns a cluster to every UAV. The algorithm adjusts the UAVs' count according to the number of relay nodes. The UAVs move towards minimum distant locations every time to reduce their travel distance while deploying relay nodes.

The sink node confirms the network recovery when it receives recovery reports from all network nodes except the failed ones.

Employing aircraft for detecting and recovering the network is a new research dimension with its challenges. This paper provides a framework for future researchers to mitigate further challenges related to network partition detection, recovery, and employing UAVs in different recovery phases. Different techniques for collecting and deploying the sensor nodes by UAVs have not been explored as a part of this study, which could further be investigated in future research.

The structure of the remaining paper is: Section 4 and Section 5 depict the system model and proposed algorithm Section 6. presents the

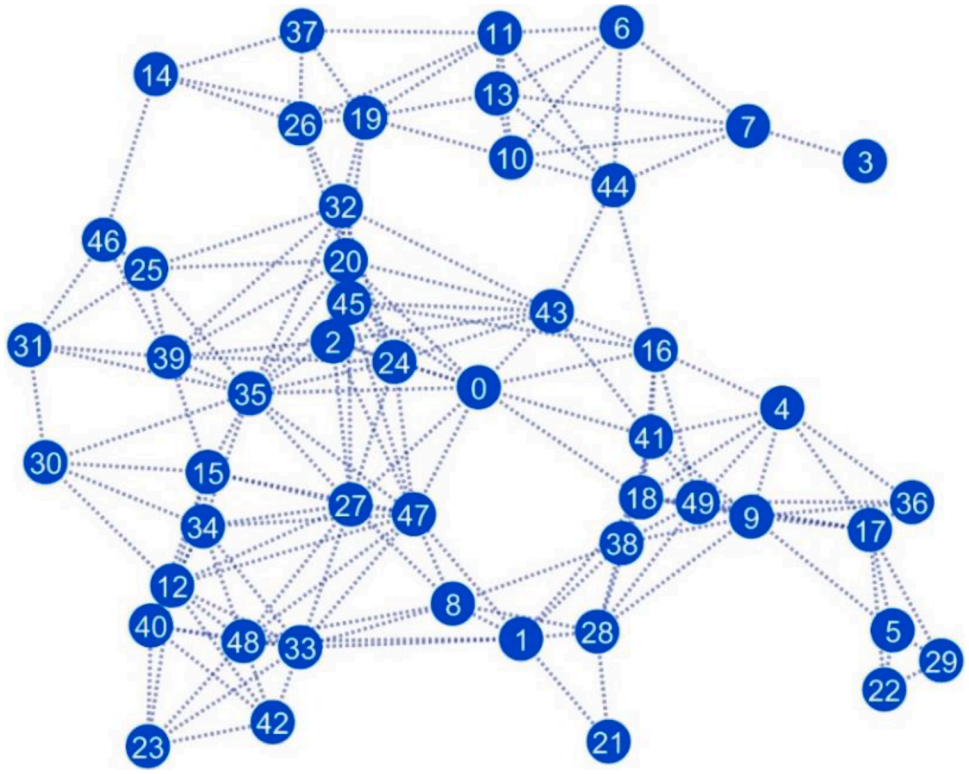


Fig. 2. A Sample Connected Network Topology.

Algorithm 1
Failure Detection.

Total number of network nodes =N
for(i= 0 to N-1)
Receive HEARTBEATs from $node_i$.
if ($node_i$ misses HEARTBEAT from k nodes), where k = 0 to N-1.
$node_i$ sends failure report to sink.
end if
end for
for(i= 0 to N-1)
if ($node_i$ receives ECHO message from the sink)
Send REPLY message including OHNL to sink.
end if
end for
for(i = 0 to N-1)
if(sink receives REPLY from $node_i$)
sink updates Adj with newly obtained OHNL.
end if
else if(Missing reply from $node_i$)
sink marks entries for $node_i$ in Adj as -1.
sink includes $node_i$ and its positional coordinates in FNA.
end if
end for
sink forwards FNA and updated Adj to the UAV control center.

simulation results and analysis. The conclusion and future scope are discussed in Section 7.

4. System Model

We consider a randomly deployed network consisting of sensors, actor nodes, and the sink node. All nodes are stationary; however, they can move whenever required except the sink node and UAV control center. The sink node collects data from sensor nodes in a direct or multi-hop manner. All nodes are aware of their location and maintain one-hop neighbor information by sending periodic HEARTBEAT, BEACON, or HELLO messages. During network initialization, nodes send their one-hop neighbor information and positional coordinates to the sink node. The multi-hop information transmission process consumes more energy of network nodes present near the sink than other nodes, leading to the quick failure of these nodes.

The entire network is depicted as a graph G = (V, E), where V represents a set of network nodes (sensors or actors), and E represents the edges between the nodes in each other's communication range. Various events such as forest fire, earthquakes, etc., can partition the network into multiple disconnected partitions. It is essential to know whether the network is partitioned and information about the partitions before proceeding with the recovery process. Therefore, UAV-NetRest employs UAVs to evaluate the extent of failure. After getting the complete information from UAVs, the UAV control center determines whether the network is partitioned or not. If the network is partitioned, the sink collects information about the partitions with the help of the UAV control center and starts the recovery process. The system model consists of the following assumptions without loss of generality:

Nodes are aware of their geographical locations.

All sensor, actor, relay, and UAV nodes have the same communication range (R) and initial energy.

It is not possible to recharge sensor and actor nodes. However, the UAVs, UAV control center, and sink can be recharged.

Packet collisions and retransmissions because of link errors are not considered.

Network nodes remain intact during the recovery process.

5. Proposed Approach

UAV-NetRest first considers a connected network in which every node can communicate with its one-hop neighbors. Sink collects information about the network topology during the initialization phase. We consider a scenario of simultaneous multiple node failures. Node failures can be because of physical damage, onboard equipment failure, or external events such as floods, fire, etc. Nodes can detect failure when periodic messages from their neighbors are absent and report it to the sink node. The entire network partition detection and recovery process

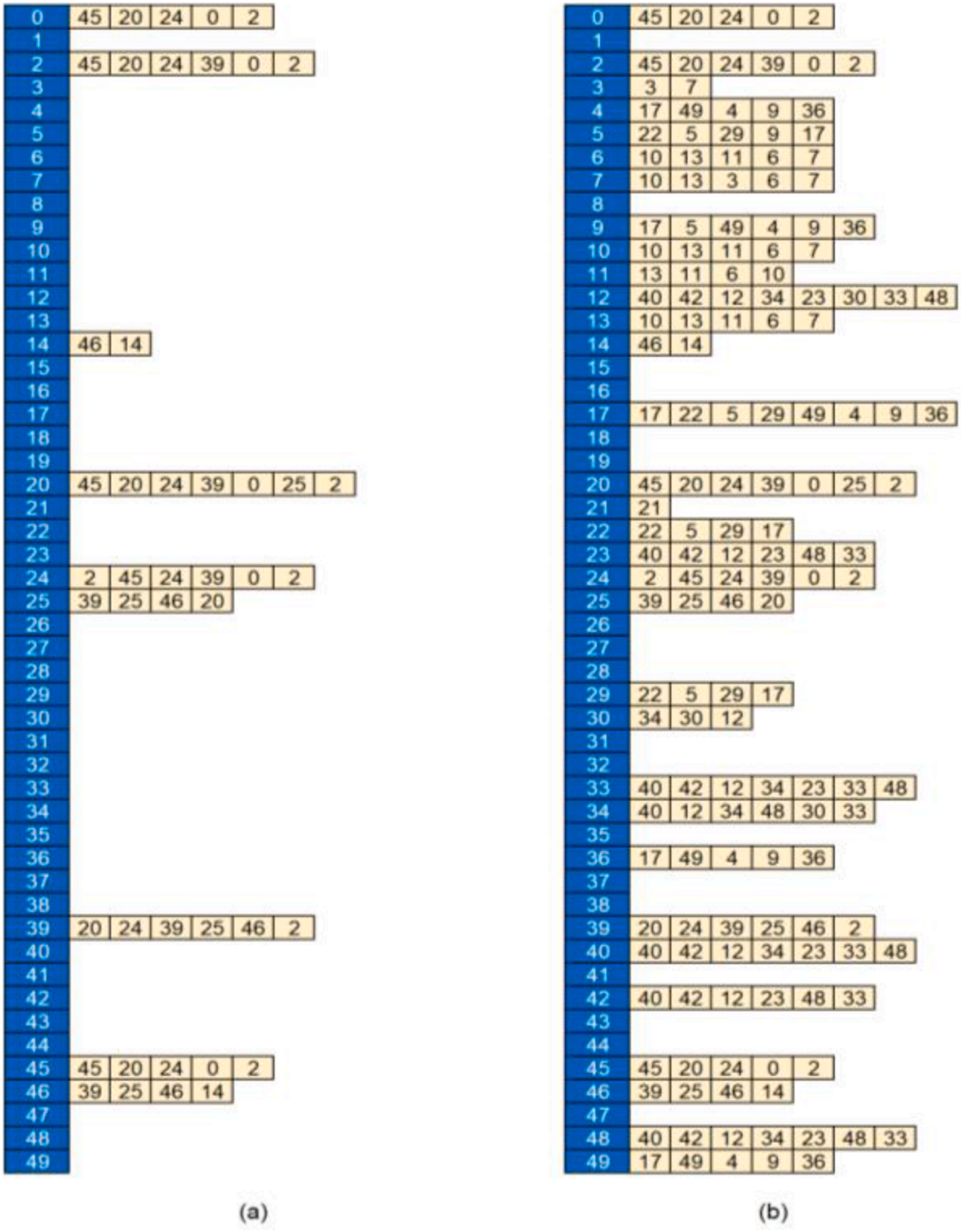


Fig. 3. (a) Adj at sink node after node failures, (b) Adj updated by the UAV control center in the Network Partition Detection phase.

are divided into five phases: Network Configuration, Failure Detection, Network Partition detection, Network Recovery, and Relay Node Placement. A detailed discussion about these phases is given below:

5.1. Network Configuration

In this phase network initializes, and nodes start sending and receiving periodic HELLO messages to the nodes in the network. When a node receives a HELLO packet from other network nodes, it includes their node_ids into its OHNL. Afterward, these network nodes send their configuration report consisting of node_id, OHNL, and positional co-ordinates (X_{coord}, Y_{coord}) to the sink node. Sink accumulates data from the network nodes and stores it in an adjacency list (Adj). Adj consists of n rows, where n is the number of nodes in the network, and each row consists of a OHNL of deployed nodes Fig. 2. shows an example network of 50 nodes.

Additionally, sink constructs Network Configuration Table (NCT) having Node_id, X_{coord}, and Y_{coord} attributes. Nodes coordinate with the sink to communicate essential updates such as configuration, failure, and recovery reports. Besides this, the network operates according to the standard procedure, and nodes communicate by sending HEARTBEATs after a specific pre-defined interval (t_{del}).

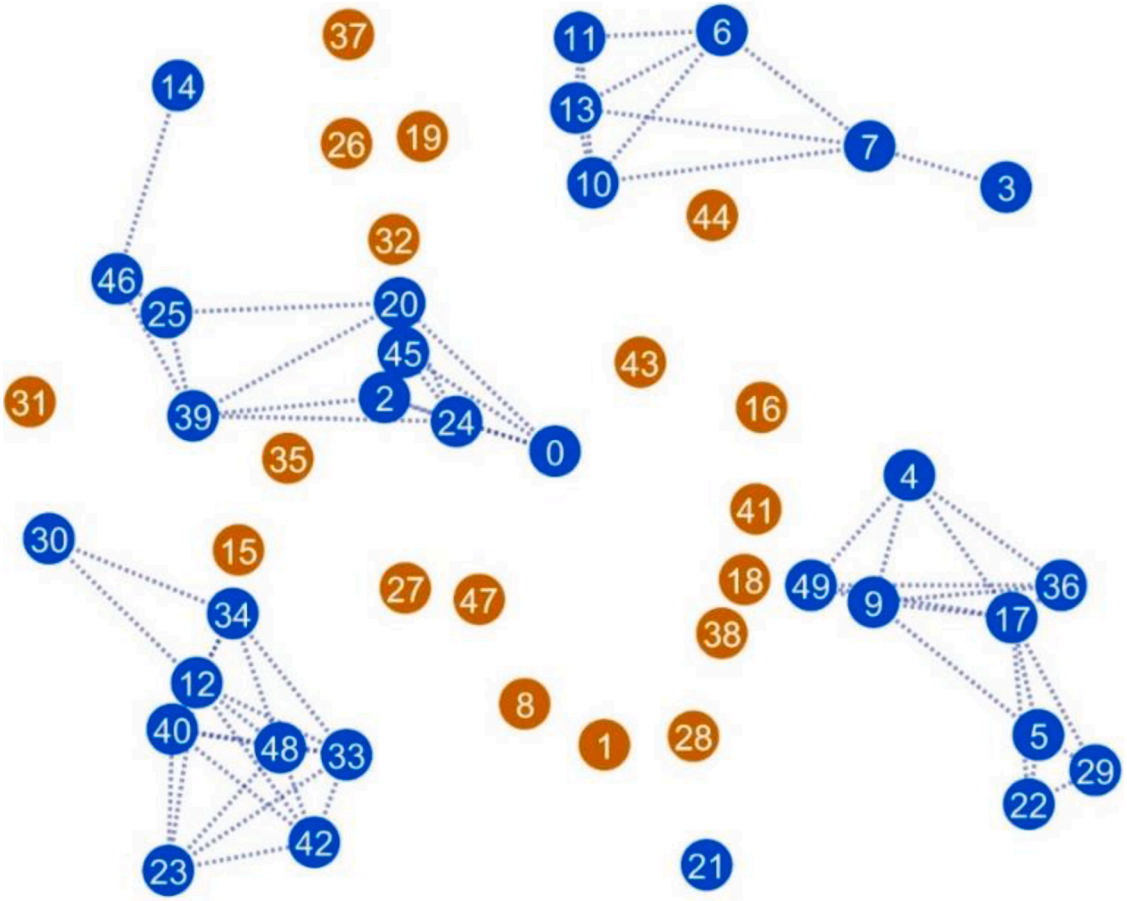


Fig. 4. Network after multiple node failures.

5.2. Failure Detection

When a node detects the missing HEARTBEATs from its one-hop neighbors, it sends a failure report to the sink node consisting of node_id/node_ids of the non-responding nodes. As a result of the damaged network, some nodes may not send failure reports whenever they detect missing HEARTBEATs from their neighbors because of the lost path with the sink node. A scenario after multiple node failures is given in Fig. 4.

Therefore, when the sink receives failure reports, it sends ECHO messages to all the network nodes. When nodes receive an ECHO message from the sink, they immediately send a REPLY message to the sink consisting of their node_ids and OHNL. Lines 1-11 in Algorithm 1 describe this process. Sink node updates Adj with newly obtained OHNL. The sink node checks the nodes from where it has not received the REPLY message and includes it into FAILED NODES ARRAY (FNA). FNA consists of node_id of failed nodes in the first column and their positional coordinates in the second and third columns Fig. 3.(a) shows the updated Adj. Sink has OHNL of a few nodes, which may lie in its partition, and it considers other nodes as the failed. However, not all the nodes in FNA are failed. They could be present in different partitions.

The number of nodes in FNA is denoted as n_{fail}. The sink node marks the entries for the failed nodes in Adj as -1 as described in Eq. (1). In this equation, n refers to the total number of nodes present in the network and FNA[i][0] provides the node_id of failed nodes. This newly created FNA is then forwarded along with an updated Adj to the UAV control center to identify whether the network is fragmented or not. Lines 12-21 in Algorithm 1 represent the steps taken by the sink node. The failure detection process encompasses transmitting different messages by the sink and network nodes and updating Adj. Therefore the complexity of the proposed algorithm is O(n).

$$\text{Adj}[\text{FNA}[i][0]][j] = -1 \quad \text{where} \begin{cases} i = 0 \text{ to } n_{fail} - 1 \\ j = 0 \text{ to } n - 1 \end{cases} \tag{1}$$

5.3. Network Partition Detection

The UAV control center receives FNA and Adj from the sink node and employs UAVs to determine whether these nodes are failed or appear as failed because of the missing path with the sink node. In this paper, n_uav denotes the number of UAVs present in the network Equation 2. determines UAV's count (n_uav_d), for employing in the detection process, which depends on n_{fail} and n_uav.

$$\begin{cases} n_uav_d = n_uav, & \text{if } n_{fail} >= n_uav \\ n_uav_d = n_{fail}, & \text{if } n_{fail} < n_uav \end{cases} \tag{2}$$

$$k = n_uav_d \tag{3}$$

It is simple to assign UAVs when the number of UAVs is equal to or more than the failed node count. However, the UAV assignment is complex when the number of failed nodes is more than UAVs. In this case, the factors to consider are grouping failed nodes based on their proximity to reduce the total distance traveled by UAV. It is vital to minimize the distance traveled by UAVs because this process consumes more energy than other processes such as message transmission or performing computations. Therefore, we have employed a k-means++ clustering algorithm [30] that groups the failed nodes into kclusters. The value of k is equal to n_uav_d from Eq. 2, and Eq. 3 describes it mathematically. Lines 1 – 4 in Algorithm 2 represent the clusters creation

Algorithm 2
Network Partition Detection.

1 UAV control center receives FNA from the sink node.
2 UAV's count (n_uav_d) is determined based on total failed nodes using (n_{fail}) using Eq. (2).
3 Failed nodes are assigned to the UAVs using k-means++, where k is determined using Eq. (3).
4 Assign a cluster FNA_i to every UAV_i, where i = 1 to k.
5 for(i=0 to k - 1)
6 UAV_i identifies minimum distant node at $cluster_i$ using equation 4.
7 UAV_i move towards the minimum distant node.
8 UAV_i sends ECHO message after reaching the node's location.
9 if(UAV_i receives REPLY message from the node)
10 UAV_i saves received OHNL in Adj_UAV_i.
11 UAV_i waits for some time interval t_uav_{del}.
12 for(j = 0 to number of nodes in FNA_i - 1)
13 for(k = 0 to number of nodes in received OHNL - 1)
14 if($FNA_i[j]$ = OHNL[k]
15 Remove $FNA_i[j]$ from FNA_i
16 end if
17 end for
18 end for
19 Repeat steps 6-7 with updated FNA_i.
20 else if(UAV_i do not receive REPLY from the node)
21 Repeat steps 6-7 with FNA_i.
22 end if
23 end for
24 UAVs return to their initial position after covering their respective clusters.
25 for (i = 0 to k - 1)
26 UAV_i share Adj_UAV_i with the UAV control center.
27 UAV control center merges Adj_UAV_i with Adj using equation 5.
28 UAV control center fills missing values in Adj with -1.
29 end for
30 UAV control center recomputes FNA from updated Adj.
31 for (i = 0 to k-1)
32 if(FNA_i is not empty)
33 Include FNA_i in FNA.
34 end if
35 end for
36 Apply DFS to find the number of connected components n_{part} in Adj.
37 The nodes in the connected components are stored in $Part_{ij}$
38 UAV control center removes failed nodes from $Part_{ij}$
39 for(i = 0 to n_{part} - 1)
40 for(j = 0 to number of nodes in FNA - 1)
41 if($Part_{ij}[i][0]$ = FNA[j])
42 Remove $Part_{ij}[i][0]$from $Part_{ij}$.
43 $n_{part} = n_{part} - 1$
44 end if
45 end for
46 end for
47 if ($n_{part} > 1$)
48 Network is partitioned
49 end if
50 else if ($n_{part} = 1$)
51 The network is connected, and some nodes failed
52 end if

process.

Afterward, a cluster is assigned to each UAV in the network. This assignment process comprises providing UAV a separate vector (FNA_i) of node_ids and positional coordinates (X_f,Y_f) of failed nodes in the cluster. The structure of FNA_i is similar to FNA. Here' i' denotes cluster_ids which range from 0 tok − 1.

We have focused on optimizing the movement of UAVs to their assigned clusters in two steps. The UAV executes the first step before moving to a particular location. In this step, UAV computes its Euclidean distance from every failed node present in its cluster and finds the node present at a minimum distance as given in Eq. (4). The current coordinates of the UAV are denoted as (X_{uav},Y_{uav}).

$$I_{nearest} = \text{argmin}\left\{\sqrt{\left(X_{uav} - X_f\right)^2 + \left(Y_{uav} - Y_f\right)^2} : f \in 0 : N_c - 1\right\} \quad (4)$$

Where $I_{nearest}$ is the index of the closest node in the cluster and N_c corresponds to the number of failed nodes in the cluster. The UAV then moves to that node's location, and upon reaching that location, it sends an ECHO message to that failed node. If the node is active and receives the ECHO message, it checks the message's source address and sends a REPLY message to the same address. This REPLY message includes their node_id and OHNL. UAV waits for some time interval t_uav_{del} to receive the message. This time interval varies for different applications. It is directly proportional to the time interval (t_{del}) after which the applications send HELLO messages to their one-hop neighbors. When UAV receives REPLY message from the presumed failed node, it removes the node_id of that node from FNA_i and saves its OHNL in a separate adjacency list (Adj_UAV_i). Adj_UAV_i refers to the adjacency list of UAV_i, where the value of 'i' ranges from 0 to k-1. Lines 5-10 in Algorithm 2 depict the process of movement and adjacency list creation by UAVs.

Afterward, UAV executes the second optimization step in its waiting time interval. It compares the node_ids in FNA_i with the OHNL received from the previously estimated failed node. If it finds the node_id of some failed node from FNA_i in OHNL, it concludes that the node is active and removes its node_id from the FNA_i. This step can remove more than one active node from FNA_i as it highly depends on the proximity of nodes in a cluster assigned to the UAV. By removing the node from FNA_i, UAV eventually reduces its total traveling distance and energy consumption. Before moving to the next location, the UAV again finds the least distant node using Eq. 4. Therefore, UAV can reduce its traveling distance by executing these two optimization steps every time. However, as mentioned before, these two steps rely on the physical proximity of nodes in the clusters. All UAVs perform these steps when they move towards their respective clusters. After covering all the nodes in their assigned cluster, UAV returns to its initial location. Lines 11-24 in Algorithm 2 describe the process of moving and checking OHNL from active nodes by UAVs.

UAVs immediately share their saved adjacency list with the UAV control center when they reach their initial location. UAV control center merges the adjacency lists from all the UAVs with the adjacency list (Adj) previously obtained from the sink node as described in Eq. (5). Fig. 3.(b) shows the updated Adj, the failed nodes in this list are reduced compared to the failed nodes in Fig. 3(a). Here, $FNA_i[j][0]$ gives the node_id of the failed nodes that UAVs were inspecting and n_{fail_c} is the number of failed nodes in the UAVs' cluster. After this detection process, the UAV control center has the complete graph of the network in Adj. It computes FNA from FNA_i received from UAVs. Lines 25-35 in Algorithm 2 describe these steps.

The next problem is to find the number of partitions and the node_ids of the active nodes in their respective partitions. Finding strongly connected components in the graph is one of the applications of DFS [29,54] and BFS [55] graph traversing algorithms. We have employed the DFS algorithm to find connected components in the updated adjacency list Adj. The DFS algorithm works in different rounds and marks all nodes 'unvisited' in a stack. The algorithm marks the first node and its adjacent nodes 'visited' and pops it out of the stack. After that, it takes the next node marked 'visited', marking its adjacent nodes 'visited', and popping it out of the stack. This process continues until the algorithm finds an 'unvisited' vertex. Therefore, the first connected component consists of nodes popped out from the stack in the previous round. Whenever the 'unvisited' nodes are encountered, DFS is called recursively to find the next connected components until all the nodes are marked 'visited' and popped out of the stack. Hence, the DFS algorithm determines the total connected components (n_{part}) and the node_ids present in each connected component.

$$\begin{cases} Adj[FNA_i[j][0]][l] = Adj.UAV_i[FNA_i[j][0]][l], \text{if } Adj.UAV_i[FNA_i[j][0]][l] != -1 \\ \text{where} \quad \begin{cases} i = 0 \text{to} k-1 \\ j = 0 \text{to} n_{fail}-1 \\ l = 0 \text{to} n-1 \end{cases} \end{cases} \quad (5)$$

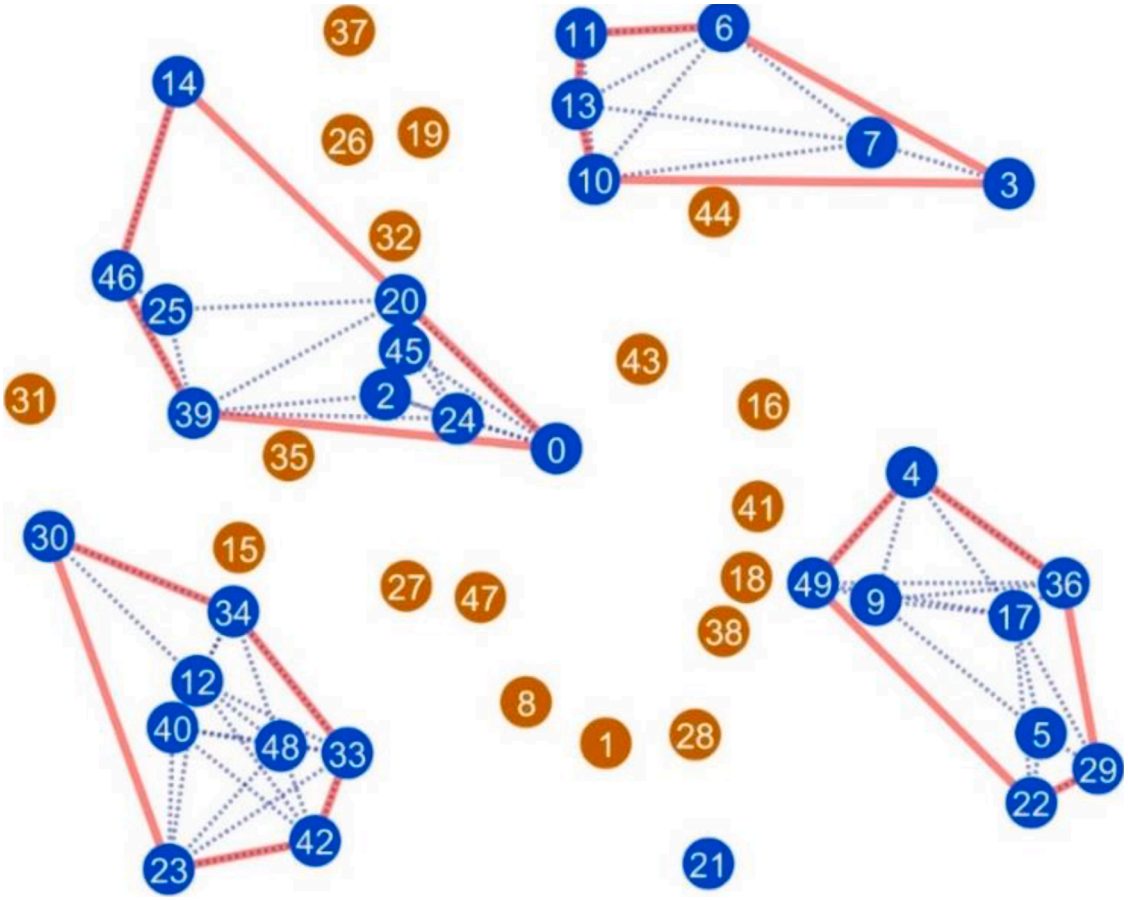


Fig. 5. Convex hulls computed for every Network Partition.

The DFS algorithm identifies all the partitions in the graph, which also includes failed nodes. The UAV control center filters out these failed nodes and creates a final vector$Part_{ij}$. The rows of $Part_{ij}$ consist of partition ids, and columns consist of node ids present in the particular partition. Lines 36-46 describes the process of filtering out failed nodes and creation of$Part_{ij}$.

These connected components refer to the partitions in the network. UAV control center concludes that the network is partitioned if the value of n_{part} comes out greater than one. In this way, our proposed algorithm differentiates between the network partition problem and general node failures, which is not seen in any earlier proposed solutions. Lines 47-52 in Algorithm 2 describe the process of detecting partitions in the graph. Finally, the UAV control center shares $Part_{ij}$ and updated FNA with the sink node. The time complexity of the DFS algorithm and computation of the total number of partitions is $O(n + E)$ and $O(n^2)$, respectively. The complexity of kmeans++ is $O(n_{fail}kI)$, where I is the number of iterations required by kmeans++ to converge. Therefore the time complexity of the network partition detection algorithm is $(O(n + E) + O(n^2) + O(n_{fail}kI))$, which is equivalent to $O(n^2)$.

5.4. Network Recovery

The sink node on receiving $Part_{ij}$ and updated FNA from the UAV control center immediately starts the recovery process. The recovery process consists of finding the locations to deploy the relay nodes to reconnect the partitioned network. As mentioned previously, finding recovery solutions by representing partitions with terminal nodes is not an efficient solution always. It is more beneficial to consider the size and shape of the segments and employ boundary nodes for the recovery process. Therefore, UAV-NetRest considers boundary nodes by computing the convex hull of every partition. A prevalent and efficient method, i.e., the Graham scan algorithm, calculates the convex hull of all network partitions [56].

$Part_bound_{ij}$ stores the convex hull of all network partitions Fig. 5. shows the convex hulls of different partitions in the sample network. Lines 1-6 in Algorithm 3 discuss the process of computing convex hulls. The next step is to find the closest partition to each partition and the path with the shortest length to join them. Firstly, the distance between each node (say, node A) in a convex hull (say, convex hull A) is calculated from all the nodes in the remaining convex hulls and stored in a vector Distance[] as given in Eq. (6). Then, the nodes with minimum distance to node A in all convex hulls (other than convex hull A) are identified using equations (7) and Eq. (8). This process is repeated for every node present in all the convex hulls. Lines 7-22 in Algorithm 3 describe the steps for finding the nearest boundary pairs and saving them in Min_dist_nodes.

$$\begin{cases} \text{Distance}[l] = \text{dist}\left(\text{Part_bound}_{ij}[i][k],\ \text{Part_bound}_{ij}[j][l]\right) & \text{if } i \neq j \\ & i = j = 0 \text{ to } n_{part} - 1 \\ \text{Where} & k = 0 \text{ to } N_i - 1 \\ & l = 0 \text{ to } N_j - 1 \end{cases} \tag{6}$$

$$I_{min} = \text{argmin}(\text{Distance}[l]),\ l = 0 \text{ to } N_j - 1 \tag{7}$$

$$\begin{cases} \text{Min_dist_nodes}[b] = (\text{Part_bound}_{ij}, [i][k], \text{Part_bound}_{ij}[j][I_{min}], \text{min}(\text{Distance}[l]) \\ & b = 0 \text{ to total}_{bp} - 1 \\ \text{Where} & k = 0 \text{ to } N_i - 1 \\ & l = 0 \text{ to } N_j - 1 \end{cases} \tag{8}$$

Algorithm 3
Network Recovery.

1 Sink node receives $Part_{ij}$ and updated FNA from the UAV control center.
2 Sink starts the recovery process by computing the convex hull of every partition
3 for(i = 0 to n_{part}- 1)
4 Apply Graham scan algorithm to $Part_{ij}[i]$.
5 Save convex hull of $Part_{ij}[i]$ in $Part_bound_{ij}[i]$.
6 end for
7 Sink find outs the least distant boundary node pairs in different convex hulls
8 C = 0
9 for(i = 0 to n_{part} - 1)
10 for(j = 0 to n_{part} - 1)
11 if(i!=j)
12 for(k = 0 to number of nodes in $Part_bound_{ij}[i]$- 1)
13 for(l=0 to number of nodes in $Part_bound_{ij}[j]$ - 1)
14 Distance[l] = dist($Part_bound_{ij}$ [i][k], $Part_bound_{ij}$ [j][l]
15 end for
16 I_{min} = argmin(Distance[])
17 Min_dist_nodes[C] = ($Part_bound_{ij}[i][k]$, $Part_bound_{ij}[j][I_{min}]$, min(Distance[])
18 C = C + 1
19 end for
20 end if
21 end for
22 end for
23 Prepare a new network graph Adj_span from Min_dist_nodes.
24 Find MST and boundary node pairs (a_i,b_i)for Adj_span using Kruskal algorithm.
25 Compute the number of relays required between each node pair using Eq. (9).
26 Compute the total number of relays required to connect all node pairs (a_i,b_i) using Eq. (10).
27 The positions to deploy the relay nodes can be calculated using equations 11-14.
28 for(i=0 to n_{span} − 1)
29 for(j=0 to N_{rel}[i] − 1)
30 Compute locations u_j and v_j between a_i and b_i using equations 11-14.
31 Save (u_j,v_j) in RELAY table.
32 end for
33 end for
34 After that, the sink node stores the left and right neighbors of relay nodes in RELAY table.
35 Sink node attempts to reduce the total number of relay nodes.
36 for(i = 0 to N_{total} - 1)
37 for(j = 0 to N_{total}- 1)
38 if(i != j && dist(Relay[i], Relay[j]) < R)
39 if(dist(Relay[i], left_n[j]) < R && dist(Relay[i], right_n[j]) < R)
40 Remove Relay[j] from RELAY table.
41 end if
42 else if(Relay[j], left_n[i]) < R && dist(Relay[j], right_n[i]) < R)
43 Remove Relay[i]from RELAY table
44 end if
45 end if
46 end for
47 end for

Here N_i and N_j refers to the number of nodes in i^{th} and j^{th} convex hull, respectively and $total_{bp}$ denotes total boundary pairs. Min_dist_nodes is a two-dimensional vector that stores minimum distant boundary node pairs and the distance between them. The term 'dist' in equations refers to the euclidean distance between the nodes calculated using the formula given in equation 4.

Calculating the combinations of minimum distance boundary pairs of all the convex hulls in advance reduces the number of edges in the network graph Adj_span constructed by adding edges from Min_dist_nodes. UAV-NetRest computes Minimum Spanning Tree (MST) after preparing Adj_span. We have applied the Kruskal algorithm because it works better for the sparse graphs than the Prims algorithm [57]. The output of the MST algorithm gives the final list of node pairs (a_i,b_i) where new relay nodes will be deployed. The n_{span} and R represents the total node pairs and communication range of network and relay nodes. Lines 23-24 in Algorithm 3 represent the computation of MST Fig. 6. shows the minimum distance boundary pairs computed using the Kruskal algorithm for the sample network. The number of relay nodes that can be deployed at the edges between the node pairs (a_i, b_i)can be computed using Eq. (9), and Eq. (10) gives the total number of relay nodes (N_{total}). The N_{rel}[i] in Eq. (9) represents the number of relay nodes required to connect the i^{th} node pair.

$$N_{rel}[i] = dist(a_i, b_i)/R - 1 : 1 \le i \le n_{span} \tag{9}$$

$$N_{total} = \sum_{i=1}^{n_{span}} N_{rel}[i] \tag{10}$$

The positions to deploy these relay nodes between the node pairs (a_i, b_i) can be calculated using equations(11-14). In Eq. (12) and Eq. (13) (x, y) refers to the positional coordinates of the node pairs(a_i,b_i). Lines 25-33 in Algorithm 3 describe the steps to find the number of relay nodes and locations to deploy the relay nodes.

$$d_i = dist(a_i, b_i) : 1 \le i \le n_{span} \tag{11}$$

$$w_i = ((b_i \cdot x - a_i \cdot x) / d_i) : 1 \le i \le n_{span} \tag{12}$$

$$z_i = ((b_i \cdot y - a_i \cdot y) / d_i) : 1 \le i \le n_{span} \tag{13}$$

$$\begin{cases} u_j = a_i \cdot x + j * R * w_i \\ v_j = a_i \cdot y + j * R * z_i \\ RELAY_k = (u_j, v_j) \\ \text{where} \begin{cases} 1 \le i \le n_{span} \\ 1 \le j \le N_{rel}[i] \\ 1 \le k \le \sum_{i=1}^{n_{span}} N_{rel}[i] \end{cases} \end{cases} \tag{14}$$

The sink node also stores information about the neighbors of each relay node. As relay nodes can be deployed in a straight line, the sink stores information of two neighbors (left_n and right_n), between which the relay node can be deployed in the RELAY table as depicted in line 34 in Algorithm 3. These neighbors can be the network nodes or new relay nodes. The recovery algorithm further attempts to reduce the relay node count. This optimization step to reduce the relay node count depends on the physical proximity of relay nodes and their neighbors on both sides. Let us suppose that two relay nodes Relay 1 and Relay 2. The algorithm checks whether the distance between the Relay 1 and Relay 2 is less than their communication range (R). If this condition comes true, the algorithm checks the distance of Relay 1 and Relay 2 with both neighbors of Relay 2 and Relay 1, respectively. If the distance between Relay 1 and the neighbors of Relay 2 is less than the communication range, then Relay 2 can be removed from the relay nodes list. Similarly, if the previous condition comes out false, then the algorithm checks whether the neighbors of Relay 1 are in the communication range of Relay 2. If they are in the communication range, then Relay 1 can be removed from the list. If both the relay nodes are not in the communication range of other relay nodes' neighbors, they are required for the network recovery and cannot be removed. However, it is essential to note that reduction in the total count of relay nodes solely depends on their position in the network. This process is repeated for all the possible pairs of relay nodes in the RELAY table. Lines 35-47 in Algorithm 3 describe the process of reducing the relay nodes. Let P and B be the total number of partitions and boundary nodes for the time complexity calculations. The value of B is significantly less as compared to the total number of network nodes (n). The time complexity of the Graham Scan algorithm and the Kruskal algorithm is O(nlogn) and O(ElogB), where B and E denote the boundary nodes and the number of edges between the boundary nodes. The computation of the least distant boundary node pairs requires $O(P^2*B^2)$ time. The positional coordinates of the new relay nodes are computed in O(n) time. Let R be the computed relay node count, the step to reduce the relay node count requires $O(R^2)$ time. Therefore the complexity of the recovery algorithm is $O(nlogn + n + ElogB + (P^2*B^2) + R^2)$ which is equivalent to $O(nlogn + ElogB + (P^2*B^2) + R^2)$.

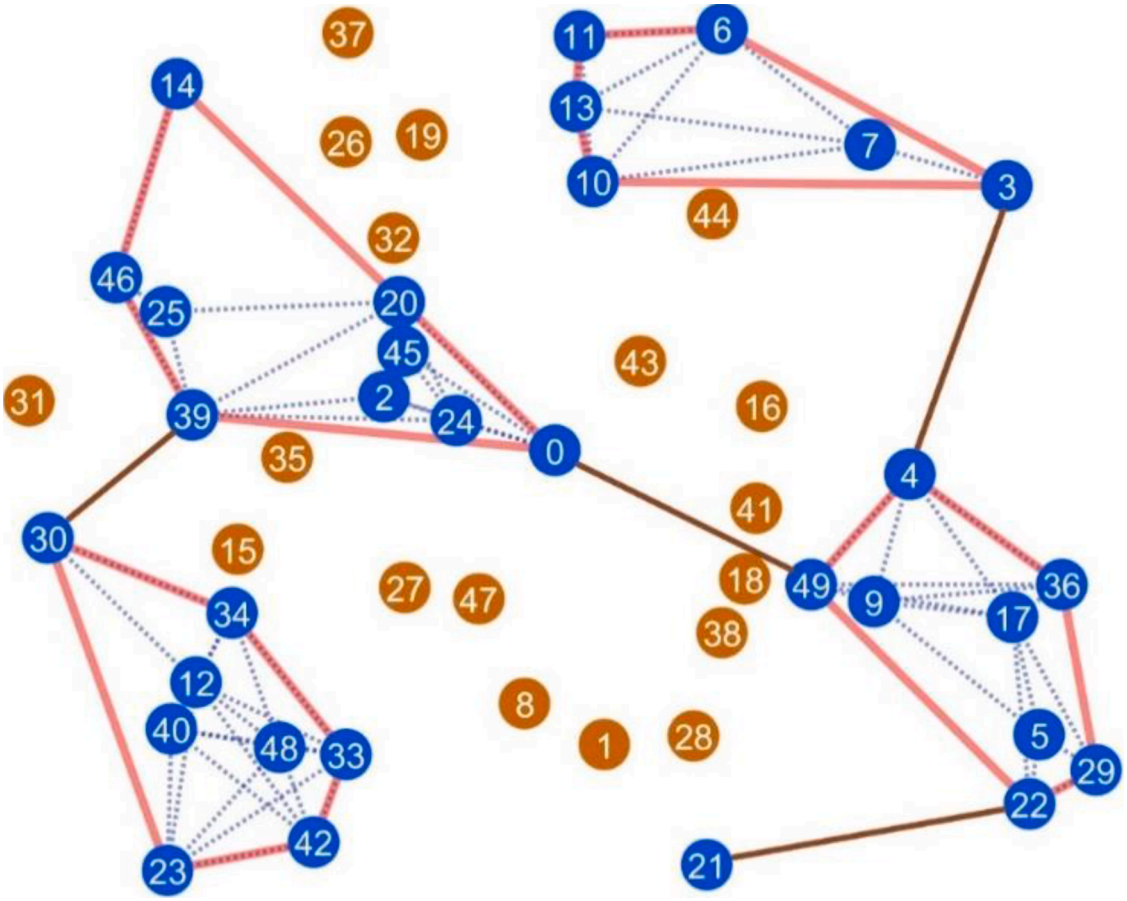


Fig. 6. Nearest Boundary Pairs of different Convex Hulls.

Algorithm 4
Optimized Relay Placement.

1 Sink compares the nodes in RELAY table with total count of failed nodes
2 if(count(RELAY) <= count(FNA))
3 Nodes_deploy = Deploy nodes at newly computed locations.
4 end if
5 else if(count(RELAY) > count(FNA))
6 Nodes_deploy = Deploy nodes at failed node locations.
7 end if
8 UAV control center starts recovery after taking instructions from sink.
9 if(Nodes_deploy <= n_uav_d)
10 Assign a UAV for every new node's location.
11 end if
12 else if(Nodes_deploy > n_uav_d)
13 k = n_uav_d
14 Apply k − means + + clustering for grouping Nodes_deploy.
15 Assign a cluster to each UAV.
16 end if
17 UAVs compute the least distant node in their respective cluster and move towards it.
18 On reaching that location, UAVs deploy a relay node.
19 UAVs repeat step 17 to find out the next location.
20 After deploying relay nodes, UAVs return to their location.
21 Network nodes, when they detect a new node, send an UPDATE message to the sink node.
22 On receiving this message, the sink node sends an ECHO message to the other surviving nodes.
23 Network nodes, when they receive the ECHO message, send their OHNL in a REPLY message.
24 if(count(REPLY messages) = Total network nodes – count(FNA))
25 Sink confirms network recovery
26 Sink updates NCT and Adj
27 end if

5.5. Relay Node Placement

After finalizing the number of relay nodes, the sink compares this count with the total sum of failed nodes. The number of nodes required to reconnect depends on the total number of partitions and their proximity. For example, the partitions created because of the failure of 9-10 nodes can be reconnected by employing 3-4 nodes using the proposed UAV-NetRest. Similarly, UAV-NetRest can compute 9-10 relay nodes required to reconnect the partitions created from the failure of 2-3 nodes. Therefore, this comparison is another optimization step that effectively prepares the proposed algorithm UAV-NetRest for small-scale and large-scale failures. Lines 1-7 in Algorithm 4 describe this comparison process. If the number of required relay nodes is less than or equal to the total count of failed nodes, the algorithm proceeds with deploying nodes at the previously computed locations from Eq. 14. Afterward, the sink communicates these locations with the UAV control center. On the other hand, if the count of failed nodes is less than the total relay nodes, then sink commands UAV control center to deploy relay nodes at locations of failed nodes saved in updated FNA after the Partition Detection phase. Hence, according to the failure scenario, UAV-NetRest proceeds with deploying the minimum number of relay nodes.

UAV control center takes the instructions from the sink node and proceeds with the final recovery process. The control center employs UAVs to deploy relay nodes at the given locations. As done in the partition detection process, it compares the number of UAVs and relay nodes. If relay nodes are less than or equal to the number of UAVs, assigning a relay node position to every UAV is easy. However, if the required relay nodes are more than the available UAVs, the node grouping according to their positional coordinates is required. Hence, similar to the partition detection phase, kmeans++ algorithm is

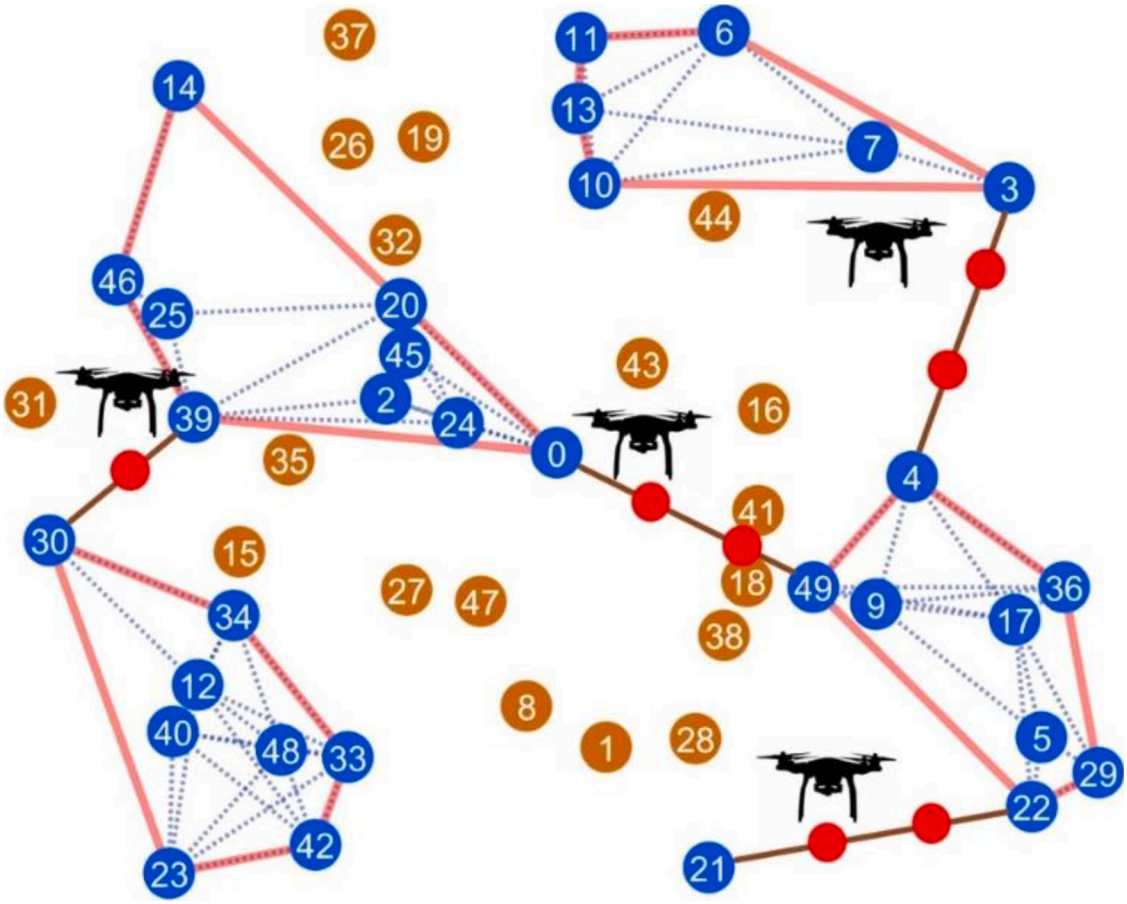


Fig. 7. UAVs deploy Sensor Nodes along the edges connecting Boundary Pairs.

Table 1
Simulation Parameters.

Parameter	Value
Simulation Area	400m × 400m - 1000m × 1000m
Nodes	10 - 100
Transmission Range	70m - 150m
Packet Size	512bytes
Mobility model of nodes	On-demand mobility
Wireless Channel	Wireless LAN IEEE 802.11b
Radio Medium	UnitDiskRadioMedium
Routing Protocol	Ad-hoc On-demand Distance Vector (AODV)
Mobility model of UAVs	Linear mobility
UAVs speed	$2ms^{-1}$ - $30ms^{-1}$
UAVs altitude	10m
Simulation Time	1000s-3000s

employed, which groups these nodes into clusters. The value ofk, in this case, is equal to the number of available UAVs. Lines 8-16 in Algorithm 4 represent the clustering process.

After computing the clusters, each cluster is assigned to an individual UAV. UAVs find the nearest position for deploying the relay node every time they move toward their given cluster locations. UAVs deploy relay nodes upon reaching their locations and proceed towards the next nearest location. After deploying the relay nodes, UAVs move back to their initial location. Lines 17-20 depict the node deployment process by UAVs Fig. 7. illustrates the node deployment process by multiple UAVs. Whenever the network nodes receive HEARTBEAT messages from the deployed relay nodes, they send their new one-hop neighbor list to the sink node. The sink node gets the new OHNL from some nodes, sends an ECHO message to the remaining nodes in the network, and requests them to send their OHNL. Whenever network nodes receive an ECHO message, they send a REPLY message consisting of their current OHNL to the sink. Sink confirms that the network is recovered when it gets the REPLY messages from all the nodes other than the failed ones. It updates NCT and Adj according to the new information obtained from the network nodes. Lines 21-27 represent the recovery confirmation process of the UAV-NetRest. The complexity of kmeans++ is O(kRI), where R and I refer to the relay node count and the number of iterations required by kmeans++ to converge. The network recovery confirmation requires O(n) time. The overall time complexity of the relay deployment algorithm is O(RkI + n).

The total time complexity of the proposed algorithm is $O(n) + O(n^2) + O(nlogn + ElogB + (P^2*B^2) + R^2) + O(RkI + n)$, which is equivalent to $O(n^2 + (P^2*B^2))$.

6. Simulation Results and Analysis

OMNeT++ 5.4.1 network simulator is used to test and analyze the performance of the proposed approach UAV-NetRest. UAV-NetRest is implemented on the Application layer of INET 4.0 project over OMNeT++ 5.4.1 Table 1. describes the parameters used for simulating the network. UAV-NetRest is tested for 10-100 nodes for different topologies in the simulation area ranging from 400m × 400m to 1000m × 1000m. The communication range of these network nodes varies between 70m-150m. The sink node is present in the center, and the UAV control center exists near the sink. The UAVs are either present in the center or outside the network area. We have tested our algorithm for both combinations. The performance of the proposed algorithm is evaluated based on the simulation settings described in Table 1 Fig. 8.-10 shows the simulation results of the proposed algorithm according to

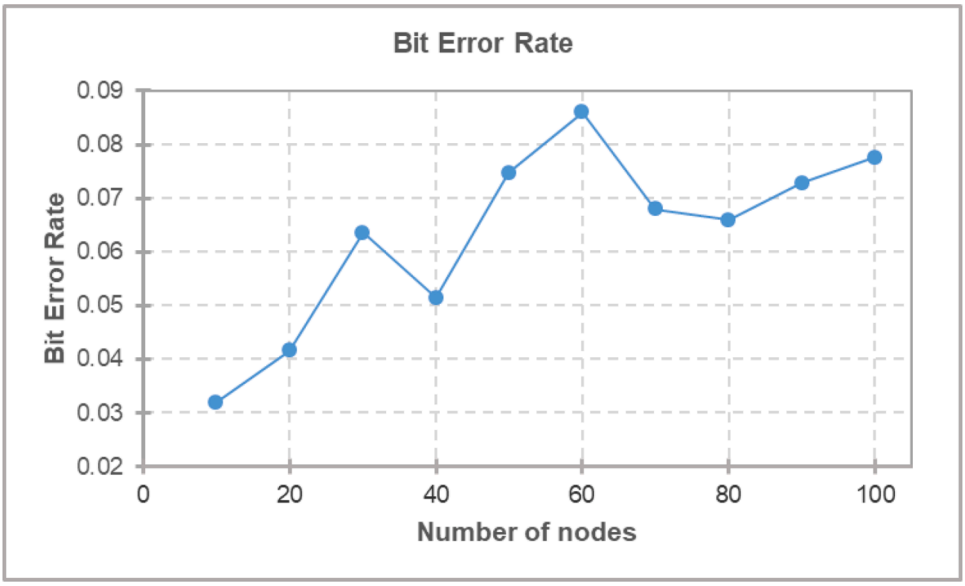


Fig. 8. BER w.r.t no of nodes.

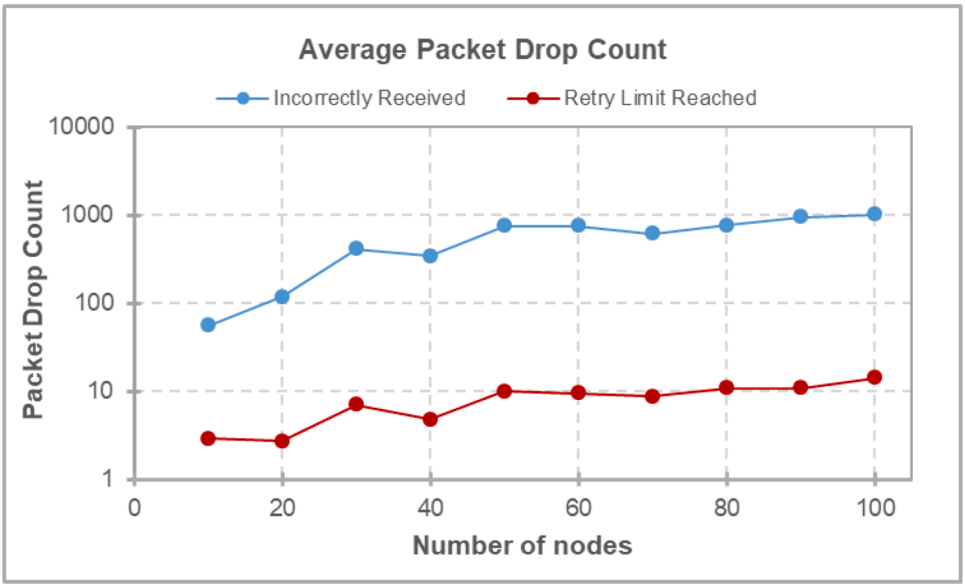


Fig. 9. Packet drop count w.r.t no of nodes.

the physical layer, data link layer, and network layer parameters Fig. 8. shows the Bit error rate (BER) results with respect to the number of nodes to quantify the reliability of UnitDiskRadioMedium. BER is the total erroneous bits received divided by the number of transmitted bits in a specific time interval. As shown in Fig. 8, the value of BER ranges from 0.031 to 0.085 Fig. 9. shows packet drop count at the link layer for reasons such as incorrectly received packets and retry limit reached. The packet drop count for incorrectly received packets is more than the packet drop count for the retry limit reached. The figure shows that the packet drop count increases with more nodes Fig. 10. illustrates end-to-end delay, i.e., the transmission time taken by the packet across the network to reach from the source to destination. The end-to-end delay entirely depends on the position of nodes in the network and the failure scenario. As shown in Fig. 10, the value of end-to-end delay ranges from 194ms to 601ms. The proposed algorithm uses IEEE 802.11b mac protocol for communication [25,58,59]. Hence, Fig. 11 analyses the average energy consumed by UAVs while communicating with the ground nodes. The energy consumed by UAVs ranges from -1.45J to -6.12J.

6.1. Comparison with Existing Approaches

The UAV-NetRest is compared with three existing techniques Basic Deployment (BD) [33], Center Deployment with Convex Hull (CDCH) [33], and Cluster-based Relay Placement Algorithm (CRP) [10], to evaluate its efficacy Fig. 12.(a) shows an example network, and Fig. 12 (b) shows the network after multiple node failures Fig. 13.-14 shows the deployment example to compare the recovery procedure of existing techniques with the proposed algorithm based on the failure scenario shown in Fig. 12 (b). These approaches lack partition detection and are unable to differentiate between fundamental node failures and network partitioning. Therefore, we have compared their recovery approach with the recovery algorithm of UAV-NetRest.

The utility of all these approaches has been assessed regarding the

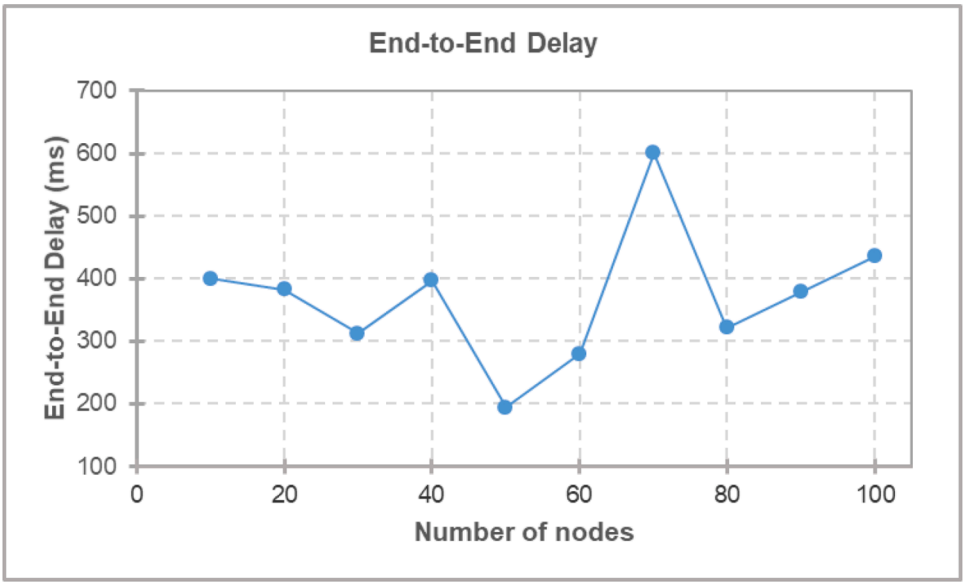


Fig. 10. End-to-End delay w.r.t no of nodes.

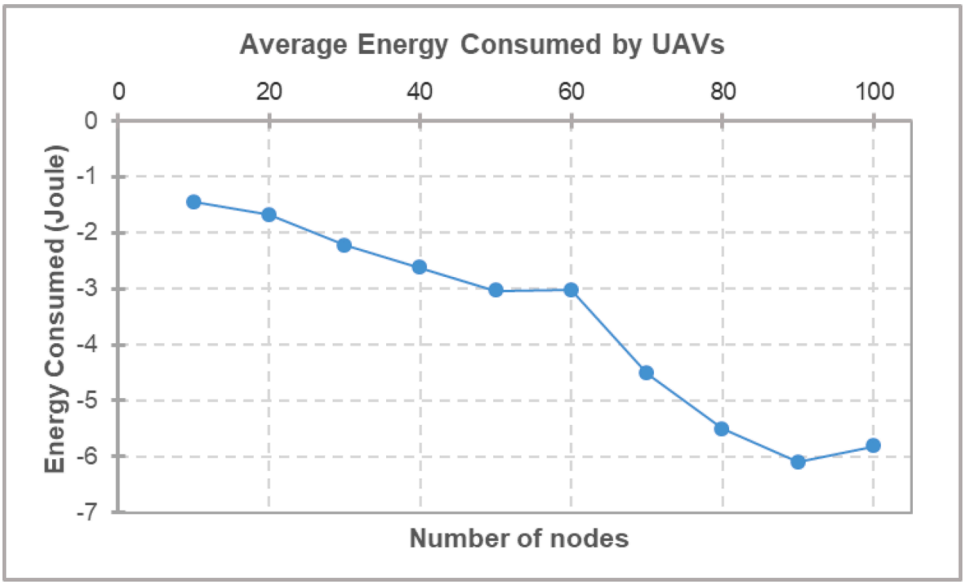


Fig. 11. UAVs' Energy Consumption w.r.t no of nodes.

number of deployed relay nodes, network recovery time, the number of messages transmitted, and the total distance traveled by UAVs or relay nodes during recovery. A brief discussion of these techniques is presented below:

Basic Deployment (BD) [33]**:** This approach identifies the outer partitions by applying the Graham scan convex hull algorithm. The relay nodes are deployed in the identified convex hull in a circular fashion. Afterward, the algorithm employs new relay nodes to connect the inner segments towards the nearest deployed relay node Fig 13.(a) shows that the BD algorithm requires 11 extra relay nodes for network recovery. It is assumed that relay nodes are present nearby the sink node and move towards the nearest unoccupied locations.

Center Deployment with Convex Hull (CDCH) [33]**:** The algorithm identifies outer segments by applying a convex hull algorithm and selects a representative node from each outer segment. The relay nodes are deployed from these representative nodes towards the center of deployment (CoD). After that, the representative nodes from the inner segments are computed and connected to the nearest deployed node by engaging more relay nodes Fig 13.(b) shows that the CDCH algorithm requires 11 extra relay nodes for network recovery.

Cluster-based Relay Placement (CRP) [10]: This approach finds out the nearest partitions by comparing the centroids of the partitions. After finding out the two least distant centroids, the algorithm applies a convex hull algorithm to identify the boundary of the selected partitions and connects the minimum distant boundary nodes by deploying additional relay nodes. After connecting two partitions, new centroids are computed, and this process is repeated until the total partition count becomes one. CRP requires 7 extra relay nodes for network recovery, as shown in Fig. 14(b).

The recovery algorithm of the proposed UAV-NetRest is similar to the CRP algorithm because both algorithms focus on finding minimum distant boundary pairs. However, both algorithms have different procedures for failure detection and identification of minimum distant boundary node pairs in different partitions. The CRP works when one or

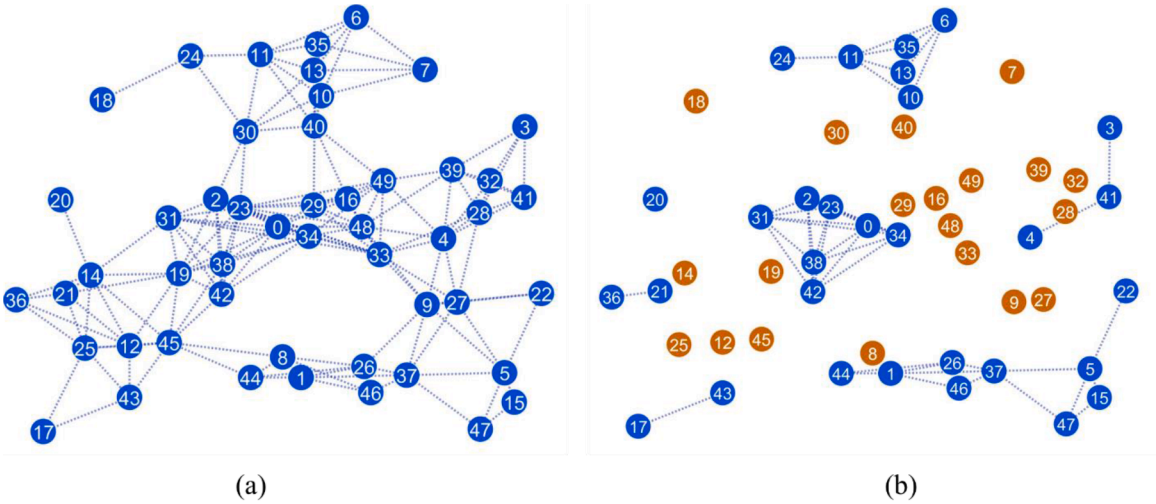


Fig. 12. Example Connected Network.

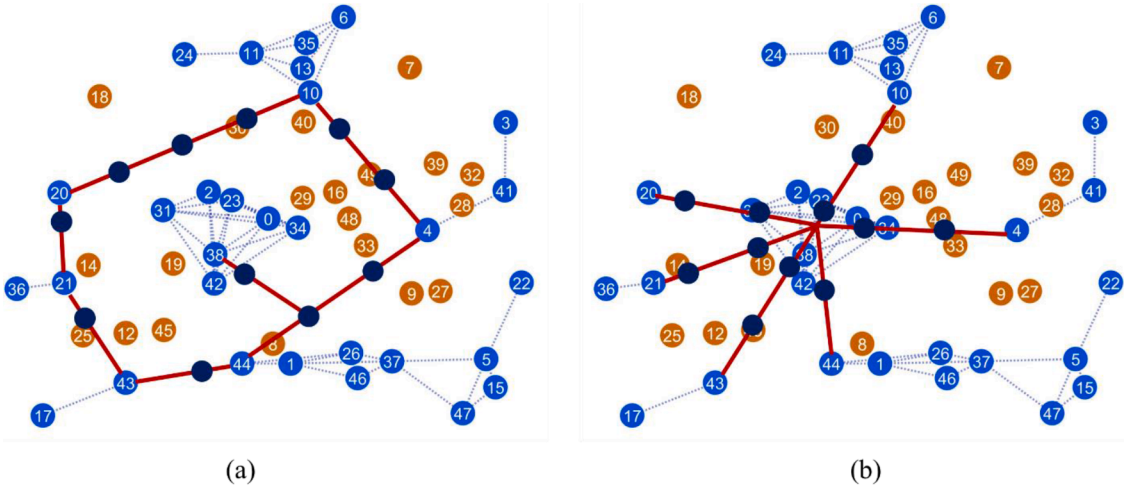


Fig. 13. Deployment example of (a) BD (b) CDCH.

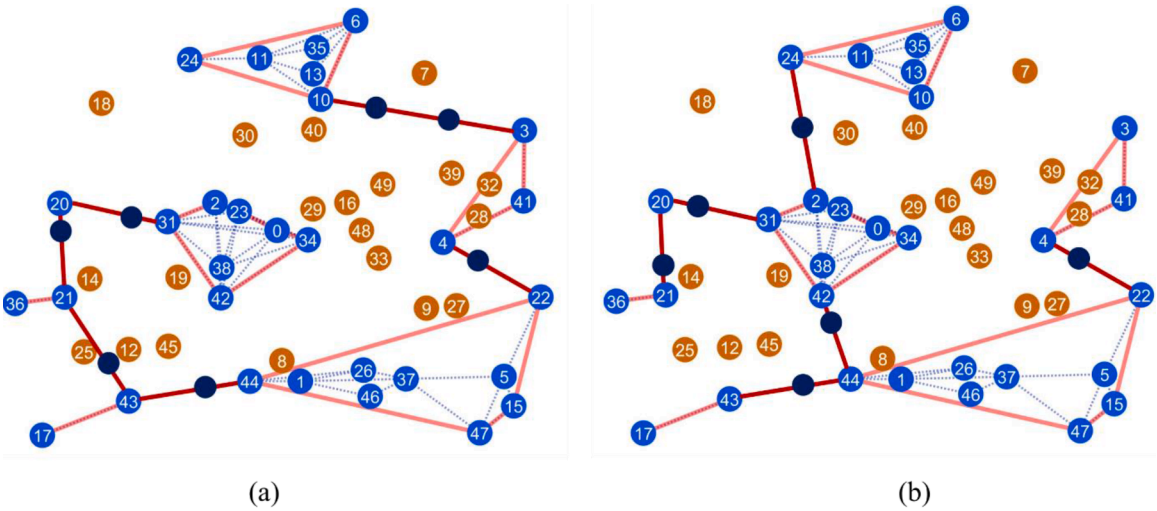


Fig. 14. Deployment example of (a) CRP (b) UAV-NetRest.

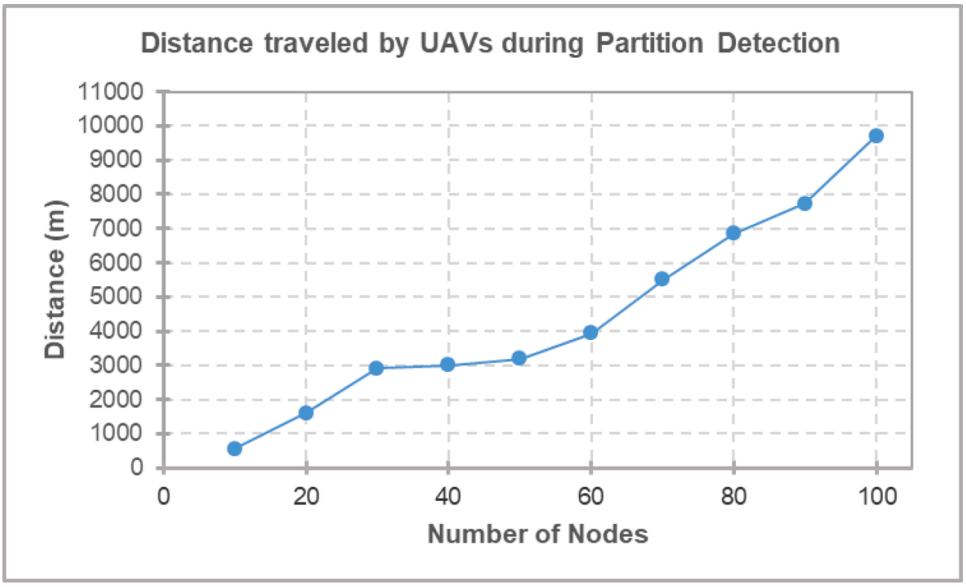


Fig. 15. DT w.r.t. no of nodes during detection.

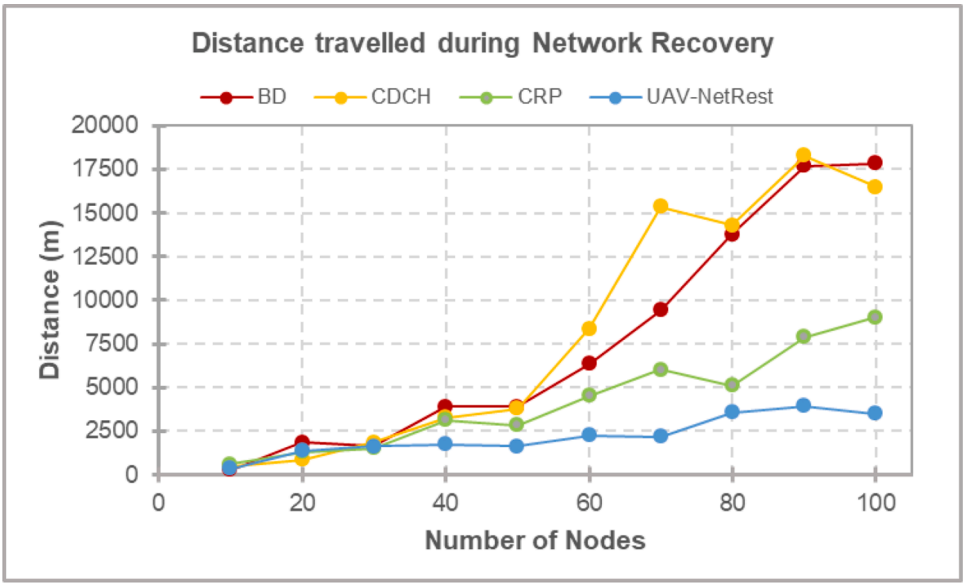


Fig. 16. DT w.r.t no of nodes during recovery.

more cut-vertex nodes fail and uses centroids of the partitions to identify nearest boundary node pairs. However, the UAV-NetRest can recover the network for both cut-vertex and non-cut-vertex sensor node failures. UAV-NetRest identifies minimum distant boundary node pairs with the help of the Kruskal minimum spanning tree algorithm Fig 14.(b) shows that the UAV-NetRest requires 6 extra relay nodes for network recovery. In order to apply BD, CDCH, and CRP algorithms in our contemplated network recovery problem, the relay nodes are placed near the sink node initially and move towards the computed deployment location during the recovery process. It is assumed that enough relay nodes are present in the network for connectivity restoration.

UAV-NetRest works in different phases, as described earlier. Initially, the sink accumulates all the necessary information about the network nodes. Whenever network nodes detect missing HEARTBEATS from their neighbors, they send UPDATE messages to the sink node. The sink node, on receiving these UPDATE messages, again collects the information from the network nodes. The sink constructs the failed node data from the updated network information and shares all this information with the UAV control center. UAV control center employs UAVs to collect complete information about the network. After gathering information about the network, UAVs share the network status with the control center. The UAV control center determines whether the network is partitioned or not and sends the result to the sink. If the number of partitions in the network exceeds one, then the sink starts the recovery procedure. After computing the required relay nodes for recovery, the sink provides the locations for deployment to the UAV control center. UAV control center sends UAVs to deploy sensor nodes at given locations.

Maximum connectivity restoration approaches focus on minimization of (i) total travel distance, (ii) largest individual tour length, (iii) deployed node count, (iv) average tour length, (v) hop count, (vi) average tour length, (vii) energy consumed, (viii) number of messages, and maximization of (i) average node degree, (ii) coverage. Our proposed work mainly focused on the involvement of an aircraft unit for

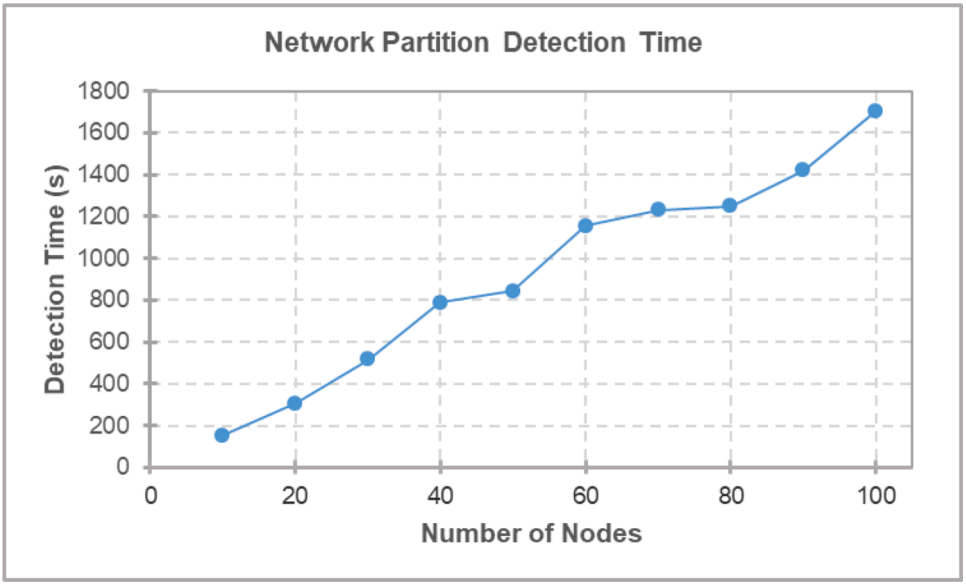


Fig. 17. NPDT w.r.t. number of nodes Fig. 18 NPDT w.r.t. velocity of UAVs.

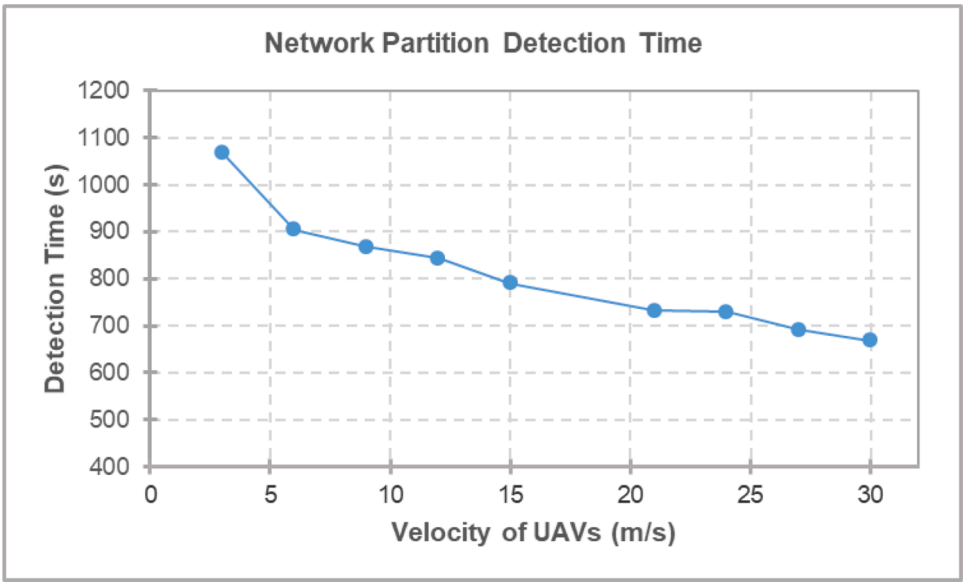


Fig. 18. NPDT w.r.t. velocity of UAVs.

partition detection and recovery. UAVs travel to different locations and wait for some intervals while detecting failed nodes or deploying additional nodes for recovery. Therefore, the main objectives of the UAV-NetRest are minimizing the (i) Distance traveled by UAVs, (ii) Network Partition Detection Time (NPDT), (iii) number of messages, (iv) number of deployed relays, and (v) Network Recovery Time (NRT). The details of the simulation results based on these parameters are given below:

1. Distance Traveled (DT): This parameter analyzes the distance traveled by UAVs in the detection and deployment phase. In both phases, UAV-NetRest works on reducing the distance traveled by UAVs in two steps. In the first step, UAVs compute the minimum distant node before moving in their assigned clusters and then proceed towards that node. When UAVs reach the location of that nodes, they check whether the node is failed or not. If some node is active, UAVs compare the failed node list with the OHNL of that node. If the UAV finds some failed node in the OHNL, it removes it from the failed node list and does not move toward its location. UAVs repeat the first step in detection and deployment phases before moving towards the next node's location. The second step is repeated only in the detection phase. In this paper, the total distance traveled by three UAVs in both detection and recovery phases is computed by varying the number of nodes from 10 to 100 Fig. 15. shows the distance traveled by UAVs w.r.t. number of nodes in the partition detection phase, and Fig. 16 represents the distance traveled by UAVs w.r.t. number of nodes during recovery Fig. 15. shows that the distance traveled by UAVs during the detection phase is very high since UAVs traverse many failed node locations because the sink node does not have information about the nodes in other partitions and considers them as failed nodes Fig. 16. compares the distance traveled by UAVs in UAV-NetRest with the distance traveled by relay nodes in BD, CDCH, and CRP algorithms. BD and CDCH have higher values for distance traveled as compared to the other two approaches. It is clear from the figure that the distance traveled by UAVs in UAV-NetRest is minimum. Since UAVs move towards the minimum distant location every time.

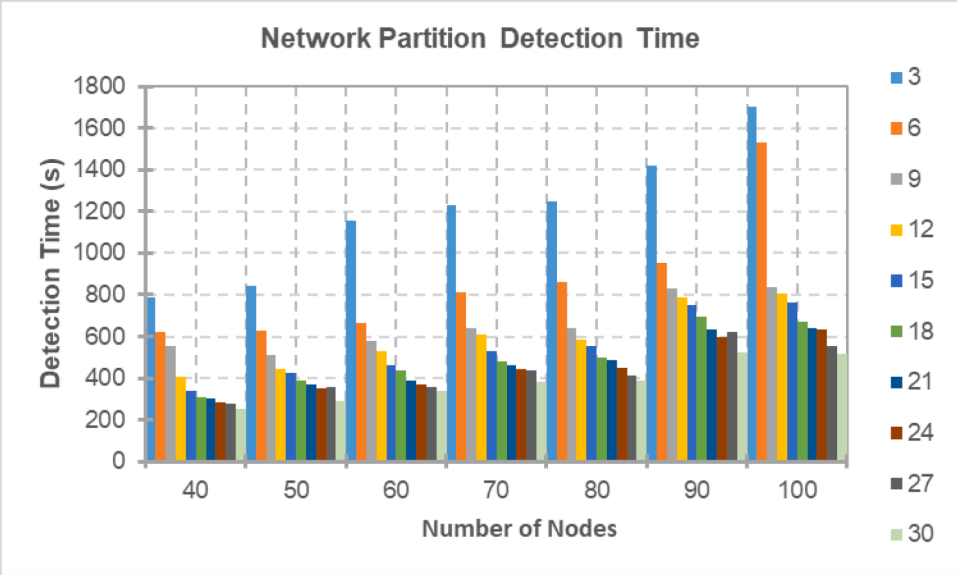


Fig. 19. NPDT w.r.t. number of nodes with UAVs varying from 3 to 30.

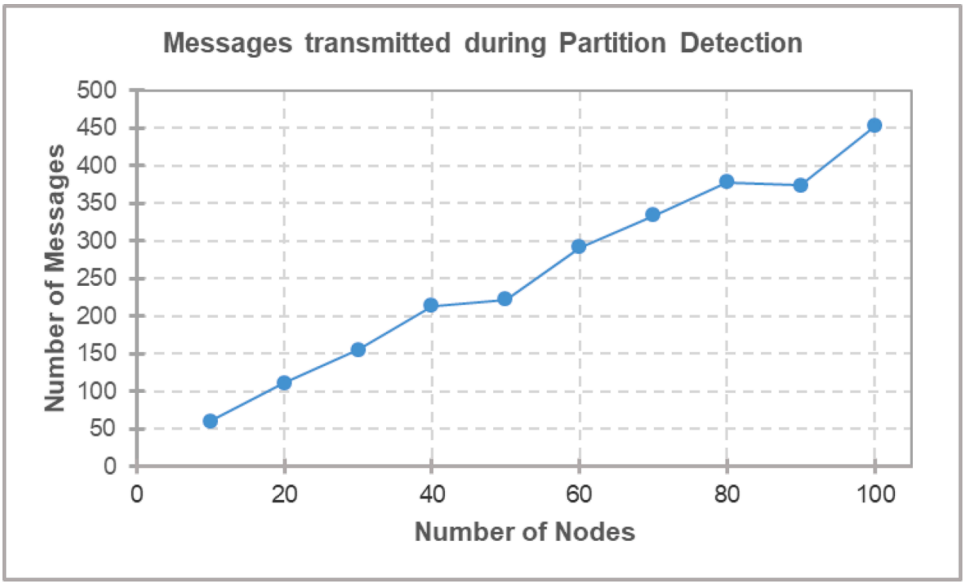


Fig. 20. MT w.r.t no. of nodes during detection.

The DT results in Fig. 15 entirely depend on the number of nodes that the sink considers as failed and the distance between these nodes Fig. 16. shows the distance traveled by UAVs during recovery. It depends on the number of nodes that UAVs deploy and their deployment locations. Therefore, in some cases, the DT does not increase with the number of nodes, such as distance traveled by UAVs for 90 nodes is more than 100 nodes for the UAV-NetRest algorithm in Fig. 16. BD, CDCH, and CRP algorithms also show these variations. Mainly in both Fig. 15 and Fig. 16, the distance traveled by UAVs or relay nodes grows with an increase in the number of network nodes

2. Network Partition Detection Time (NPDT): Network Partition Detection is a fundamental step before recovery. Almost all proposed approaches assume that the information about the partitions is already available. In some works, it is assumed that aircrafts such as drones, robots, mini-robots, etc., collect all the information about the network [14,15]. Some distributed solutions are also proposed for partition detection[10,11]. However, these solutions can not differentiate between multiple node failures and partition detection. Our work is not based on any assumptions. Therefore, we have included NPDT as a performance evaluation parameter for the proposed solution. It is the time between the multiple node failures to the partition detection by the sink node with the help of UAVs. The NPDT is tested based on three parameters. The first one is NPDT w.r.t number of nodes in the network, keeping the number of UAVs constant. The number of UAVs employed for simulating this scenario is three, and nodes vary from 10-100 Fig. 17. depicts the NPDT performance w.r.t number of nodes. This plot illustrates that detection time increases with more nodes. However, it also depends on the number of failed nodes, waiting time interval of UAVs, and the time interval between which UAVs return to their initial location and share their adjacency lists with the UAV control center. The second one is NPDT w.r.t. the velocity of UAVs. The velocity of UAVs varies from 5m/s to 30m/s Fig. 18. shows the results obtained for NPDT w.r.t. velocity of UAVs. NPDT is inversely proportional to the velocity of UAVs.

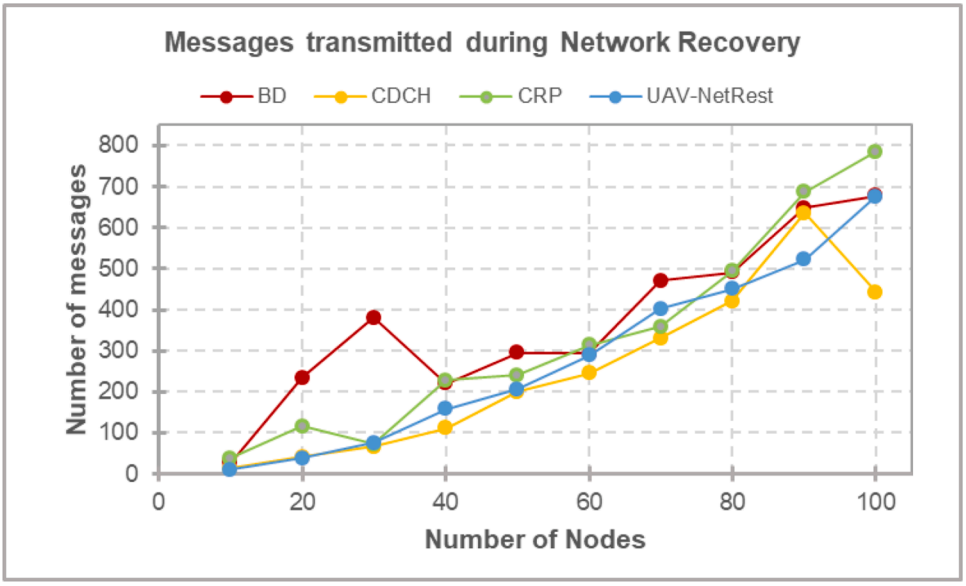


Fig. 21. MT w.r.t no. of nodes during recovery.

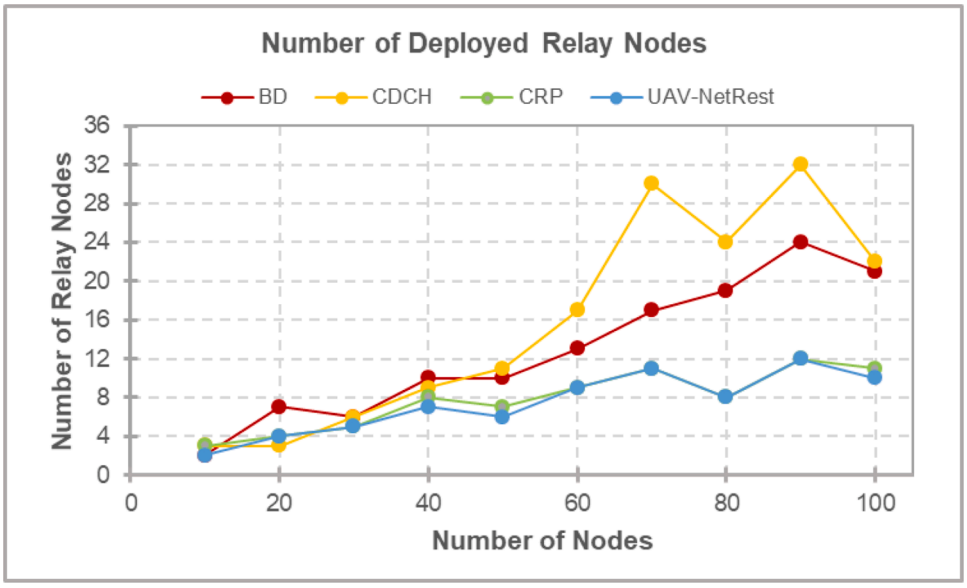


Fig. 22. Number of relays w.r.t number of nodes.

The third scenario is NPDT w.r.t. number of nodes by varying the number of UAVs. The results are collected for different topologies of a varying number of nodes between 40-100 nodes. The failed nodes are fixed to 30%-35% of network nodes while increasing or decreasing the UAV count. UAV-NetRest automatically adjusts the UAV's count while employing them in the detection process when they are less than or greater than the number of failed nodes. The number of UAVs for simulating this scenario varies from 3 to 30 Fig. 19. shows the results for NPDT w.r.t number of nodes by varying UAV's count. This graph clearly illustrates that NPDT decreases with an increase in the number of UAVs. NPDT also depends on the time variations when UAVs come back to their initial position, and the UAV control center accumulates the adjacency lists from all UAVs. Therefore, the reduction in NPDT is less after 12 or 15 UAVs. When the number of UAVs grows, the UAV control center utilizes more time in merging their adjacency lists. The velocity of UAVs is 10m/s for the results obtained in Figs. 17 and 19.

3. Messages transmitted for Detection and Recovery (MTDR): MTDR represents total messages transmitted between the intervals when a node detects two missing HEARTBEATS messages from other network nodes and sends UPDATE messages to the sink node. The sink node accordingly takes help from the UAV control center for detection and recovery and confirms the network recovery finally Fig. 20. shows the number of messages transmitted (MT) during the detection and recovery process for different topologies of 10-100 nodes. The plot shows that the number of messages increases for more nodes, with some exceptions. The number of messages transmitted for 80 nodes is more than the number of messages transmitted for 90 nodes. One major factor affecting the increase or decrease in transmitted messages is the OHNL of failed nodes and the number of failed nodes. If some failed nodes have more active neighbors, then the UPDATE messages for these nodes can be more. Likewise, UPDATE messages can decrease if failed nodes have less active neighbors Fig. 21. shows the comparison results for the number of messages transmitted (MT) during network recovery in BD, CDCH, CRP, and UAV-NetRest approaches. The performance of the CDCH, CRP, and

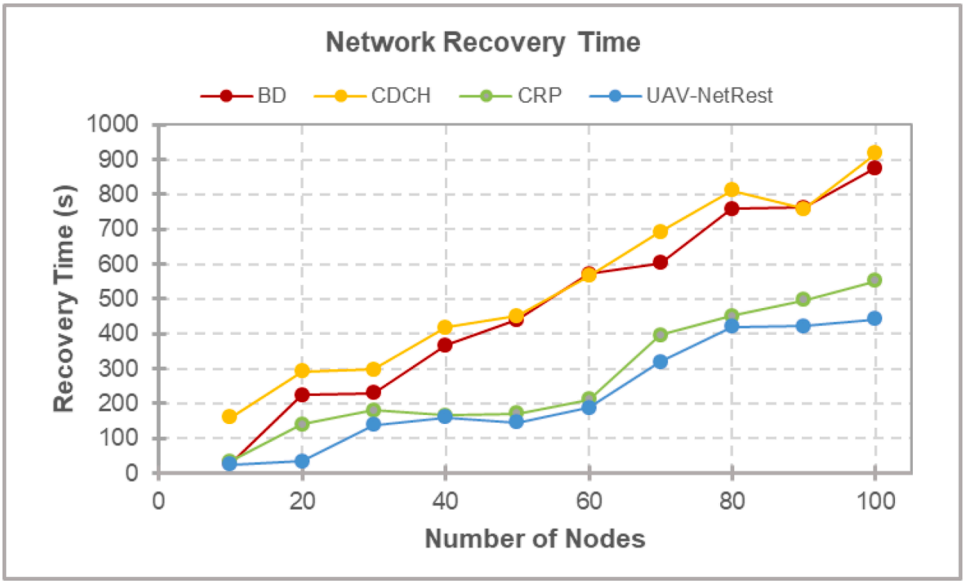


Fig. 23. Recovery time w.r.t number of nodes.

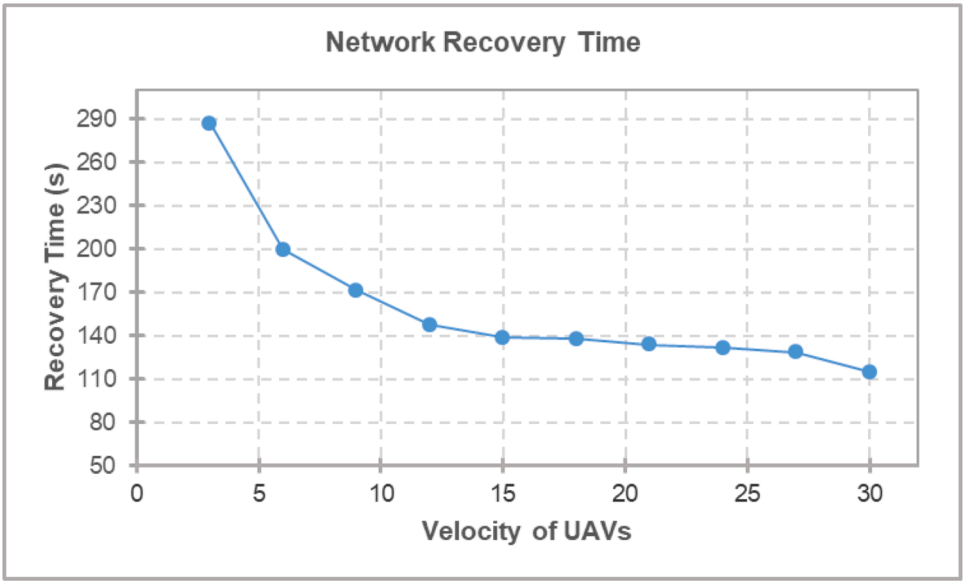


Fig. 24. Recovery time w.r.t velocity of UAVs.

UAV-NetRest algorithms is similar, with slight variations. Initially, the BD algorithm shows higher values for the number of messages transmitted. However, its performance improves as the network size grows.

4. The number of deployed relays: The relay nodes are deployed on the minimum distant boundary pair nodes resulting from the spanning tree. UAV-NetRest also tries to minimize further the number of relay nodes, which entirely depends on the physical proximity and topology of the network and deployed nodes Fig. 22. shows the number of deployed relays variations by varying nodes from 10 to 100. The plotted graph shows that the number of relay nodes increases with an increase in network nodes. However, it ultimately depends on the number of partitions created, the distance between the nearest boundary pairs of these partitions, and whether the optimizations steps for reducing the number of relay nodes are applicable or not. The number of deployed relay nodes in UAV-NetRest is compared with the relay node count in BD, CDCH, and CRP algorithms. In this case, UAV-NetRest and CRP are the best performers as compared to the other two techniques. CDCH algorithm requires the maximum number of relay nodes for the same network size.

5. Network Recovery Time (NRT): The paper is focused on network partition detection and recovery. Hence, recovery time is considered an important parameter to be evaluated here. It is the time after the partition detection phase. Between these time intervals, the sink recovers the network by employing UAVs to deploy nodes at different locations and then confirms the network recovery. Similar to NPDT, NRT is tested w.r.t. the number of nodes, the number of nodes by varying UAV's count, and velocity of UAVs. In the first scenario, simulations are performed for different network topologies by varying nodes from 10 to 100. In this scenario, three UAVs are employed to deploy relay nodes, and the velocity of these UAVs is 10m/s Fig. 23. represents the results obtained for NRT w.r.t. number of nodes. NRT depends on the number of deployed nodes and their locations, the time interval after which nodes start sending messages, and time variations of the network nodes when they send recovery updates and REPLY messages to the sink node. Therefore, the results depicted in Fig. 23 show high values of NRT for

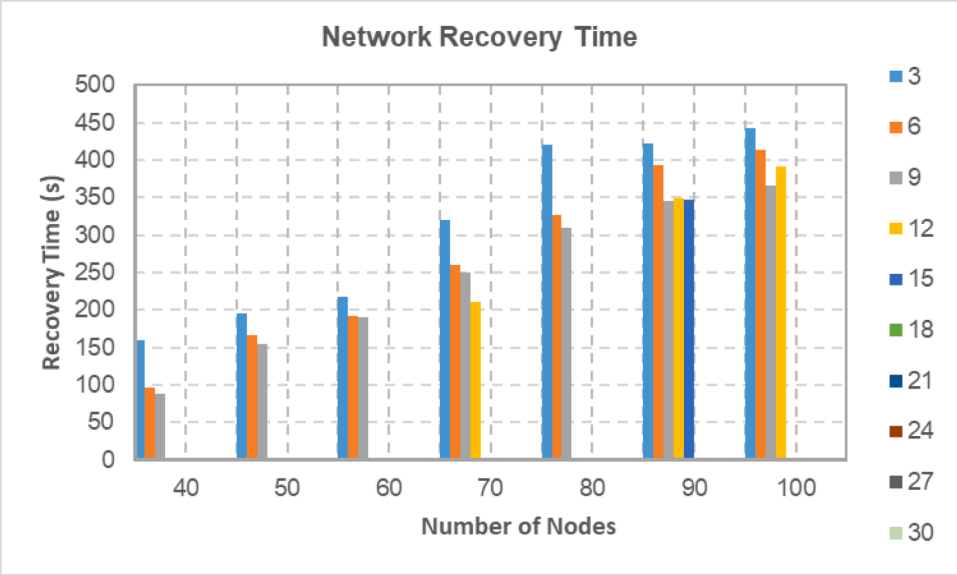


Fig. 25. NRT w.r.t number of nodes with UAVs varying from 3 to 30.

fewer nodes, such as NRT for 40 nodes is more than NRT for 50 nodes, respectively. The NRT performance of the proposed UAV-NetRest is compared with BD, CDCH, and CRP algorithms, and UAV-NetRest shows the best results. CRP shows second-best results, and the NRT performance for BD and CDCH is similar Fig. 23. shows that the value of NRT increases as the network size grows for all four algorithms, with some deviations that happen because of various factors such as the number of deployed relay nodes and their locations. The second scenario tests the NRT w.r.t velocity of UAVs Fig. 24. shows the results of NRT w.r.t velocity of UAVs. Similar to the detection time in Fig. 18, Recovery time is inversely proportional to the velocity of UAVs.

In the third scenario, NRT is computed w.r.t. number of nodes by varying the UAV's count. The number of UAVs varies from 3 to 30. However, similar to the detection phase UAV-NetRest automatically adjusts the UAVs required for the relay node deployment from the given number of UAVs Fig. 25. shows the results of NRT w.r.t. number of nodes by varying the UAV's count from 3 to 30. The velocity of UAVs is 10m/s for this scenario. It can be observed from the plot that less number of UAVs are utilized in the relay deployment phase and recovery time reduces with the increase in the number of UAVs. The recovery time also depends on when UAVs return to their initial positions and network nodes send recovery reports to the sink node. Hence in some cases, recovery time can show some deviations, such as the value of NRT for 100 nodes is more for 12 UAVs than 9 UAVs.

Therefore, from the obtained results, it can be concluded that network partitioning is a very unpredictable event in an unattended network. The algorithm's performance depends on the number of partitions, the distance between the partitions, the number of damaged and active nodes, and the number of available UAVs.

7. Conclusion

Incorporating UAVs to lessen the limitations of WSNs comes up with more challenges. In this paper, we focused on engaging UAVs in the partition detection and recovery process. This paper divides the network partition recovery process into five phases: Network configuration, failure detection, network partition detection, network recovery, and optimized relay placement. The proposed algorithm covers all aspects of network partition recovery and employs UAVs to detect network partitions and deploy relay nodes to reconnect these partitions. The simulation results are obtained for detection and recovery time, distance traveled by UAVs during partition detection and recovery, the number of messages transmitted, and the number of deployed relay nodes. The simulation results for these parameters are obtained for different randomly generated WSN topologies. These results entirely depend on the network state after network partitioning and the number of UAVs employed. The performance of the UAV-NetRest's recovery algorithm is compared with three existing recovery techniques. The simulation results demonstrate that the UAV-NetRest recovers the network with significantly less recovery time and distance traveled by UAVs. The algorithm deploys minimum relay nodes as compared to other techniques. The performance of UAV-NetRest in terms of messages transmitted during recovery is similar to the other three approaches.

UAVs assisted partition detection and recovery requires optimal path planning, obstacle avoidance mechanism, data gathering by UAVs, and optimized area exploration techniques for UAVs. The obstacles can be objects in the environment or other moving aircraft. The stringent energy limitations of UAVs can also impact the detection and recovery process. In the future, we aspire to focus on research problems related to employing UAVs in the network domain. Further research can be carried out to understand the techniques for collecting and deploying sensor nodes by UAVs. Another study area can be designing an efficient coordination and communication mechanism between UAVs and ground nodes to reduce the UAVs' effort in the detection phase. Furthermore, restoring k-connectivity in the partitioned network is also an interesting optimization problem and research direction.

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.

Acknowledgment

We are grateful to the Science and Engineering Research Board (SERB) - Department of Science and Technology (DST), Government of India, for sponsoring this research project.

References

[1] Y.K. Joshi, M. Younis, Restoring connectivity in a resource constrained WSN, J. Netw. Comput. Appl. 66 (2016) 151–165, https://doi.org/10.1016/j.jnca.2016.03.009.

[2] M. Younis, I.F. Senturk, K. Akkaya, S. Lee, F. Senel, Topology management techniques for tolerating node failures in wireless sensor networks: A survey, Comput. Networks. 58 (2014) 254–283, https://doi.org/10.1016/j.comnet.2013.08.021.

[3] V. Ranga, M. Dave, A.K. Verma, Network partitioning recovery mechanisms in WSANs: A survey, Wirel. Pers. Commun. 72 (2013) 857–917, https://doi.org/10.1007/s11277-013-1046-7.

[4] S. Lee, M. Younis, M. Lee, Connectivity restoration in a partitioned wireless sensor network with assured fault tolerance, Ad Hoc Networks 24 (2015) 1–19, https://doi.org/10.1016/j.adhoc.2014.07.012.

[5] S. Chouikhi, I. El Korbi, Y. Ghamri-Doudane, L. Azouz Saidane, Centralized connectivity restoration in multichannel wireless sensor networks, J. Netw. Comput. Appl. 83 (2017) 111–123, https://doi.org/10.1016/j.jnca.2017.01.040.

[6] Y.K. Joshi, M. Younis, Exploiting skeletonization to restore connectivity in a wireless sensor network, Comput. Commun. 75 (2016) 97–107.

[7] Y.K. Joshi, M. Younis, Mobility-based internetworking of disjoint segments. 2014 27th Bienn. Symp. Commun., IEEE, 2014, pp. 193–197, https://doi.org/10.1109/QBSC.2014.6841212.

[8] U. Baroudi, M. Aldarwbi, M. Younis, Energy-Aware Connectivity Restoration Mechanism for Cyber-Physical Systems of Networked Sensors and Robots, IEEE Syst. J. 14 (2020) 3093–3104, https://doi.org/10.1109/JSYST.2020.2970649.

[9] M. Hauspie, J. Carle, D. Simplot, M. Hauspie, J. Carle, D. Simplot, P. Detection, A. N. Using, J. Carle, D. Simplot, Partition Detection in Mobile Ad-hoc Networks Using Multiple Disjoint Paths Set To cite this version : HAL Id : hal-00616860 Partition Detection in Mobile Ad-Hoc Networks Using Multiple Disjoint Paths Set, (2011).

[10] R. Kumar, T. Amgoth, Adaptive cluster-based relay-node placement for disjoint wireless sensor networks, Wirel. Networks. 26 (2020) 651–666, https://doi.org/10.1007/s11276-019-02171-8.

[11] S. Shriwastav, D. Ghose, Round-table negotiation for fast restoration of connectivity in partitioned wireless sensor networks, Ad Hoc Networks 77 (2018) 11–27, https://doi.org/10.1016/j.adhoc.2018.04.008.

[12] K. Akkaya, A. Thimmapuram, F. Senel, S. Uludag, A. Thimrnapuram, F. Senel, S. Uludag, Distributed recovery of actor failures in wireless sensor and actor networks, in: IEEE Wirel. Commun. Netw. Conf. WCNC, IEEE, 2008, pp. 2480–2485, https://doi.org/10.1109/wcnc.2008.436.

[13] F. Senel, M.F. Younis, K. Akkaya, Bio-inspired relay node placement heuristics for repairing damaged wireless sensor networks, IEEE Trans. Veh. Technol. 60 (2011) 1835–1848.

[14] I. Senturk, S. Yilmaz, K. Akkaya, Connectivity restoration in delay-tolerant sensor networks using game theory, Int. J. Ad Hoc Ubiquitous Comput. 11 (2012) 109–124, https://doi.org/10.1504/IJAHUC.2012.050268.

[15] W. Lalouani, M. Younis, N. Badache, Optimized repair of a partitioned network topology, Comput. Networks. 128 (2017) 63–77, https://doi.org/10.1016/j.comnet.2017.02.003.

[16] F. Senel, M. Younis, Novel relay node placement algorithms for establishing connected topologies, J. Netw. Comput. Appl. 70 (2016) 114–130, https://doi.org/10.1016/j.jnca.2016.04.025.

[17] F. Senel, M. Younis, Relay node placement in structurally damaged wireless sensor networks via triangular steiner tree approximation, Comput. Commun. 34 (2011) 1932–1941, https://doi.org/10.1016/j.comcom.2011.05.010.

[18] S. Lee, M. Younis, Recovery from multiple simultaneous failures in wireless sensor networks using minimum Steiner tree, J. Parallel Distrib. Comput. 70 (2010) 525–536.

[19] S. Lee, M. Younis, Optimized relay node placement for connecting disjoint wireless sensor networks, Comput. Networks. 56 (2012) 2788–2804.

[20] G.H. Lin, G. Xue, Steiner tree problem with minimum number of Steiner points and bounded edge-length, Inf. Process. Lett. 69 (1999) 53–57, https://doi.org/10.1016/s0020-0190(98)00201-4.

[21] R. Kumar, T. Amgoth, D. Das, Obstacle-Aware Connectivity Establishment in Wireless Sensor Networks, IEEE Sens. J. 21 (2021) 5543–5552, https://doi.org/10.1109/JSEN.2020.3032144.

[22] W. Lalouani, M. Younis, N. Badache, Interconnecting isolated network segments through intermittent links, J. Netw. Comput. Appl. 108 (2018) 53–63, https://doi.org/10.1016/j.jnca.2018.02.010.

[23] B. Alzahrani, O.S. Oubbati, A. Barnawi, M. Atiquzzaman, D. Alghazzawi, UAV assistance paradigm: State-of-the-art in applications and challenges, J. Netw. Comput. Appl. 166 (2020), 102706, https://doi.org/10.1016/j.jnca.2020.102706.

[24] S. Zaidi, M. Atiquzzaman, C.T. Calafate, Internet of Flying Things (IoFT): A Survey, Comput. Commun. 165 (2021) 53–74, https://doi.org/10.1016/j.comcom.2020.10.023.

[25] S. Hayat, E. Yanmaz, R. Muzaffar, Survey on Unmanned Aerial Vehicle Networks for Civil Applications: A Communications Viewpoint, IEEE Commun. Surv. Tutorials. 18 (2016) 2624–2661, https://doi.org/10.1109/COMST.2016.2560343.

[26] O. Sami Oubbati, M. Atiquzzaman, T. Ahamed Ahanger, A. Ibrahim, Softwarization of UAV networks: A survey of applications and future trends, IEEE Access 8 (2020) 98073–98125, https://doi.org/10.1109/ACCESS.2020.2994494.

[27] A. Mukherjee, S. Misra, N.S. Raghuwanshi, A survey of unmanned aerial sensing solutions in precision agriculture, J. Netw. Comput. Appl. 148 (2019), 102461, https://doi.org/10.1016/j.jnca.2019.102461.

[28] I. Jawhar, N. Mohamed, J. Al-Jaroodi, D.P. Agrawal, S. Zhang, Communication and networking of UAV-based systems: Classification and associated architectures, J. Netw. Comput. Appl. 84 (2017) 93–108, https://doi.org/10.1016/j.jnca.2017.02.008.

[29] R. Tarjan, Depth-first search and linear graph algorithms, SIAM J. Comput. 1 (1972) 146–160.

[30] D. Arthur, S. Vassilvitskii, K-Means++: The advantages of careful seeding, in: Proc. 18th Annu. ACM-SIAM Symp. Discret. Algorithms (SODA'07), Philadelphia, PA, 2007, pp. 1027–1035.

[31] K. Mahmood, M.A. Khan, M.U. Hassan, A.M. Shah, S. Ali, M.K. Saeed, Intelligent On-Demand Connectivity Restoration for Wireless Sensor Networks, Wirel. Commun. Mob. Comput. 2018 (2018) 1–10, https://doi.org/10.1155/2018/9702650.

[32] K. Akkaya, F. Senel, A. Thimmapuram, S. Uludag, S. Hwang, W. Chao, C. Wu, C. Dow, Distributed recovery from network partitioning in movable sensor/actor networks via controlled mobility, IEEE Trans. Comput. 59 (2010) 258–271, https://doi.org/10.1109/TC.2009.120.

[33] V. Ranga, M. Dave, A.K. Verma, Relay node placement to heal partitioned wireless sensor networks, Comput. Electr. Eng. 48 (2015) 371–388, https://doi.org/10.1016/j.compeleceng.2015.09.014.

[34] V. Ranga, M. Dave, A.K. Verma, Restoration of lost connectivity of partitioned wireless sensor networks, Cogent Eng 3 (2016) 1–22, https://doi.org/10.1080/23311916.2016.1186263.

[35] K. Akkaya, F. Senel, Detecting and connecting disjoint sub-networks in wireless sensor and actor networks, Ad Hoc Networks 7 (2009) 1330–1346, https://doi.org/10.1016/j.adhoc.2009.01.003.

[36] A. Abbasi, M. Younis, U. Baroudi, Restoring connectivity in wireless sensor-actor networks with minimal topology changes, IEEE Int. Conf. Commun. (2010) 1–5, https://doi.org/10.1109/ICC.2010.5502448.

[37] A.A. Abbasi, K. Akkaya, M. Younis, A Distributed Connectivity Restoration Algorithm in Wireless Sensor and Actor Networks, in: 32nd IEEE Conf. Local Comput. Networks (LCN, IEEE, 2007, 2007, pp. 496–503, https://doi.org/10.1109/LCN.2007.19.

[38] V. Ranga, M. Dave, A.K. Verma, A Hybrid Timer Based Single Node Failure Recovery Approach for WSANs, Wirel. Pers. Commun. 77 (2014) 2155–2182, https://doi.org/10.1007/s11277-014-1631-4.

[39] M. Imran, M. Younis, A. Md Said, H. Hasbullah, Localized motion-based connectivity restoration algorithms for wireless sensor and actor networks, J. Netw. Comput. Appl. 35 (2012) 844–856, https://doi.org/10.1016/j.jnca.2011.12.002.

[40] M.Y. Sir, I.F. Senturk, E. Sisikoglu, K. Akkaya, An optimization-based approach for connecting partitioned mobile sensor/Actuator Networks, 2011 IEEE Conf. Comput. Commun. Work. INFOCOM WKSHPS 2011 (2011) 525–530, https://doi.org/10.1109/INFCOMW.2011.5928869.

[41] S.J. Habib, P.N. Marimuthu, An integrated restoration framework for coverage and communication within wireless sensor networks, Int. J. Ad Hoc Ubiquitous Comput. 15 (2014) 95–105.

[42] H. Wang, X. Ding, C. Huang, X. Wu, Adaptive connectivity restoration from node failure(s) in wireless sensor networks, Sensors (Switzerland) 16 (2016), https://doi.org/10.3390/s16101487.

[43] H. Essam, M. Younis, E. Shaaban, Minimum cost flow solution for tolerating multiple node failures in wireless sensor networks, in: IEEE Int. Conf. Commun. 2015-Septe, 2015, pp. 6475–6480, https://doi.org/10.1109/ICC.2015.7249356.

[44] I.F. Senturk, K. Akkaya, S. Yilmaz, Relay placement for restoring connectivity in partitioned wireless sensor networks under limited information, Ad Hoc Networks 13 (2014) 487–503, https://doi.org/10.1016/j.adhoc.2013.09.005.

[45] Y.K. Joshi, M. Younis, Distributed approach for reconnecting disjoint segments, GLOBECOM - IEEE Glob, Telecommun. Conf. (2013) 255–260, https://doi.org/10.1109/GLOCOM.2013.6831080.

[46] Y.K. Joshi, M. Younis, Autonomous recovery from multi-node failure in Wireless Sensor Network, 2012 IEEE Glob. Commun. Conf. (2012) 652–657, https://doi.org/10.1109/GLOCOM.2012.6503187. IEEE.

[47] I. Senturk, K. Akkaya, Energy and terrain aware connectivity restoration in disjoint Mobile Sensor Networks, in: Proc. - Conf. Local Comput. Networks, LCN, 2012, pp. 767–774, https://doi.org/10.1109/LCNW.2012.6424062.

[48] X. Cheng, D.Z. Du, L. Wang, B. Xu, Relay sensor placement in wireless sensor networks, Wirel. Networks. 14 (2008) 347–355, https://doi.org/10.1007/s11276-006-0724-8.

[49] Errol L. Lloyd, Guoliang Xue, Relay Node Placement in Wireless Sensor Networks, Freud, V.1 56 (2015) 134–138, https://doi.org/10.4324/9781315791937-9.

[50] S. Lee, M. Younis, Optimized relay placement to federate segments in wireless sensor networks, IEEE J. Sel. Areas Commun. 28 (2010) 742–752.

[51] R.C. Shah, S. Roy, S. Jain, W. Brunette, Data MULEs: Modeling and analysis of a three-tier architecture for sparse sensor networks, Ad Hoc Networks 1 (2003) 215–233, https://doi.org/10.1016/S1570-8705(03)00003-9.

[52] F. Senel, M. Younis, Optimized interconnection of disjoint wireless sensor network segments using K mobile data collectors, 2012 IEEE Int. Conf. Commun. (2012) 492–496, https://doi.org/10.1109/ICC.2012.6364467.

[53] A. Abbas, M. Younis, Establishing connectivity among disjoint terminals using a mix of stationary and mobile relays, Comput. Commun. 36 (2013) 1411–1421, https://doi.org/10.1016/j.comcom.2013.06.004.

[54] N. Banerjee, S. Chakraborty, V. Raman, Improved space efficient algorithms for BFS, DFS and applications. Int. Comput. Comb. Conf., Springer, 2016, pp. 119–130.

[55] G.M. Slota, S. Rajamanickam, K. Madduri, BFS and coloring-based parallel algorithms for strongly connected components and related problems, in: Proc. Int. Parallel Distrib. Process. Symp. IPDPS, 2014, pp. 550–559, https://doi.org/10.1109/IPDPS.2014.64.

[56] T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, Computational geometry, Introd. to Algorithms, 3rd Ed, MIT Press Cambridge, MA, USA, 2009, pp. 1022–1027.

[57] R.S. Barr, R.V. Helgaon, J.L. Kennington, Minimal spanning trees: An empirical investigation of parallel algorithms, Parallel Comput 12 (1989) 45–52, https://doi.org/10.1016/0167-8191(89)90005-7.
[58] I. Bekmezci, O.K. Sahingoz, Ş. Temel, Flying Ad-Hoc Networks (FANETs): A survey, Ad Hoc Networks 11 (2013) 1254–1270, https://doi.org/10.1016/j.adhoc.2012.12.004.
[59] H. Ullah, M. Abu-Tair, S. McClean, P. Nixon, G. Parr, C. Luo, Connecting Disjoint Nodes Through a UAV-Based Wireless Network for Bridging Communication Using IEEE 802.11 Protocols, Eurasip J. Wirel. Commun. Netw. (2020) 2020, https://doi.org/10.1186/s13638-020-01727-z.

Aditi Zear is currently working as PhD research scholar in the department of computer engineering in National Institute of Technology (NIT), Kurukshetra, Haryana. She was previously associated with National Institute of Technology (NIT), Hamirpur, Himachal Pradesh as lecturer. She has obtained her M. Tech degree in Computer Science & Engineering from Jaypee University of Information Technology (JUIT) Waknaghat, Solan, Himachal Pradesh-India in 2016. She has obtained her B.Tech degree in Information Technology from University Institute of Information Technology, Shimla, Himachal Pradesh. Her research interests include Data Hiding techniques, Wireless Sensor Networks and Unmanned Aerial Vehicles.

Dr. Virender Ranga has received his Ph.D. degree in Feburary 25, 2016 from Computer Engineering Department of National Institute of Technology, Kurukshetra, Haryana, India followed by M.Tech. and B.Tech. degrees. He worked as Assistant Professor in the Computer Engineering Department, National Institute of Technology Kurukshetra, India since March 20, 2008 to October 25, 2021. Currently, Dr. Virender Ranga is working as Associate Professor in IT Department of Delhi Technological University, Delhi. He has been published more than 100 research papers in various SCI/SCIE/SCOPUS Journals and reputed International Conferences in the area of Computer Communications and Computer Security. He has been conferred by Young Faculty Award in 2016 for his excellent contributions in the field of Computer Communications. He is a member of editorial board of various reputed journals like Journal of Applied Computer Science & Artificial Intelligence, International Journal of Advances in Computer Science and Information Technology (IJACSIT), Circulation in Computer Science (CCS), International Journal of Bio Based and Modern Engineering (IJBBME) and International Journal of Wireless Networks and Broadband Technologies. Currently, he is acting as Ambassador of Asia Region for Bentham Science. He is an active reviewer of many reputed journals of IEEE Transactions, Springer, Elsevier, Talyor & Francis, Wiley and InderScience. His research area includes Wireless Sensor & Adhoc Networks Security, IoT security, FANET security, SDN security, IoRT etc.

Hindawi
Advances in Civil Engineering
Volume 2022, Article ID 1529416, 17 pages
https://doi.org/10.1155/2022/1529416

Research Article

Wind Excited Action around Tall Building Having Different Corner Configurations

Rahul Kumar Meena, Ritu Raj, and S. Anbukumar

Department of Civil Engineering, Delhi Technological University, Delhi, India

Correspondence should be addressed to Ritu Raj; rituraj@dtu.ac.in

Received 18 September 2021; Revised 25 January 2022; Accepted 29 January 2022; Published 25 February 2022

Academic Editor: Filippo Ubertini

Copyright © 2022 Rahul Kumar Meena 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.

Aeroelastic instabilities are common in square or rectangular plan shape structures due to the development of powerful vortices by the rolling motion of the separated shear layers. Windward corner modifications such as corner cut, recession, rounded, and slotted help to reduce instabilities. Codal recommendations about common shapes like square, rectangle, triangle, and circular shapes are available in different international standards, but they did not provide the detailed analysis results for the different regular shapes with corner configurations. Although analysis was performed on various plan shapes of tall buildings using the computational fluid dynamics technique through k-ε turbulence model, a very small number of studies were performed on particular shapes with different corners having same plan area and height. The mean pressure coefficients of Model-A and Model-B were compared with various international standards and wind tunnel data, respectively, for validation, showing a nearly equivalent consistency; however, international standards consist of coefficients at 0° and 90° wind angles only. Wind effects on building shapes having different corners change the characteristics of the separated shear layer and reduce the downstream wake which helps to reduce drag and lift forces simultaneously. Recent study shows that the windward pressure distribution pattern is almost independent of building size and height. Therefore incorporation of corners in building helps to reduce the forces caused by extreme wind. A very large amount of numerical simulation data about wind pressure is generated, which can be used by the designer while designing such a building for wind load. Comparison have been made between buildings having different corners under same wind speed, and Model-D (round corner) performed very well against the wind.

1. Introduction

Because of the ever-increasing population, cities' growth in the horizontal direction has reached a saturation point. As a result, tall structures are springing up all over the planet. Tall buildings are particularly vulnerable to wind loads, necessitating effective wind load design. Irwin [1] studied the bluff body aerodynamics in wind engineering. Bluff bodies subjected to strong large vortices in their wakes, wind turbulence, are another factor that has a significant impact on vortex shedding. The rectangular building is very common in plan shape, and it is generally symmetrical about both axes. This plan-shaped building is very common for residential as well as corporate buildings [2–5]. Effects of wind load for common plan shape are available in various international standards [6–10]. Wind tunnel testing and CFD technologies are used to study the impact of wind on tall buildings. Wind tunnel studies and/or CFD simulations are required to get wind driven forces in complex plan shaped structures.

Wind effects on structure are varying as per the shape of building; to evaluate the effect of wind on such structures, wind tunnel testing has been performed by many researchers; for example, Sheng et al. [11] investigated the fluctuating properties of global and local wind loads; Sun et al. [12] conducted wind tunnel test on 1040-meter building model; Irwin et al. [13] examined the effect of corners; Raj and Ahuja [14] performed an experiment on rigid models with varied cross-sectional shapes and equal

floor area; Bhattacharyya and Dalui [15] presented wind effects on "E"-shaped building; Nagar et al. [16] studied "H" and square building model; Bandi et al. [17] performed an experiment on six building models; Raj et al. [18] studied the response of "+" shape; Carassale et al. [19] studied corner effect on a square cylinder; Tanaka et al. [20] studied modified square shape; He and Song [21] experimentally investigated the response of TTU building under wind load. According to the literature, modifying the shape of an entrant's corner by roughly 10% of the building width can reduce wind effects on high-rise buildings. Some researchers also evaluated the effect of interference of principal building models; for example, Pal et al. [22] estimated drag and lift force coefficients on interfering square and fish plan shape models; Nagar et al.[16] have studied interference effects on "H" and square model; Pal and Raj [23] evaluated the full blockage condition for various plan shapes of tall building models; Pal et al. [24] studied interference effect of twin building models on square plan shape and remodelled triangle shape model; Nagar et al. [25] studied interference effect on two "+" shape models. Few important outcomes of interfering studies are as follows: pressure distribution is symmetric for isolated building, and drag and lift force coefficient magnitude is based on the position and geometric shape of the neighbouring building model. Wind tunnel test is time-consuming and costly, and testing facilities are also limited; hence, the effect of wind on high-rise structures can be found using CFD techniques.

CFD technique gives the wind effects on various shapes of high-rise buildings in quick time; it works by dividing the building geometry into small elements using meshing, and a number of studies have been carried out on various shapes, for example, Bhattacharyya and Dalui [26] on "E" shape, Raj et al. [27] on "H" shape, He and Song [28] on TTU building, Yu and Kareem [29] and Sanyal and Dalui [30] on rectangular shape, Paul and Dalui [31] and Okajima et al. [32] on rectangular shape with horizontal limbs, Paul and Dalui [33] on "+" shape, Hajra and Dalui [34] on octagonal shape, Gaur and Raj [35] on square model, Bairagi and Dalui [36] on stepped building model, and Sanyal and Dalui [37–40] on "Y" shape building model. The major findings obtained through the numerical simulation are as follows: results obtained through the numerical simulation are significantly affected by the mesh size, pressure distributed on the building model depends on the height of the building model, velocity alteration is found on the roof of the building model, and the drag force depends on the exposure to wind. The reduction in wind load mainly occurs due to the corner of short edges.

The effect of wind on the various plan-shaped tall building is studied using the computational fluid dynamics (CFD) technique at an angle range from 0° to 90° at an interval of 15°. For validation, the results obtained through numerical simulation are compared with previous wind tunnel experiments and different international standards. CFD technique is used to study the streamlines around the building models to get a clear view of flow characteristics. The distribution of local wind pressure on different faces of the building models is closely examined in order to fully understand the wind load patterns that occur at different wind incidence angles. Pressure contour on different faces on the building is discussed with reasonable extant.

2. Numerical Study

Numerical study was performed using computational fluid dynamics. Most of the studies in the past were done using the k- ε model. This model has well-established prediction capability and has been proven to be stable and numerically reliable. This model gives an accurate balance of reliability for general-purpose simulations. It is very less expensive and is mostly used to simulate the turbulent flow characteristics. k-ε turbulence is a two-equation model and provides the solution by using two transport equations, that is, turbulent kinetic energy (k) and turbulent dissipation rate (ε). It has positive advantage of not including any geometry-related parameters in the modelling. The turbulent kinetic energy and the turbulence dissipation rate are two variables introduced into the system of equations for the model.

(i) Basic equations. The basic equations used to study the fluid flow problems are Navier-Stokes and continuity equations.

(ii) Navier-Stokes equation is as follows:

$$\frac{\partial(\rho u_i)}{\partial t} = -\frac{\partial(\rho u_i u_j)}{\partial x_j} - \frac{\partial P}{\partial x_j} + \frac{\partial}{\partial_{x_j}}\left[\mu\left(\frac{\partial_{u_i}}{\partial_{x_j}} + \frac{\partial_{u_j}}{\partial_{x_i}}\right)\right] + F. \tag{1}$$

(iii) continuity equation is as follows:

$$\frac{\partial_\rho}{\partial_t} + \frac{\partial_{\rho_i}}{\partial_{x_i}} = 0. \tag{2}$$

The standard k-ε model uses the following transport equations for the turbulence kinetic energy and turbulence dissipation:

$$\rho\frac{\partial k}{\partial t} + \rho\overline{u_i}\frac{\partial k}{\partial x_j} = \tau_{ij}\frac{\partial \overline{u_i}}{\partial x_j} - \rho\varepsilon + \frac{\partial}{\partial_{x_j}}\left[\left(\mu + \frac{\mu_t}{\sigma_\varepsilon}\right)\frac{\partial \varepsilon}{\partial x_j}\right]. \tag{3}$$

Momentum equation is as follows:

$$\frac{\partial(\rho U_i)}{\partial t} = -\frac{\partial(\rho U_i U_j)}{\partial x_j} - \frac{\partial P}{\partial x_i} + \frac{\partial}{\partial_{x_j}}\left[\mu_{\text{eff}}\left(\frac{\partial_{U_i}}{\partial_{x_j}} + \frac{\partial_{U_j}}{\partial_{x_i}}\right)\right] + S_M. \tag{4}$$

2.1. Building Model. In the present study, the building is modelled at a 1 : 200 length scale. CFD simulations were used to explore the effect of wind on the modified corners on rectangular, corner cut, chamfered, and fillet shapes of almost identical plan area and equal height of 750 mm. The variation of pressure coefficient on building surfaces around the whole building models is discussed in this study, for the building models shown in Figure 1. Slight difference in the

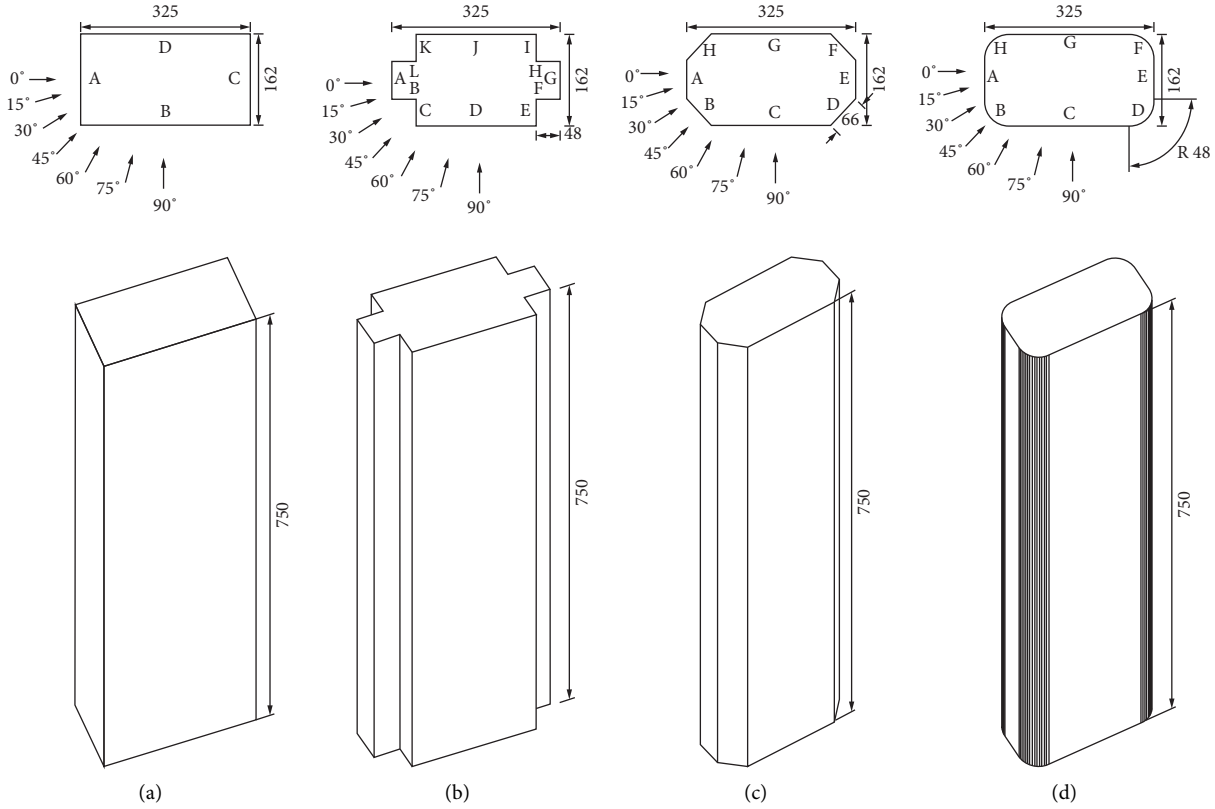


Figure 1: Plan and isometric view of (a) rectangular, (b) corner cut, (c) chamfered, and (d) fillet shapes of tall building models.

proportions for all the building models is kept to maintain the equal plan area. Four building models are used to understand the wind-induced responses on various building faces with varying skew angles of wind stream varying from 0° to 90° at 15° intervals.

2.2. Boundary Condition. The effect of wind on building models is studied using CFD simulation. CFD works by dividing a space into a grid containing a large number of cells. The grid of the cells surrounded by boundaries that simulate the surfaces, opened and enclosed space, pressure of the boundaries, and air movements within the cell is then set to a starting condition, which is closed to the real environment of the wind tunnel experiment. The inlet, top, and sidewall boundaries are considered 5 H from the model and outlet boundary is placed at 15 H behind the model for simulation to develop the flow properly as recommended by Frank et al. [41] and Revuz et al. [42]. To define a problem that results in a unique solution, it is required to specify the information on the dependent variables at the domain boundaries, as the governing equation is differential, and their solution requires integration. The boundary condition is the mathematical equivalent of the constant of integration, the value of which is required to gain a unique solution. The domain is represented in Figure 2, and the sidewall and top are kept as free slip in the CFX-Pre setup. Ground and building model surfaces in this study are considered as no-slip wall in CFX-Pre setup. No-slip is defined as follows: the velocity of the air at wall boundary is the same as the air velocity at inlet. Free slip is defined as follows: velocity components parallel to wall have finite value (computed by the solver), but the velocities normal to the wall and shear stress are both set to zero.

At the inlet, the free stream velocity is set to 10 m/s. The domain is designed according to Revuz et al. [42] (Figure 3). The velocity profile and turbulence intensity profile are graphically depicted in Figure 4 and compared with previous wind tunnel data by Raj (2015).

2.3. Meshing. Meshing is the most important part of numerical simulation. Finer mesh might not give you reliable results, but a bad mesh will surely affect the results. Meshing is of different sizes and provides capturing the flow characteristics in critical zones. Meshing divides the complex geometry into elements that can be used to discretize a domain since meshing typically consumes time. Meshing is represented in Figure 3, domain meshing is provided as tetrahedron meshing as it consumes fewer resources in computing the result, and building meshing is of different sizes so that solution can capture a clear picture of flow behaviour near the building surfaces. The edge meshing is also provided and is finer than the other meshing sizes to capture the flow characteristics. Inflation is provided to simulate the flow easily around the building model's geometry.

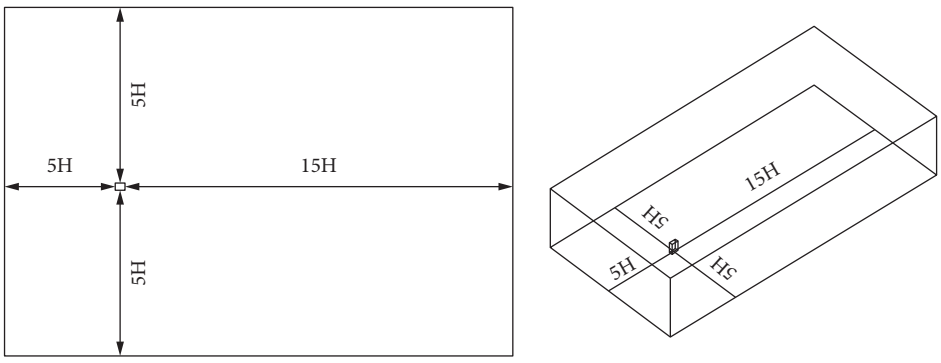
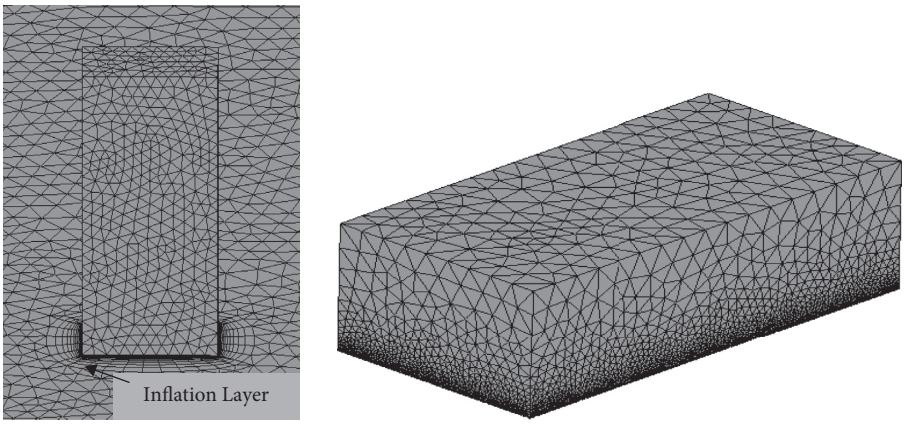


Figure 2: Domain used in CFD Simulation.



Building Meshing Domain Meshing

Figure 3: Type of meshing. (a) Building meshing. (b) Domain meshing.

2.4. Velocity and Turbulent Intensity Profile. The boundary condition used in the numerical simulation is similar to the boundary condition used by Raj [43]. To obtain a good relationship between numerical simulation and wind tunnel experiment, the boundary condition should be the same as those used in experiments, especially for inflow boundary conditions.

The variation in the wind with height is generally expressed in logarithmic form.

$$U_z = \frac{\mathbf{u}_0}{\mathbf{k}} \mathbf{ln}\left(\frac{\mathbf{Z}}{\mathbf{Z_o}}\right), \tag{5}$$

where U_z is the wind speed at height z, U_o is friction velocity, Z_o is the characteristic height of surface roughness, and k is a Karman constant with a value of 0.4.

The boundary layer wind profile is governed by the power low equation:

$$U_z = U_o\left(\frac{z}{z_o}\right)^n, \tag{6}$$

where U_z is the wind speed at height Z, U_o is the mean wind speed at a reference height Z_o, and "n" is a parameter that varies with ground roughness.

Generally, turbulence intensity for the smooth terrain is less than the turbulence intensity for rough terrain. Turbulence intensity is a nondimensional quantity derived from the variance and for the mean wind speed.

$$I_z = \frac{\sigma_z}{U_z}, \tag{7}$$

where I_z is the turbulence intensity at height z, σ_z is the standard deviation of the wind speed at height z, and U_z is the mean wind speed at reference height.

2.5. Validation. CFD results are based upon a number of factors. Therefore, it is necessary to validate the accuracy of the numerical results. For this purpose, building the various numerical results of Model-A (rectangular) and Model-B (corner cut) is compared with various international standards and experimental results.

The external pressure coefficients, C_p (face average value), for the different faces of Model-A are tabulated in Table 1 and compared between numerical results and various international standards. Results in Figure 5 and Table 1 are obtained using the k-ε turbulence models, and results are nearly the same with various international standards and experimental results.

The external pressure coefficient "C_p" is calculated using the following equation:

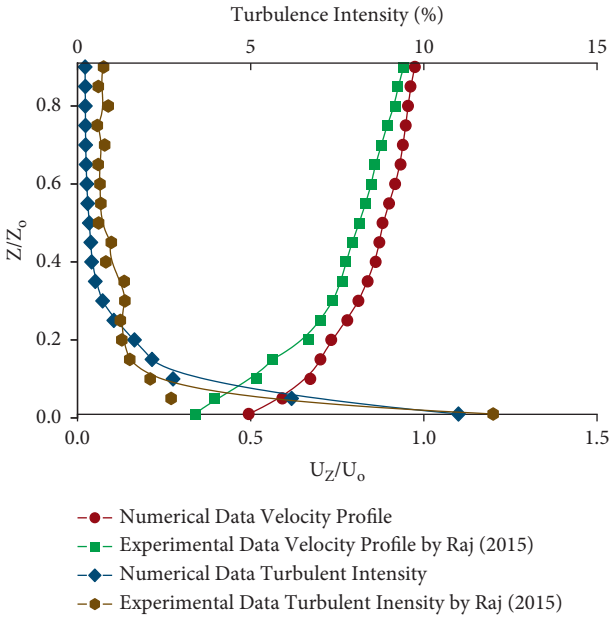


Figure 4: Variation of velocity and turbulent intensity.

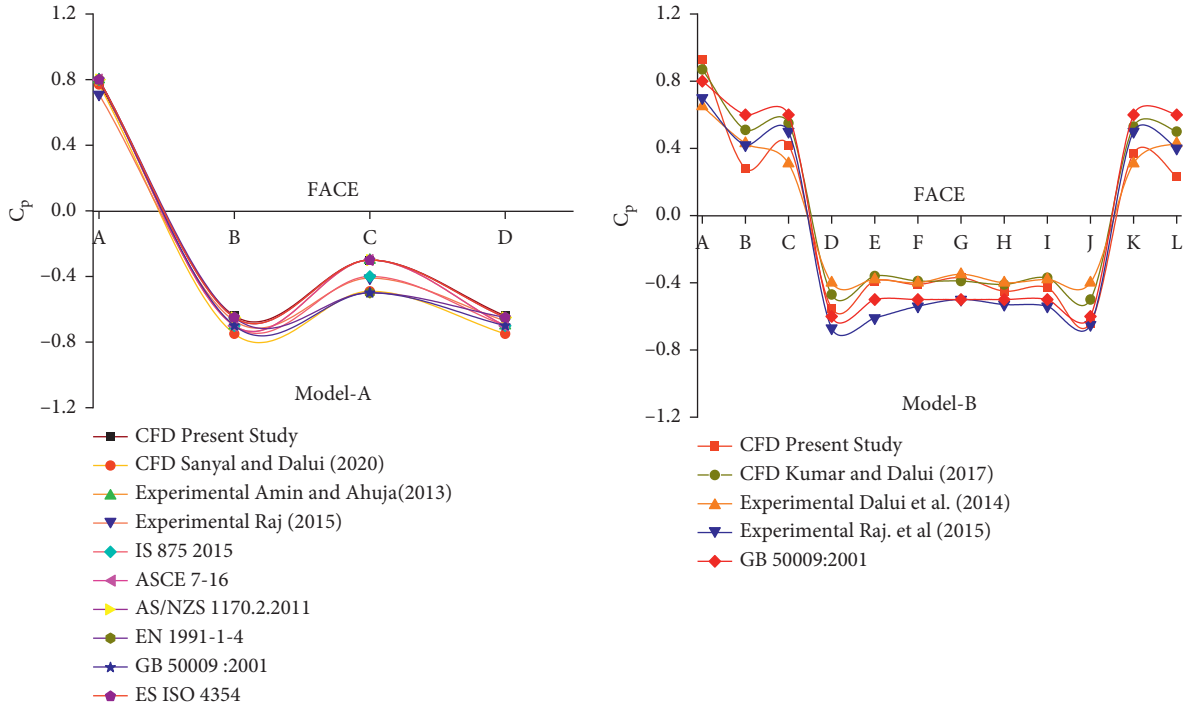


Figure 5: Average C_p for Model-A and Model- B at 0° wind incidence.

$$C_p = \frac{p - p_o}{1/2\rho U_Z{}^2}, \tag{8}$$

where p is the pressure derived from the needed point, p_o is the reference height static pressure, ρ is the air density (1.225 kg/m^3), and U_z is mean wind velocity at the building reference height.

3. Results and Discussions

3.1. Flow Pattern. The wind pressure distribution against a wall of the high-rise building is affected by various wind characteristics such as vortex, separation, stagnation, recirculation, and reattachments. Wind flow for Model-A is presented in Figure 6, while streamlines for rectangular

Table 1: Comparison of average face pressure coefficient (C_p) on the rectangular tall building.

International code	Wind angle	Windward side	Sidewall	Leeward side
CFD results	0°	0.78	−0.64	−0.30
	90°	0.70	−0.61	−0.45
IS 875 (part 3)	0°	0.7	−0.7	−0.4
	90°	0.8	−0.5	−0.1
ASCE/SEI 7-16	0°	0.8	−0.7	−0.5
	90°	0.8	−0.7	−0.5
AS/NZS 1170.2.2011	0°	0.8	−0.65	−0.5
	90°	0.8	−0.65	−0.5
EN 1991-1-4	0°	0.8	−0.5	−0.7
	90°	0.8	−0.5	−0.7
BS 6399-2	0°	0.8	−0.5	−0.7
	90°	0.8	−0.5	−0.7
GB 50009-2001	0°	0.8	−0.5	−0.7
	90°	0.8	−0.5	−0.7
NSCP 2015	0°	0.8	−0.5	−0.7
	90°	0.8	−0.5	−0.7
ES/ISO 4354: 2012	0°	0.8	−0.65	−0.7
	90°	0.8	−0.65	−0.7

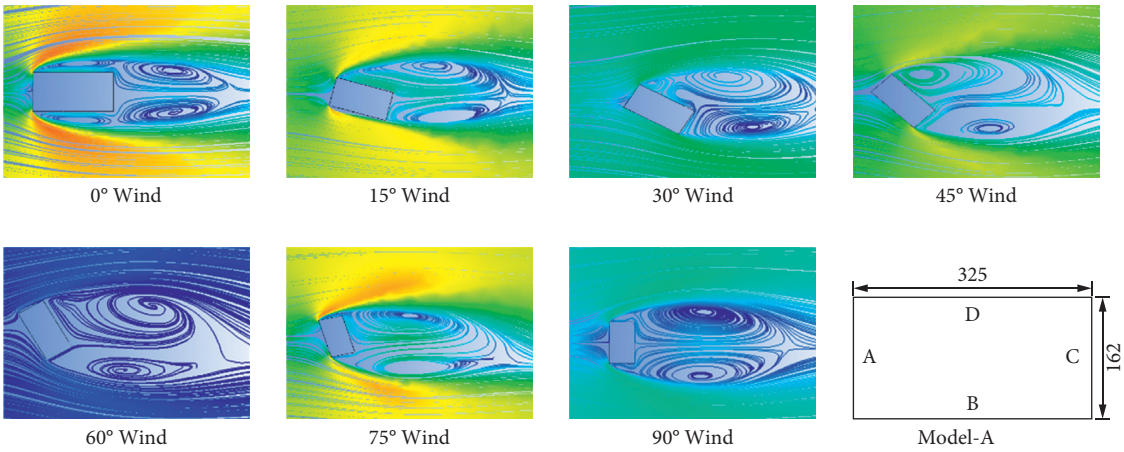


Figure 6: Flow lines for building Model-A at various wind incidence angles. (a) 0° wind. (b) 15° wind. (c) 30° wind. (d) 45° wind. (e) 60° wind. (f) 75° wind. (g) 90° wind. (h) Model-A.

model having corner cut are represented in Figure 7. The streamlines in plan for Model-C are depicted in Figure 8 and for Model-D wind flow patterns in plan are illustrated in Figure 9. The size of downstream wake depends on the wind incidence angle and is maximum for Model-A at 30°.

Pressure distribution and vortex generation are based on the geometric configuration of the building. Suction is observed in the rear side of the building with respect to the wind flow direction. Velocity at the edge is increased due to the separation of flow. The reattachment of flow is observed more on side faces at 0° and 15° wind. Streamlines patterns are unaffected by wind speed and are mostly determined by the geometry of the building and upwind conditions.

3.2. Pressure Distribution. The pressure distribution is expressed in the form of the pressure contours for various building shapes in Figures 10–13. For the case of building Model-A (Figure 10), maximum pressure on the centre of the face is 1.05, on windward face-B, at 90°, and the minimum (−0.7) is on face-B at 0°. Pressure distribution is maximum on the centre of the face, and then it is decreasing towards the edges of the faces because of the increase in the flow velocity. Contour lines become dense towards the edges of the face, indicating the changes in the pressure distribution. Wind pressure follows the same pattern for face-A at 0° and face-B at 90° because wind flow is in direct contact with both faces. For side face-B, at 0°, suction is increasing as wind flow passes the building model from windward to leeward edge.

Figure 11 represents the wind pressure distribution for building Model-B; maximum positive pressure is 1.05 on face-D at 90°, and maximum negative pressure is 0.8 on face-J at 0°. Pressure distribution on face-A at 0° and face-D at 90° is the windward face for the respective case and follows the same pressure distribution pattern. Pressure distribution on the windward face is almost independent of the width and depth of the face.

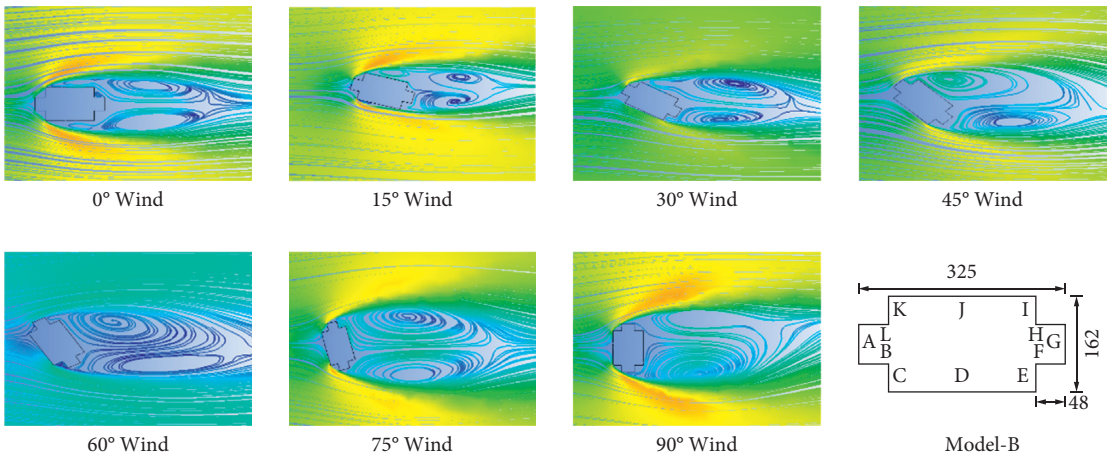


Figure 7: Flow lines for building Model-B at various wind incidence angles. (a) 0° wind. (b) 15° wind. (c) 30° wind. (d) 45° wind. (e) 60° wind. (f) 75° wind. (g) 90° wind. (h) Model-B.

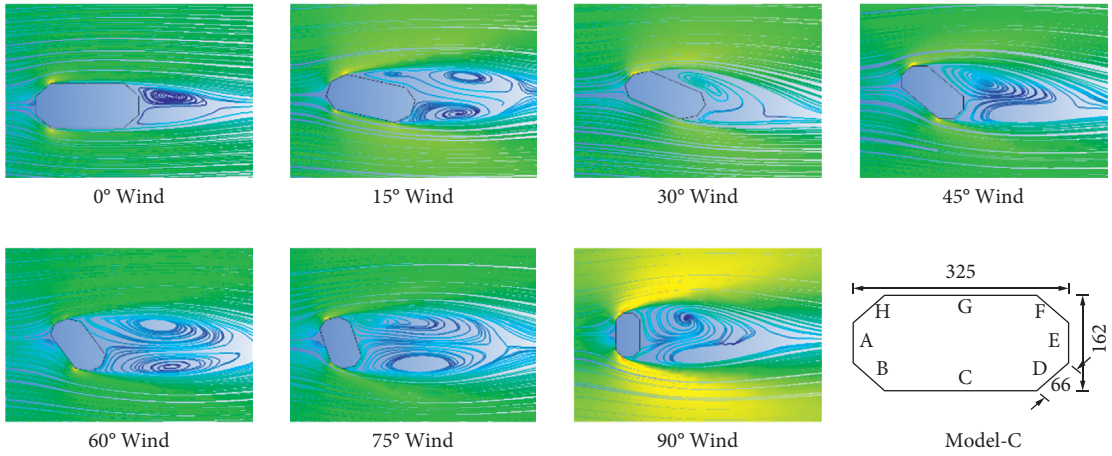


Figure 8: Flow lines for building Model-C at various wind incidence angles. (a) 0° wind. (b) 15° wind. (c) 30° wind. (d) 45° wind. (e) 60° wind. (f) 75° wind. (g) 90° wind. (h) Model-C.

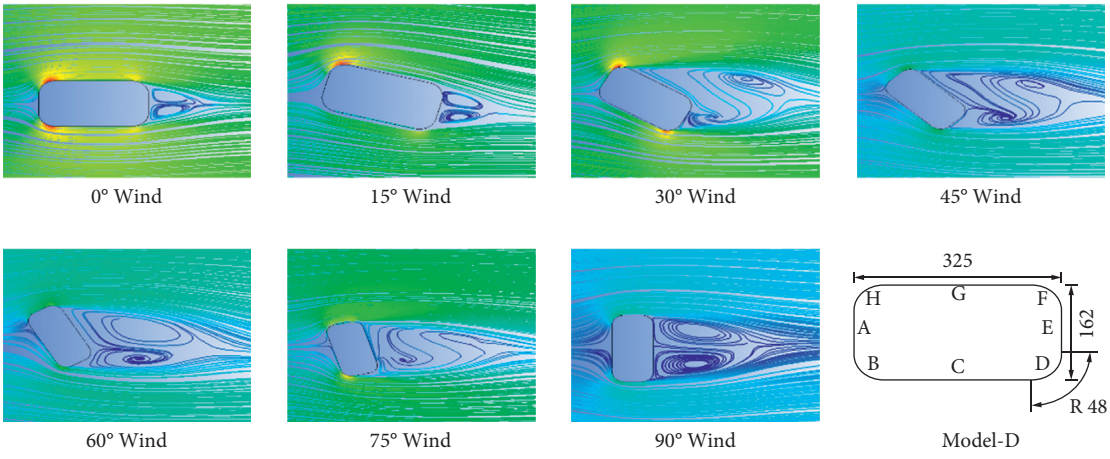


Figure 9: Flow lines for building Model-D at various wind incidence angles. (a) 0° wind. (b) 15° wind. (c) 30° wind. (d) 45° wind. (e) 60° wind. (f) 75° wind. (g) 90° wind. (h) Model-D.

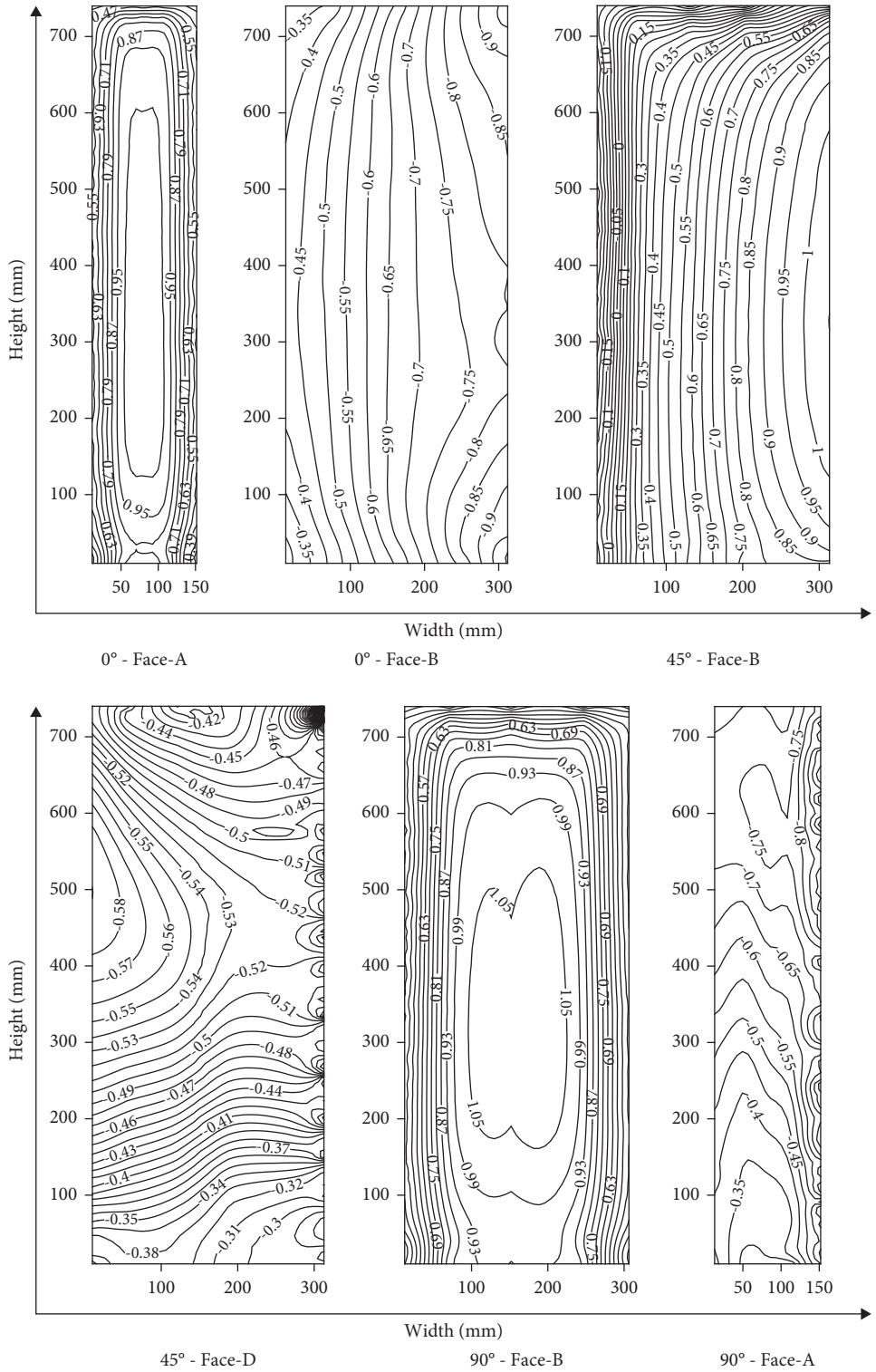


Figure 10: Distribution of mean wind pressure coefficient on different faces of Model-A.

Wind pressure distribution for building Model-C (Figure 12) at 90° on face-C (windward) has maximum pressure (1.07) at the centre of the face and minimum pressure (−2.9) on face-B at 90°. In Figure 10, the maximum pressure is 1.05 on face-C (windward) at 90°, and the minimum pressure is −0.8 on face-G (side face) at 0° wind and it decreases in negative from windward face to leeward face. Pressure generally increases with height because of increasing velocity in the approach flow as wind speed increases with height.

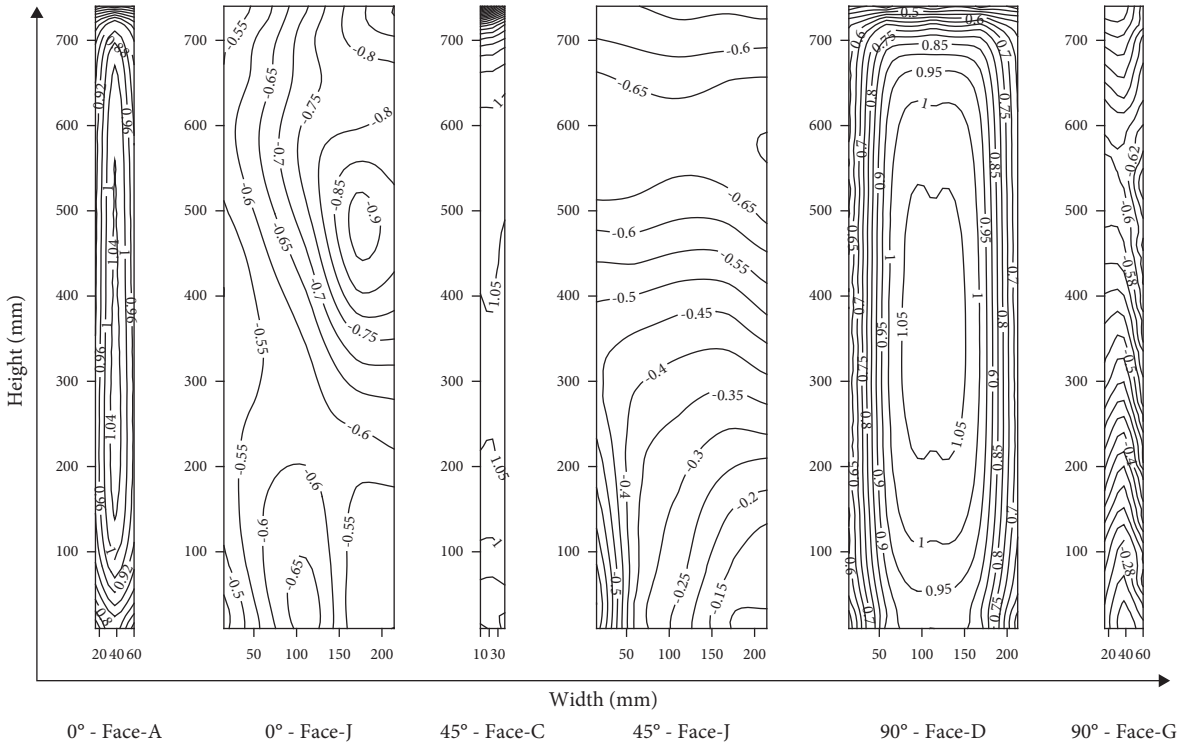


Figure 11: Distribution of mean wind pressure coefficient on different faces of Model-B.

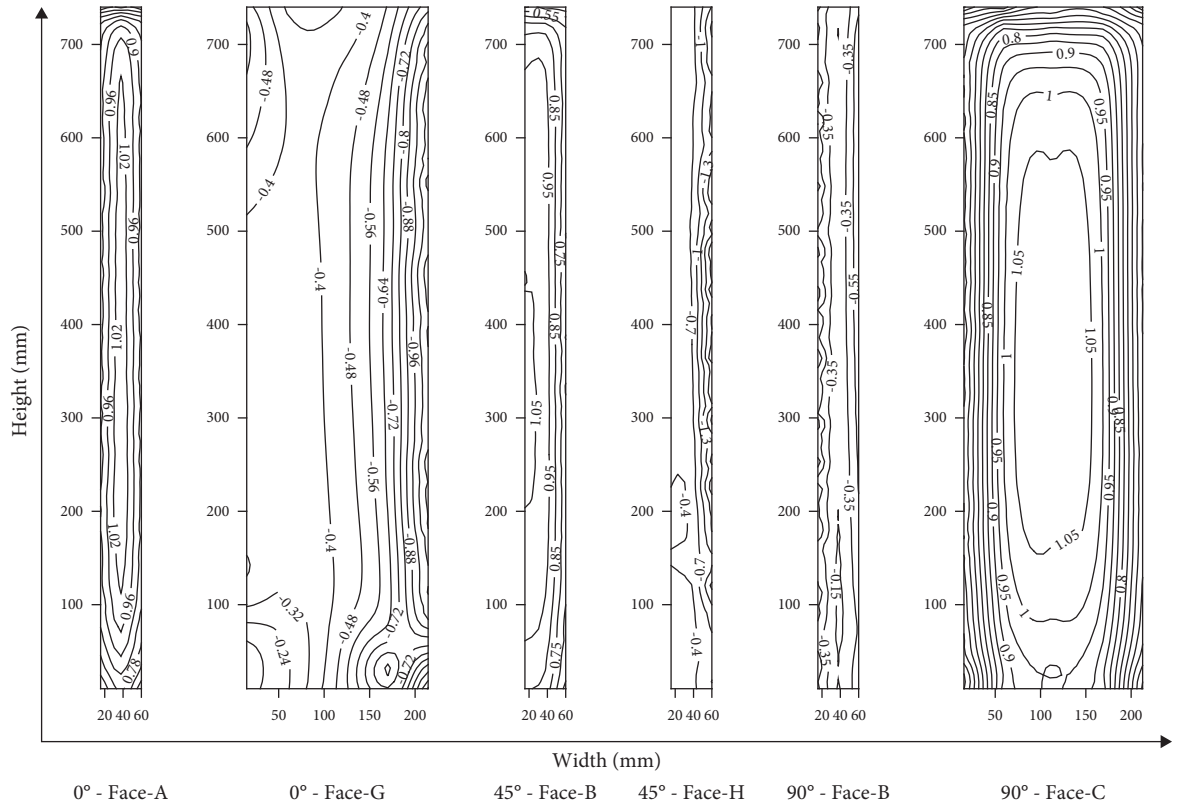


Figure 12: Distribution of mean wind pressure coefficient on different faces of Model-C.

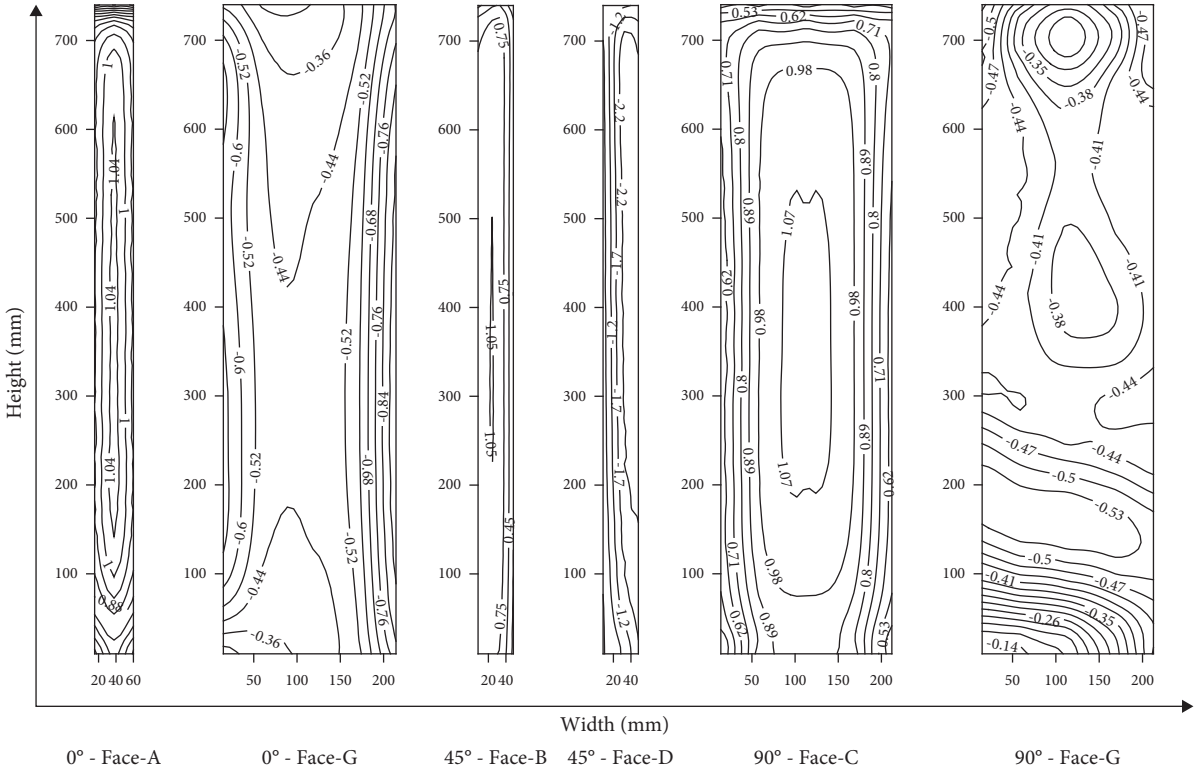


Figure 13: Distribution of mean wind pressure coefficient on different faces of Model-D.

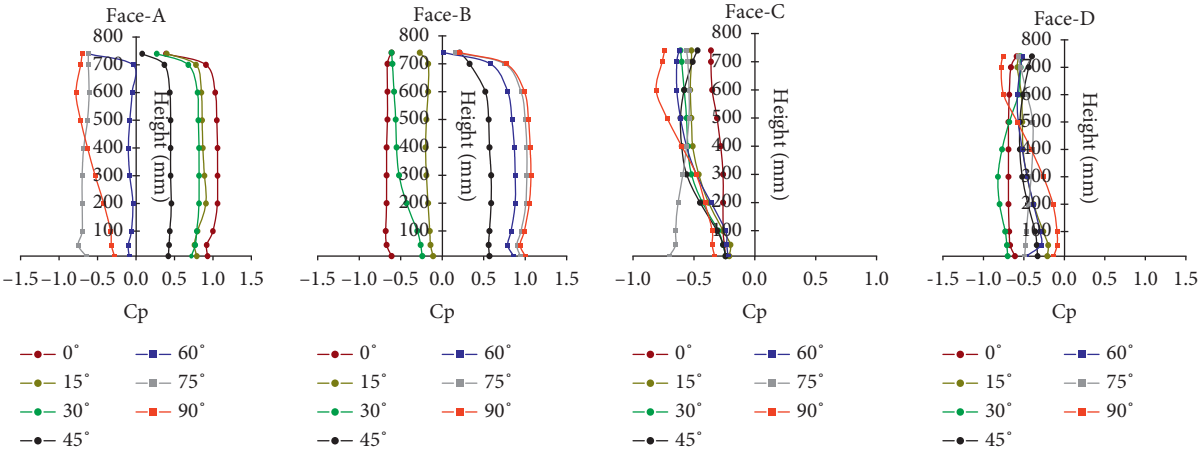


Figure 14: Pressure variation along the vertical centreline for various faces of Model-A at various wind incidence angles.

3.3. Vertical C_p. Variation in C_p along the vertical centreline for 0° to 90° wind incidence angles at an interval of 15° of each face is in Figure 14 for rectangular building Model-A. In Figure 14, face-C and face-D are subjected to negative pressure for all wind incidence angles. Face-A and face-B are subjected to maximum positive pressure at 0° and 90° wind incidence angles, respectively. Negative pressure zone is observed on face-A at 60°, 75°, and 90° wind incidence angles. Face-B has been subjected to negative pressure at 0°, 15°, and 30° wind incidence angles. In Figure 14, the vertical centreline pressure distribution on face-A is decreasing continuously from 0° to 90°, is maximum at 0° wind, and is minimum at 90° for Model-A.

For plus-shaped building Model-B, the distribution of pressure coefficients on the vertical centreline is represented in Figure 15 for 0° to 90° wind incidence angles at an interval of 15°. in Figure 15, face-E, face-F, face-G, face-H, face-I, and face-J are subjected to negative pressure on all wind incidence angles. Face-B and face-C are subjected to negative pressure at 90° wind incidence angles and subjected to

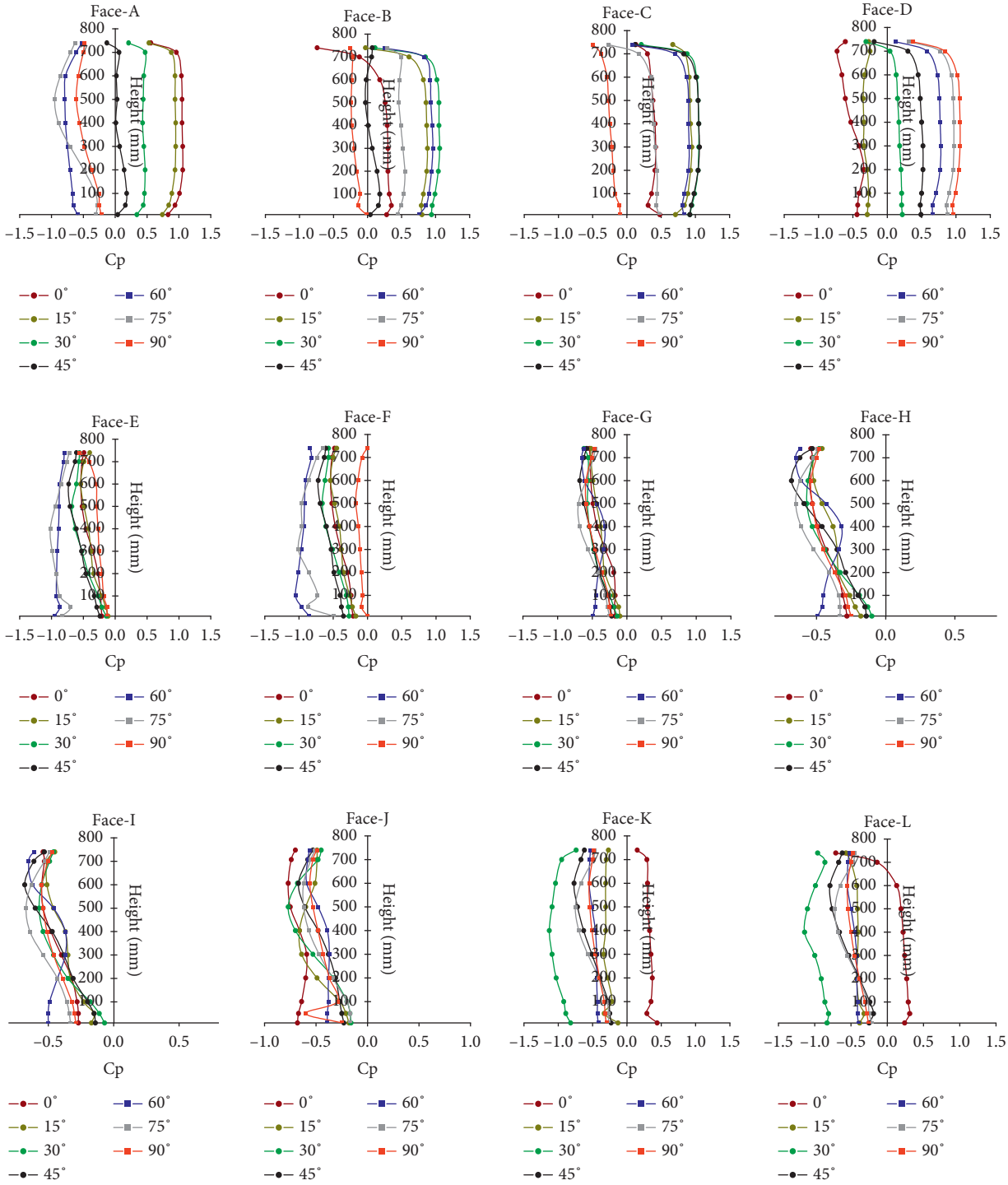


Figure 15: Pressure variation along the vertical centreline for all faces of Model-B at various wind incidence angles.

positive pressure at 0° to 75° wind incidence angles. Face-D is subjected to negative pressure at 0° and 15° wind incidence angles. At the rest of the angles, this is subjected to positive pressure. Face-K and face-L are under the effect of positive pressure at 0° wind incidence angle for the rest of the angle, and they are subjected to negative pressure.

Model-C vertical centreline pressure distribution is presented in Figure 16; for face-A, maximum C_p is at 0° wind and minimum C_p is at 75° wind incidence angle. Face-D, face-E, face-F, face-G, and face-H are under the effect of negative pressure, that is, suction for 0° to 90° wind incidence angles at an interval of 15° each. Face-E and face-F have a maximum magnitude of pressure at 0° wind and minimum magnitude of pressure at 75° because of recirculation of flow. Minimum C_p is at 0° wind incidence angle because flow separation takes place at these recessed corners.

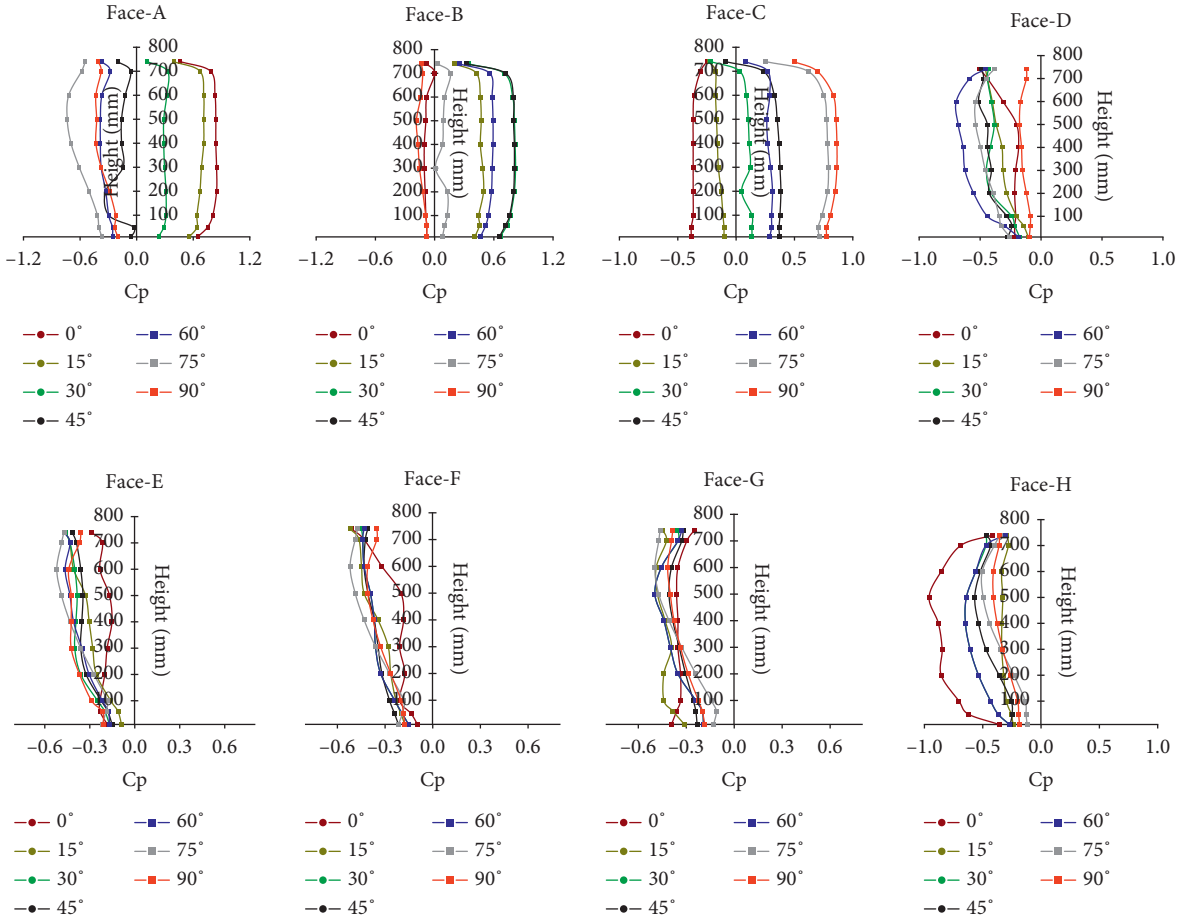


Figure 16: Pressure variation along the vertical centreline for all faces of Model-C at a various wind incidence angles.

For recesses/chamfer Model-D, vertical centreline pressure distribution pattern is shown in Figure 17, face-A, under the direct exposure to wind at 0° wind, so the maximum vertical centreline pressure is at 0° wind and the maximum on face-C is at 90° wind incidence angle. Face-A is under the reattachment of flow from 45° to 90° wind incidence angle. Face-D, face-E, face-F, face-G, and face-H are under the effect of negative pressure, that is, suction for 0° to 90° wind incidence angles at an interval of 15° each. Recessed corner face-B has maximum vertical centreline pressure at 0° and minimum at 90° wind incidence angle, as well as recessed corner face-H.

3.4. Drag and Lift Force. Wind load has long been a subject of research for wind engineers. Wind-produced responses on structures are difficult to quantify due to the complexity of structural geometry and field category. In modern design practice, wind loads can be calculated using wind tunnel tests and CFD technique. In this study, force coefficients in X-direction (C_{fx}) and Y-direction (C_{fy}) are obtained using equations (9) and (10) Therefore, values of F_x and F_y are obtained using ANSYS simulation.

Figure 18 depicts the variation of force coefficient in X-direction (C_{fx}) and Y-direction (C_{fy}) with respect to varying wind incidence angles. Maximum C_{fx} and C_{fy} are at 45° wind incidence angle for Model-C and Model-D.

$$C_{f_x} = \frac{F_x}{\left(0.5\rho_{U_h{}^2}.A_p\right)}, \tag{9}$$

$$C_{f_y} = \frac{F_y}{\left(0.5\rho_{U_h{}^2}.A_p\right)}. \tag{10}$$

Figure 19 shows the variation of moment coefficient in X-direction (C_{mx}) and Y-direction (C_{my}) with respect to varying wind incidence angles. C_{mx} is maximum for Model-C and Model-D at 30° wind incidence angle. C_{my} is maximum at 45° wind incidence angle for Model-C and Model-D.

$$C_{m_x} = \frac{M_x}{\left(0.5\rho_{U_h{}^2}.A_p.1/2.H\right)}, \tag{11}$$

$$C_{m_y} = \frac{M_y}{\left(0.5\rho_{U_h{}^2}.A_p.1/2.H\right)}. \tag{12}$$

The moment coefficient in the X-direction and Y-direction is obtained using equations (11) and (12); here, ρ is the density of air and U_h is the reference velocity at the building height, A_p is the area projected in the direction of

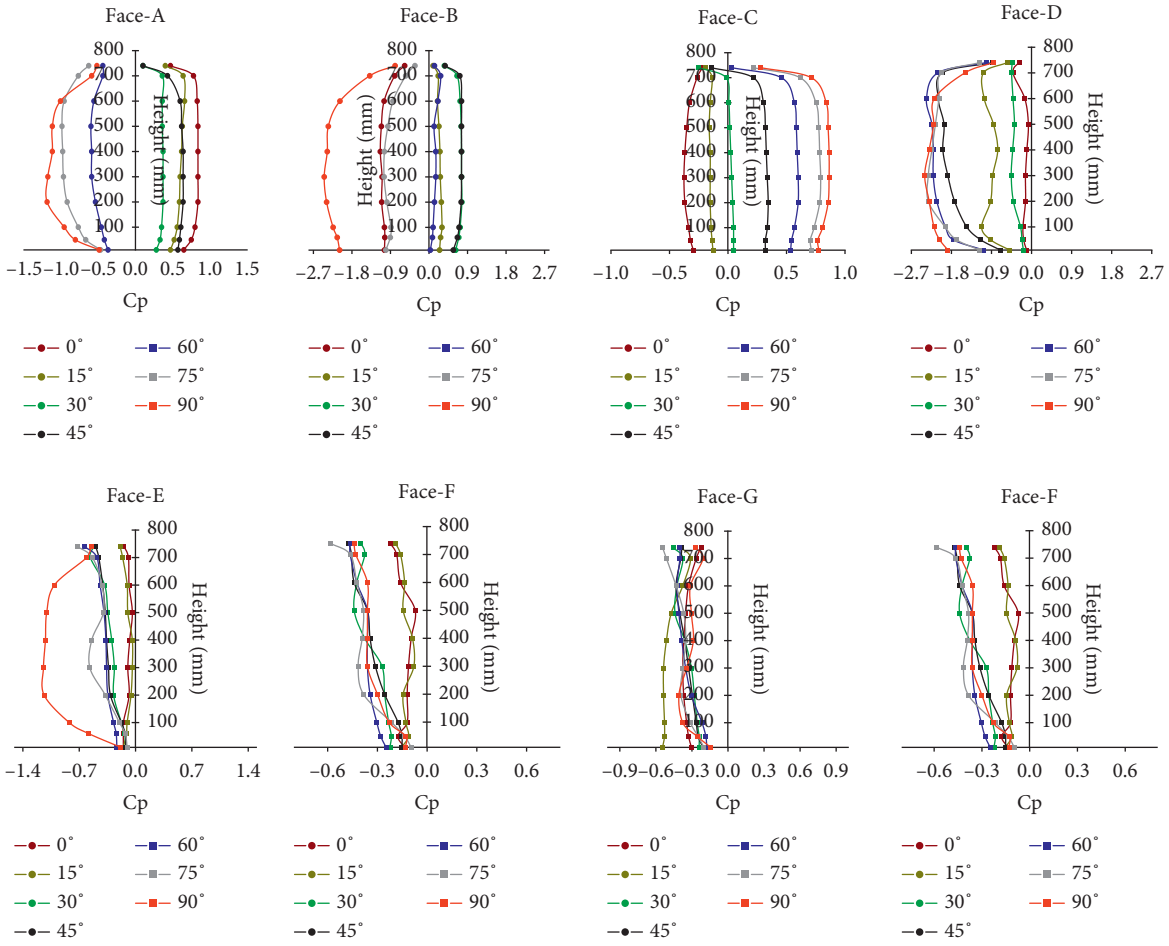


Figure 17: Pressure variation along the vertical centreline for all faces of Model-D at various wind incidence angles.

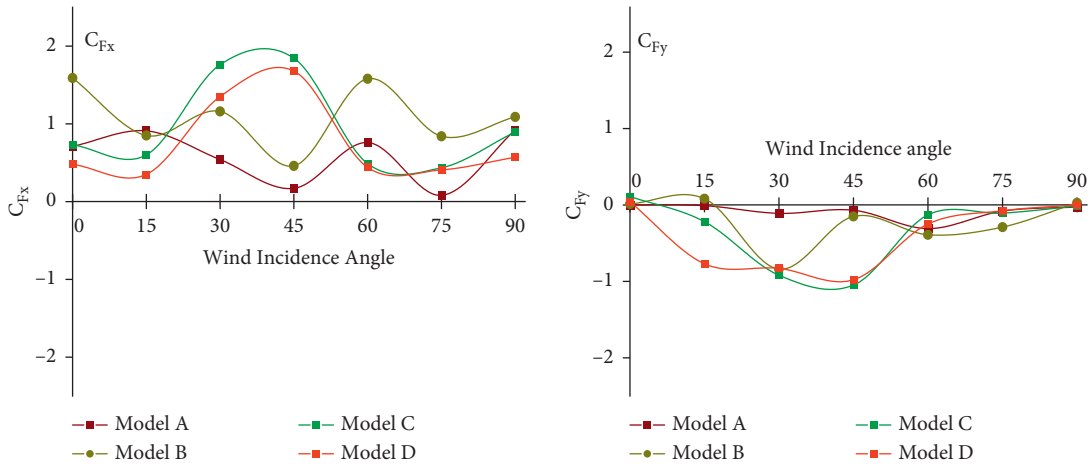


Figure 18: Variation of force coefficient at X-axis and Y-axis at various wind incidence angles.

the wind, and H is the height of the building model. C_{mx} is the moment coefficient of the building model in the X-direction, and C_{my} is the moment coefficient of the building model in the Y-direction.

3.5. Face Average $C_{p,mean}$. Coefficient of pressure is a non-dimensional ratio of wind-induced pressure on a building to the velocity pressure of the wind speed at the reference height. Pressure coefficient depends on the shape of the

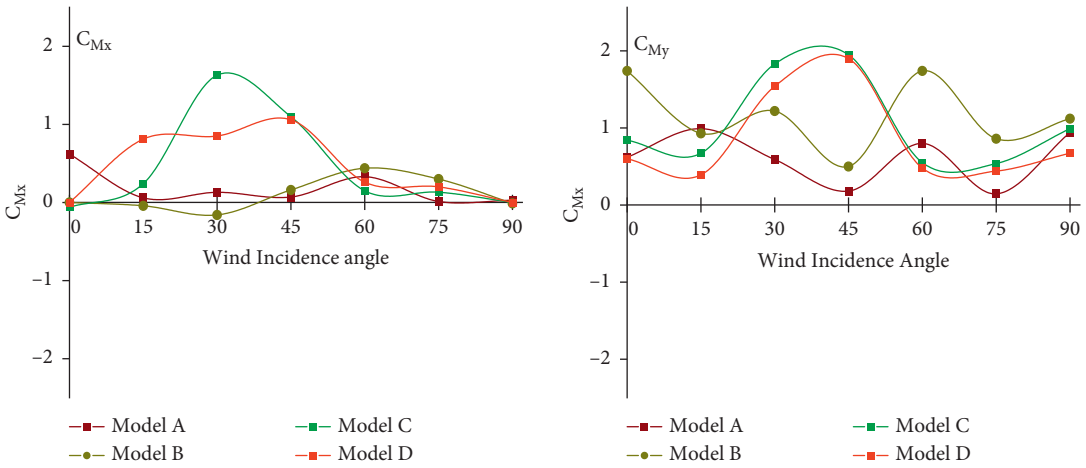


Figure 19: Variation of moment coefficient at X-axis and Y-axis at various wind incidence angles.

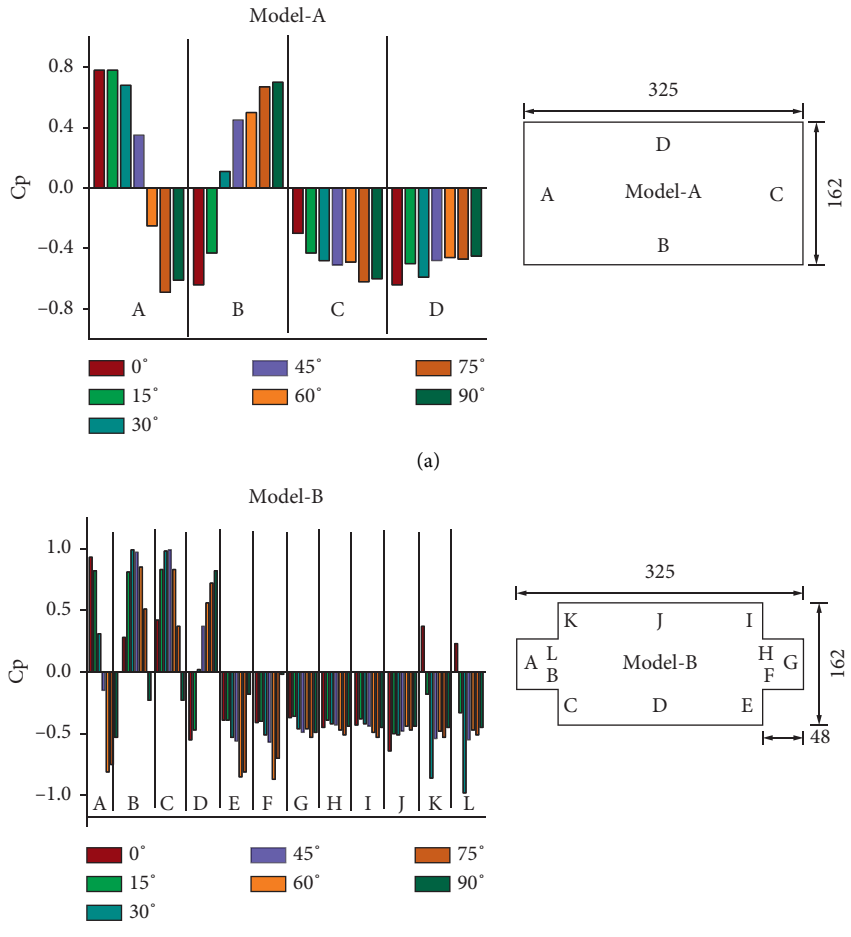


Figure 20: Continued.

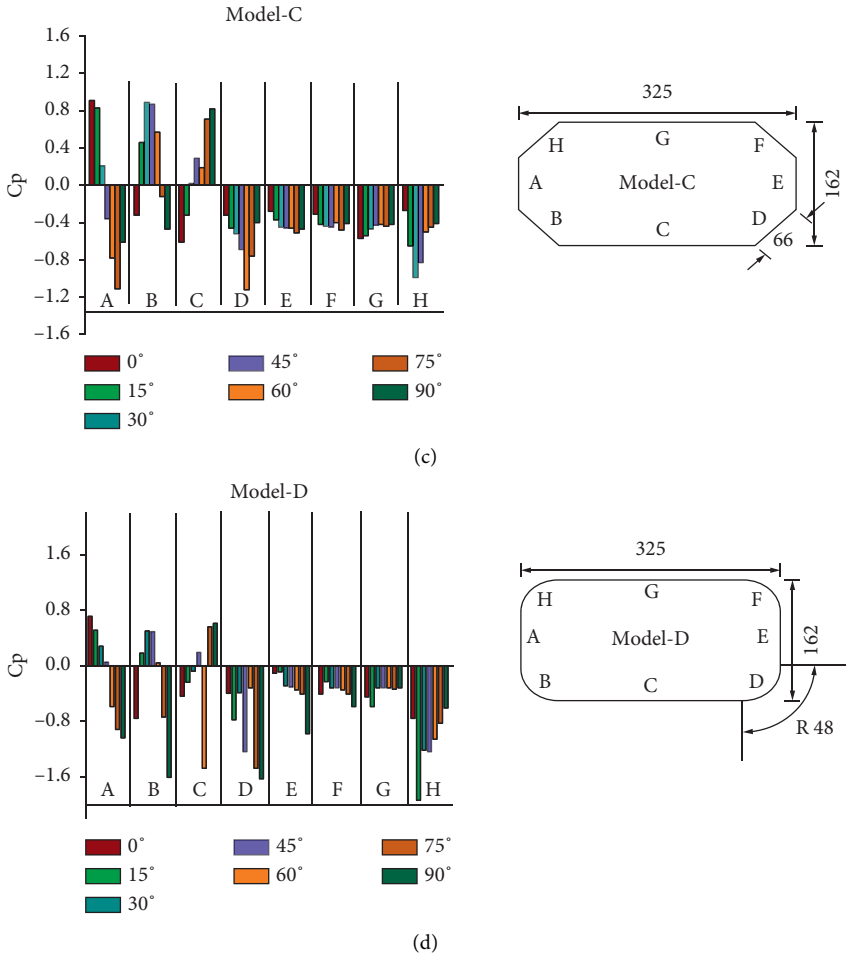


FIGURE 20: Mean pressure coefficient for all faces of (a) Model-A, (b) Model-B, (c) Model-C, and (d) Model-D at different wind incidence angles.

building wind incidence angle and the profile of the wind velocity. Figure 20 presents mean pressure coefficients for all building models at various wind incidence angles of 0° to 90° at an interval of 15°. Model-A, face-A, has maximum positive mean pressure of 0.78 and then it starts to decrease till 45° wind to 0.35. Model-A is attacked by various wind skew angles, that is, beyond 45°, $C_{p,\text{mean}}$ changes on face-A from positive to negative. Model-A, face-B, has negative $C_{p,\text{mean}}$ at 0° and 15° wind attack; after 15° it increases with respect to wind incidence angles, that is, changing with respect to wind incidence angle. In case of face-C, $C_{p,\text{mean}}$ is maximum for 0° wind attack and maximum for 45° wind. For face-C, $C_{p,\text{mean}}$ is increasing in negative with wind incidence angle. For face-D, maximum $C_{p,\text{mean}}$ is at 0° wind attack and minimum $C_{p,\text{mean}}$ is at 90° wind attack. $C_{p,\text{mean}}$ of Model-B, Model-C, and Model-D *are* represented in Figure 20.

4. Conclusions

This paper presents the wind loads effect on the various plan-shaped tall buildings by varying the corner shapes. Numerical simulation is performed using ANSYS CFX, k-ε turbulence model at 1 : 200 length scale. Face average C_p is compared and found in a range with various international standards and experimental studies for modelling in the numerical simulation. A clear view of streamlines has been presented at different wind incidence angles. The notable outcomes of the current study are summarized as follows:

(i) The validation study demonstrated very prominent results in line with various international standards and wind tunnel results; however, several upgradations are required in codal provision as the wind data is not updated with time.

(ii) Model-A without any corner configuration has maximum C_p at face-A for 0° wind incidence angle; after incorporation of the various corners in building model, it has been observed that Model-B is subjected to maximum windward pressure at 0°, that is, 20% greater wind impact than Model-A, and Model-C and Model-D have chamfer and fillet corners, respectively; however, less effect of wind on windward side is observed in case of Model-D. Furthermore, it is clear that incorporating the round corner helps to reduce the effect of wind flow on building.

(iii) In case of 15° wind incidence angle, windward face-A experiences similar wind response for Model-B and Model-C; however, Model-C with cross corner helps to reduce overall wind forces on building as compared to Model-B having cut corner configuration.

(iv) Drag force coefficients are compared between all the buildings models having various corner configurations; it was found that Model-C and Model-D experienced the same drag forces with little variation. Minimum drag forces in X-axis were experienced by Model-A at 75° wind incidence angle.

(v) Although lift force coefficient is calculated for all models and it has been observed that maximum lift force is measured on Model-C at 45° wind incidence angle, it was also concluded that the building having corner configuration has experienced greater lift force as compared to the building without corner configurations. Lift and drag forces are majorly affected by providing the corners on buildings.

(vi) Moment of X and Y axes is plotted with respect to drag and lift forces and found identical to base forces plotted against X and Y axes, respectively. It can be useful while calculating the overall response of buildings having different corner configuration. Building with normal rectangular shape experienced lowest CMY as compared to other building models having different corners.

Data Availability

All data, models, and code generated or used during the study appear in the submitted article. Data are available in the form of contour plots and graphs.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to express their sincere gratitude to Delhi Technological University, Delhi, India, for providing research facilities and funding to conduct the research work.

References

[1] P. A. Irwin, "Bluff body aerodynamics in wind engineering," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 96, no. 6-7, pp. 701–712, 2008.

[2] G. Solari, *Wind Actions and Effects on Structures*, Springer, New York, NY, USA, 2019.

[3] E. Simiu and D. Yeo, *Wind Effects on Structures*, John Wiley & Sons, Hoboken, NJ., USA, 4TH edition, 2019.

[4] B. S. Taranath, *Wind and Earthquake Resistant Buildings*, Rouledge, England, UK, 2004.

[5] C. C. B. Ted Stathopoulos, *Wind Effects on Buildings and Design of Wind-Sensitive Structures*, Springer, New York, NY, USA.

[6] IS-875 (Part-3), *Indian Standard Design Load (Other Than Earthquake) for Building and Structures-Code of Practice*, BIS, Old Delhi, India, 2015.

[7] AS/Nzs11702, *Structural Design Actions - Part 2*, Wind actions. Standards Australia/Standards New Zealand, Sydney, Austrlia, 2011.

[8] Asce716, *Minimum Design Loads and Associated Criteria for Buildings and Other Structures*, Structural Engineering Institute of the American Society of Civil Engineering, Reston, 2017.

[9] GB 50009-2001, *National Standard of the Peoples' Republic of China, Ministry of Construction (MOC) and the Genral Administration of Quality Superivison, Inspection and Quarantine (GAQSIQ) of the People's Republic of China*, Beijing, China, 2002.

[10] H. Kong Building Department, *Code of Practice on Wind Effects in Hong Kong 2019*, Building Department Hong Kong, Hong Kong First Issue, 2019.

[11] R. Sheng, L. Perret, I. Calmet, F. Demouge, and J. Guilhot, "Wind tunnel study of wind effects on a high-rise building at a scale of 1:300," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 174, pp. 391–403, 2018.

[12] X. Sun, H. Liu, N. Su, and Y. Wu, "Investigation on wind tunnel tests of the Kilometer skyscraper," *Engineering Structures*, vol. 148, pp. 340–356, 2017.

[13] P. Irwin, J. Kilpatrick, and A. Frisque, "Friend or foe, wind at height," *CTBUH 2008, 8th World Congr. - Tall Green Typology a Sustain. Urban Futur. Congr. Proc.*, pp. 336–342, 2008.

[14] R. Raj and A. K. Ahuja, "Wind loads on cross shape tall buildings," *Journal of Academia Industrial Resear*, vol. 2, no. 2, pp. 111–113, 2013.

[15] B. Bhattacharyya and S. K. Dalui, "Experimental and numerical study of wind-pressure distribution on irregular-plan-shaped building," *Journal of Structural Engineering*, vol. 146, no. 7, Article ID 04020137, 2020.

[16] S. K. Nagar, R. Raj, and N. Dev, "Experimental study of wind-induced pressures on tall buildings of different shapes," *Wind Struct. An Int. J.*vol. 31, no. 5, pp. 441–453, 2020.

[17] E. K. Bandi, Y. Tamura, A. Yoshida, Y. Chul Kim, and Q. Yang, "Experimental investigation on aerodynamic characteristics of various triangular-section high-rise buildings," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 122, pp. 60–68, 2013.

[18] R. Raj, S. Jha, S. Singh, and S. Choudhary, "Response analysis of plus shaped tall building with different bracing systems under wind load," *International Journal of Advanced Research in Engineering & Technology*, vol. 11, no. 3, pp. 371–380, 2020.

[19] L. Carassale, A. Freda, and M. Marrè-Brunenghi, "Experimental investigation on the aerodynamic behavior of square cylinders with rounded corners," *Journal of Fluids and Structures*, vol. 44, pp. 195–204, 2014.

[20] H. Tanaka, Y. Tamura, K. Ohtake, M. Nakai, and Y. Chul Kim, "Experimental investigation of aerodynamic forces and wind pressures acting on tall buildings with various unconventional configurations," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 107-108, no. 108, pp. 179–191, 2012.

[21] J. He and C. C. S. Song, "A numerical study of wind flow around the TTU building and the roof corner vortex," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 67-68, no. 68, pp. 547–558, 1997.

[22] S. Pal, R. Raj, and S. Anbukumar, "Comparative study of wind induced mutual interference effects on square and fish-plan shape tall buildings," *Sādhanā*, vol. 46, no. 2, 2021.

[23] S. Pal and R. Raj, "Evaluation of wind induced interference effects on shape remodeled tall buildings," *Arabian Journal for Science and Engineering*, vol. 46, no. 11, pp. 11425–11445, Article ID 0123456789, 2021.

[24] S. Pal, R. Raj, and S. Anbukumar, "Bilateral interference of wind loads induced on duplicate building models of various shapes," *Latin American Journal of Solids and Structures*, vol. 18, no. 5, 2021.

[25] S. K. Nagar, R. Raj, and N. Dev, "Proximity effects between two plus-plan shaped high-rise buildings on mean and RMS pressure coefficients," *Scientia Iranica*, 2021.

[26] B. Bhattacharyya and S. K. Dalui, "Investigation of mean wind pressures on "E" plan shaped tall building," *Wind Structures An International Journal*, vol. 26, no. 2, pp. 99–114, 2018.

[27] R. Raj, T. Rana, T. Anchalia, and U. Khola, "Numerical study of wind excited action on H Plan-shaped tall building," *International Journal on Emerging Technologies*, vol. 11, no. 3, pp. 591–605, 2020.

[28] T. Uchida and Y. Ohya, "A numerical study of stably stratified flows over a two-dimensional hill - Part I. Free-slip condition on the ground," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 67-68, no. 68, pp. 493–506, 1997.

[29] D.-h. Yu and A. Kareem, "Numerical simulation of flow around rectangular prism," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 67-68, no. 68, pp. 195–208, 1997.

[30] P. Sanyal and S. K. Dalui, "Effects of courtyard and opening on a rectangular plan shaped tall building under wind load," *International Journal of Advanced Structural Engineering*, vol. 10, no. 2, pp. 169–188, 2018.

[31] R. Paul and S. Dalui, "Shape optimization to reduce wind pressure on the surfaces of a rectangular building with horizontal limbs," *Periodica Polytechnica: Civil Engineering*, vol. 65, no. 1, pp. 134–149, 2021.

[32] A. Okajima, D. Yi, A. Sakuda, and T. Nakano, "Numerical study of blockage effects on aerodynamic characteristics of an oscillating rectangular cylinder," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 67-68, no. 68, pp. 91–102, 1997.

[33] R. Paul and S. K. Dalui, "Prognosis of wind-tempted mean pressure coefficients of cross-shaped tall buildings using artificial neural network," *Periodica Polytechnica: Civil Engineering*, vol. 64, no. 4, pp. 1124–1143, 2020.

[34] S. Hajra and S. K. Dalui, "Numerical investigation of interference effect on octagonal plan shaped tall buildings," *Jordan Journal of Civil Engineering*, vol. 10, no. 4, pp. 462–479, 2016.

[35] N. Gaur and R. Raj, "Aerodynamic mitigation by corner modification on square model under wind loads employing CFD and wind tunnel," *Ain Shams Engineering Journal*, vol. 283, 2021.

[36] A. K. Bairagi and S. K. Dalui, "Estimation of wind load on stepped tall building using CFD simulation," *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, vol. 45, no. 2, pp. 707–727, 2021.

[37] P. Sanyal and S. K. Dalui, "Effect of corner modifications on Y' plan shaped tall building under wind load," *Wind Structures ofAn International Journal*, vol. 30, no. 3, pp. 245–260, 2020.

[38] P. Sanyal and S. K. Dalui, "Effects of side ratio for "Y" plan shaped tall building under wind load," *Building Simulation*, vol. 14, no. 4, pp. 1221–1236, 2020.

[39] P. Sanyal and S. K. Dalui, "Effects of internal angle between limbs of "Y" plan shaped tall building under wind load," *Journal of Building Engineering*, vol. 33, Article ID 101843, 2021.

[40] P. Sanyal and S. K. Dalui, "Comparison of aerodynamic coefficients of various types of Y-plan-shaped tall buildings," *Asian Journal of Civil Engineering*, vol. 21, no. 7, pp. 1109–1127, 2020.

[41] J. Franke, A. Hellsten, H. Schlünzen, and B. Carissimo, *Guideline for the CFD Simulation of Flows in the Urban Environment: COST Action 732 Quality Assurance and Improvement of Microscale Meteorological Models*, COST Office, Brussels , Belgium, 2007.

[42] J. Revuz, D. M. Hargreaves, and J. S. Owen, "On the domain size for the steady-state CFD modelling of a tall building," *Wind and Structures*, vol. 15, no. 4, pp. 313–329, 2012.

[43] R. Raj, *Effects of Cross-Sectional Shapes on Response of Tall Buildings under Wind Loads*, Ph.D. Thesis, IIT Roorkee, Roorkee, India, 2015.

Word Sense Disambiguation using Cooperative Game Theory and Fuzzy Hindi WordNet based on ConceptNet

GOONJAN JAIN, Department of Applied Mathematics, Delhi Technological University, Delhi, India
D. K. LOBIYAL, School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, India

Natural Language is fuzzy in nature. The fuzziness of Hindi language was captured in the **Fuzzy Hindi WordNet (FHWN)**. FHWN assigned membership values to fuzzy relationships by consulting experts from various domains. However, these membership values need to be corrected. In the proposed work, we compute the membership values of fuzzy semantic relations using ConceptNet. Later, we perform WSD of Hindi text using cooperative game theoretic approach. We used the Shapley Value centrality measure where we predict which coalition of players (word senses) proves to be the most beneficial. We tested and compared our algorithm with the existing state-of-the-art approaches of Hindi on three datasets and results are better on all the three datasets. One more notable aspect is that the results are quite stable even if the fuzzy membership values of fuzzy graphs changes.

CCS Concepts: • **Computing methodologies** → **Information extraction**; *Vagueness and fuzzy logic*; *Theory of mind*; *Stochastic games*; • **Theory of computation** → *Graph algorithms analysis*;

Additional Key Words and Phrases: Word sense disambiguation, Shapley Value, Fuzzy Hindi WordNet, ConceptNet, cooperative game theory

ACM Reference format:
Goonjan Jain and D. K. Lobiyal. 2022. Word Sense Disambiguation using Cooperative Game Theory and Fuzzy Hindi WordNet based on ConceptNet. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 21, 4, Article 79 (February 2022), 25 pages.
https://doi.org/10.1145/3502739

1 INTRODUCTION

One of the essential tasks in **Natural Language Processing (NLP)** is **Word Sense Disambiguation (WSD).** It deals with understanding the true meaning of a word in the text. The efficiency of various NLP tasks like Machine Translation [Vickrey et al. 2005, Carpuat and Wu 2005, 2007; Chan et al. 2007], Sentiment Analysis [Rentoumi et al. 2009, Sumanth and Inkpen 2015, Hung and Chen 2016], and Information Retrieval [Zhong and Ng 2012, Schadd and Roos 2015] and the like relies on the efficiency of WSD.

Various approaches for WSD have been proposed in the literature. These approaches are categorized based on the process of training the system for disambiguation. The four conventional approaches for WSD are briefly described below:

This work was supported in part by the Council of Scientific and Industrial Research (CSIR), New Delhi, India with file no. 09/263(1018)/2014-EMR-I.
Authors' addresses: G. Jain, Department of Applied Mathematics, Delhi Technological University, Delhi, India; email: Goonjan_jain@hotmail.com; D. K. Lobiyal, School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, India; email: lobiyal@gmail.com.
ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.
© 2022 Association for Computing Machinery.
2375-4699/2022/02-ART79 $15.00
https://doi.org/10.1145/3502739

1. Dictionary- or Knowledge-Based Approaches. Dictionary- or knowledge-based approaches rely on dictionaries, thesauri, or lexical knowledge bases, without any corpus base. These approaches use information from one or many knowledge resources. The sense similarity scores are then computed to disambiguate words in the given text [Mihalcea, 2005]. Knowledge-based approaches are picking up the pace because of their ease of use as no training of corpus is required and the availability of different knowledge bases that capture different aspects of knowledge, e.g., PurposeNet [Sanghal et al. 2013], HowNet [Dong et al. 2010], MindNet [Vanderwend et al. 2005]. Graph-based algorithms have shown prominence to solve the problem of WSD [Sinha and Mihalcea 2007; Agirre et al. 2009].

2. Semi-Supervised Approaches. Semi-supervised approaches make use of unlabeled data for training. The approach is designed using information from a corpus. The corpus is split into two parts. One part is smaller in size and is labeled while the other part is vast and is unlabeled. This small amount of data is treated as seed data for processing [Taghipour, and Ng 2015; Gonen and Goldberg, 2016; Yuan et al. 2016].

3. Supervised Approaches. Supervised approaches use semantically annotated corpora to train the systems that are used to assign a correct sense to a word in the context. The performance of supervised approaches has been the best amongst all types of approaches. However, these approaches suffer from a drawback of using annotated corpora. For a resource-rich language, such as English, the corpora are available, but for other resource-scarce languages, such as Hindi, collecting sense-tagged corpora is itself a challenging task [Zhong and Ng 2010; Wang et al. 2014]. Almost all the machine learning algorithms have been tried including feature selection, parameter optimization, ensemble learning, and so on. Techniques used under supervised approaches are: Naïve Bayes Classification, based on Bayes Theorem; Decision tree to facilitate classification of rules over the training dataset; Support Vector Machine; and Neural networks.

4. Unsupervised approaches. Unsupervised methods abstain from training and instead work directly on raw unannotated corpora [Mihalcea 2005; McCarthy et al. 2007; Tripodi and Pelillo 2017]. Word embeddings have become important to represent words as fixed size vectors by taking the context into consideration. AutoExtend [Rothe and Schütze 2015, 2017] and Most Suitable Sense Annotation (MSSA) [Ruas et al. 2019] combine knowledge bases and word embeddings for WSD task.

The performance of supervised approaches has superseded other approaches in general domains. However, unsupervised and semi-supervised methods have started gaining attention from the research community as they have shown the ability to perform closer to the state-of-the-art supervised systems. Though we encounter various approaches for WSD of English text, minimal work has been done for WSD of Hindi text.

Hindi is one of the two official languages of India and is ranked fourth in the world with 615 million speakers, just after English.[1] In spite of that, Hindi is still in its infancy stage with respect to Natural Language Processing applications and research. For disambiguation of Hindi text, Hindi WordNet [Narayan et al. 2002, Bhattacharya 2010] was developed. Hindi WordNet is a sense inventory where words are organized as sets of synonyms, called *synsets*. Words in a synset are those words that bear similar meanings. The total number of unique words in Hindi WordNet

[1] http://www.ethnologue.com.

are 98,651, and total numbers of synsets are 38,687. Synsets and words are related with one another with various semantic and lexical relations such as is-a, kind-of, and part-whole relations. The semantic relations in Hindi WordNet includes synonymy, hyponymy, hypernymy, meronymy, holonymy, entailment, and the like.

Usage of natural language is fuzzy in nature. Taking the fuzzy nature of Hindi language into consideration, Jain and Lobiyal [2016] introduced **Fuzzy Hindi WordNet (FHWN)**. It emphasizes that it is not always necessary that word senses are either connected or not connected at all. Word senses may be related to one another with some degree. The degree with which word senses relate with one another is represented by a membership value ranging between 0 and 1. In the same paper, Jain and Lobiyal [2016] also proposed a fuzzy graph centrality based WSD approach for Hindi text using the FHWN. However, we identified two major gaps in the development of FHWN and the WSD approach applied thereafter. We discuss these gaps in 1.1.

1.1 Gaps in the Fuzzy Hindi WordNet-Based Approach for WSD

We identified two major gaps in the approach proposed by Jain and Lobiyal [2016] for WSD of Hindi text using FHWN. The first gap is with the approach used for assignment of fuzzy membership values to the fuzzy relationships and second gap is with the measures used for WSD of Hindi text. We describe these gaps below:

1. Unreliable fuzzy membership values.

Jain and Lobiyal [2016] took experts' opinions into consideration to assign membership values for fuzzy semantic relations. They consulted selected renowned Hindi lexicographers, experts in linguistics, computer science, fuzzy logic, and Hindi language. Based on the experts' recommendations, membership values were assigned to the relationship between pairs of words. Words that were related in Hindi WordNet were kept intact with membership value of one (1).

With the human-based method of assignment of fuzzy membership values, we observed the following limitations:

(a) **Validity.** Humans can only state if the words are related to one another or not. Quantifying this relationship is beyond humans' capability.
(b) **Certainty.** The values assigned by experts are subjected to the opinion of the expert. Therefore, they are judgment based rather than probability. For example, the relationship strength between अलमारी (alamārī) and रैक (raika) was assigned a numerical value of 0.7. However, it is difficult for experts to explain why the value is not 0.68 or 0.72. The values assigned by experts are arbitrary. (*English Translations:* अलमारी – almirah, रैक – rack)
(c) **Knowledge.** For assignment of membership values, only a few experts can be consulted. Thus, the membership values are based on the perceptions of those experts and are subjective. However, for initial investigation, approach may be justified.
(d) **Language.** To develop the fuzzy WordNet for any other language, we need to perform the same exercise again. Bringing together the experts from different fields is quite costly and time consuming, too.

The above limitations were resolved using our proposed algorithm:

(a) **Validity.** The membership value for the fuzzy relationship between two words is based on uncertainty of information. If the information between two words is concentrated in fewer intermediate nodes in the graph, then there is less uncertainty, thus high membership value. However, in case the information of the relationship between two words is diffused

among a large number of nodes, this means that there is more uncertainty and thus lower fuzzy membership value. The details about the proposed algorithm are given in 4.2.

(b) **Certainty.** The values are calculated using a computer program; thus, the output remains the same. There is no uncertainty in the computed values. The values can change only if there is change in the ConceptNet, otherwise the values will remain static.

(c) **Knowledge.** The knowledge contained in ConceptNet has been collected from a variety of resources, including crowd-sourced resources, such as Wiktionary[2] and Open Mind Common Sense [Singh et al. 2002]; purpose-driven games, such as Verbosity [Ahn et al. 2006]; and expert-created resources, such as WordNet [Miller 1995, Fellbaum 1998]. Thus, the domain and the knowledge are vast and more accurate.

(d) **Language.** The task for development of fuzzy WordNet of any other language can be undertaken by using the proposed algorithm. ConceptNet supports 10 core languages and 68 common languages.

2. Dangling behavior of graph centrality measures with varying fuzzy membership values.

Change in fuzzy membership values should change the value of the centrality measure. It is desirable. Therefore, incorrect fuzzy membership value will result in an incorrect centrality measure. We performed numerous experiments to confirm this behavior. We add illustrations to showcase the dangling behavior in Section 6.

We acknowledge the fact that the fuzzy membership values are still not stable and we have a long way to reach that stage. Thus, we searched for a metric that does not change its behavior drastically with the change in fuzzy membership values. The proposed algorithm uses Game-Theory-based Shapley Value as a centrality measure. The group graph centrality measure is used to compute the contribution by each coalition of players (in our case, *words*). Thereafter, average marginal contribution by a player that takes over all possible ways to form the grand coalition is considered. The Shapley Value is not completely immune to the incorrect fuzzy membership values. However, Shapley Value minimizes the impact of change in membership values of fuzzy relationships because it takes only the marginal contribution. The results using Shapley Value centrality measures are promising and stable. The stability of the Shapley Value is demonstrated in the Section 7.5.

1.2 Contributions of the Proposed Work

In the proposed work, we perform two tasks. The first task is the assignment of fuzzy membership values to the fuzzy relationships between pair of words. The second task is to perform WSD of Hindi text. The proposed approach of assigning membership values is based on the principle of *Lexical Priming* [Hoey 2005]. Lexical Priming was proposed by Hoey in his book titled *Lexical Priming* [Hoey 2005]. The idea is that words never exist in isolation, instead they are used in correlation with each other. Every usage of a word is loaded with the context and co-texts in which it is encountered. Thus, our knowledge of the word includes the fact that it co-occurs with certain other words in some kinds of context [Hoey 2005]. This theory has been validated and utilized in the literature, too [Pace-Sigge 2013; Rapp and Zock 2014]. We have followed this principle in establishing relatedness between two words. We fetch a ConceptNet subgraph consisting of those two words as the end nodes amongst which we are seeking the relationship if it exists. Using this graph, membership values are assigned using graphical connectivity measures. To deduce the correct sense, we use cooperative game-theory-based Shapley Value as the centrality measure. Shapley value is used to identify the most influential sense of a word in the given context.

[2] https://en.wiktionary.org/wiki/Wiktionary:Main_Page.

1.3 Organization of this Article

The rest of this article is structured as follows: Section 2 gives a glimpse of some of the approaches proposed for WSD in recent years. The preliminary explanation of essential topics is given in Section 3. The detailed methodology to assign membership values to the relationship between two words is given in Section 4. The detailed design and implementation methodology of WSD algorithm is explained in Section 5. Section 6 contains elaborate illustrations of the proposed algorithm. The results of the evaluation of our work by undertaking various experiments and its comparison with other available methods are reported in Section 7. Finally, in Section 8, we conclude the work along with the possible future extensions of the work.

2 RELATED WORK

In this section, we give a brief review of the work done in the literature for WSD.

Various dictionary-based and unsupervised approaches have been proposed in the literature that tries to identify the actual sense of a word. The actual sense is identified by exploiting the information derived from the context of a word. Algorithms proposed by Lesk [1986], Banerjee and Pederson [2002], Patwardhan et al. [2003], and Basile et al. [2014] exploited pairwise semantic similarity among ambiguous words and the words in their context.

Graph-based approaches have gained the attention of researchers in the last few years. It is because of the native ability of graphs to organize contextual information as well as to compute relations between word senses. This ability enables researchers to extract structural as well as semantic properties of the text. Few of the research work based on this approach in the literature are Navigli and Lapata [2007, 2010], Sinha and Mihalcea [2007], Manion and Sainudiin [2014], Agirre et al. [2014], Moro et al. [2014], Chaplot et al. [2015], and Pina and Johansson [2016].

We see minimal work in the literature that uses ConceptNet for WSD. Three of the approaches based on ConceptNet are Blending [Havasi et al. 2010], ExtGM [Chen and Liu, 2011], and implicit-information-based WSD [Jain and Lobiyal, 2020]. Blending [Havasi et al. 2010] uses a common-sense reasoning technique, called *blending*, that combines knowledge from SemCor, WordNet, ConceptNet, Extended WordNet, and their ambiguated versions. ExtGM [Chen and Liu, 2011] first disambiguates concepts in ConceptNet. It tags each concept in a given assertion with one of the senses. It then combines ConceptNet and WordNet to create an enriched knowledge base. The algorithm proposed by Jain and Lobiyal [2020] finds implicit terms that connect words of a sentence using ConceptNet. Thereafter, set intersection of the sentence profile and word sense profile is computed. The intersection set with the highest cardinality is the deduced sense of the ambiguated word.

There is a gain in momentum in the amalgamation of NLP and Game Theory. Tripodi, Pelillo [2017], proposed a game theoretic approach for WSD. Their approach is based on non-cooperative game theory, whereas our proposed approach is based on cooperative game theory. We follow the principle that all the words in a sentence collaborate with each other to deduce the correct sense for the sentence as well as their individual sense.

Though Hindi being the fourth largest used language in the world, the availability of digital resources is minimal. Therefore, Hindi is considered as a scarce resource language, and not much work has been done for WSD of Hindi text. Various Lesk based approaches have been proposed for WSD of Hindi text [Singh, Siddiqui 2014, Gautam, Sharma 2016, Sajini, Kallimani 2017, Tripathi et al. 2020]. We have described each of them individually below.

Singh and Siddiqui [2014] proposed a Lesk-based algorithm for Hindi WSD task and to evaluate the effects of semantic relations on the task. The first step of creation of a context vector of an ambiguous word was done by extracting words appearing in $\pm$ n window size, where the ambiguous

word is in the middle of the window. This vector defines the context of the ambiguous word. To perform WSD, sense definition vector is created for every ambiguous word by extracting words occurring in the sense definitions that include synsets, their glosses, and semantic relations. Then, in the algorithm, a match is performed between the context vector and the sense definitions vector. The result of the match is a numerical value that takes the frequency of the matching words where the ambiguous word is not counted in the matching. The sense with the maximum score is declared as a correct sense.

Gautam and Sharma [2016] define the context of an ambiguous word as consisting of bigrams and trigrams. Then an overlap between these bigrams and trigrams and the word sense id is computed. The sense with the highest overlap is declared as the correct sense. Sajini and Kallimani [2017] redefined the context of a sentence to include all the content words of the sentence, along with the words in the sentence preceding it and the one following it. Then they found the overlap between the sentence context and the WordNet senses. Tripathi et al. [2020] used a novel scoring method to assign a sense score to each token of the Hindi sentence. The sense score is calculated based on the gloss, hypernym, hyponym, and synonym of the combinations of different sense of tokens.

The approach proposed by Tayal et al. [2015] converted the text in the Hindi language to Hyperspace analogue to language vectors. Then Fuzzy c-mean clustering algorithm is applied to form clusters denoting the various contexts in which polysemous words may occur. Bhingardive and Bhattacharya [2016] proposed an approach for bilingual WSD using IndoWordNet, which is a linked structure of WordNets of major Indian languages. They followed two unsupervised approaches. In the first approach, IndoWordNet was used for bilingual WSD. It employed the expectation maximization algorithm to find the sense distribution. In the second approach, they used the word and sense embeddings to find the most frequent sense.

Similar to the English language, graph-based approaches are being extensively used for WSD of Hindi text too. Some of the proposed approaches are Jain and Lobiyal [2014, 2015, 2016]. In the approach proposed by Jain and Lobiyal [2014], a semantic graph for each interpretation of the given sentence is constructed by establishing semantic relations between the pairs of words present in the sentence. The cost of the spanning tree for each semantic graph is calculated. The interpretation with the minimum cost for the spanning tree is identified as the correct interpretation.

Jain and Lobiyal [2015] proposed an approach in which they create the interpretation graph for each interpretation of the sentence. Then the network agglomeration for all the interpretation graphs is computed, and the highest value graph is declared as the correct interpretation. Jain and Lobiyal [2016] proposed Fuzzy Hindi WordNet (FHWN). FHWN is an extension of Hindi WordNet in which few fuzzy semantic relations were added to reflect the fuzziness of language usage. They also proposed fuzzy centrality measures. The WSD of Hindi text was performed by computing fuzzy centrality measures on the sentence graph.

Word-embedding-based approach for WSD was proposed by Kumari and Lobiyal [2019]. They used two word2vec architectures–skip gram and continuous bag-of-words models–to develop word embeddings. The correct sense of a word is deduced using cosine similarity.

We identified research gaps while doing the literature survey. We see massive potential in Fuzzy Hindi WordNet-based approaches. Since the change in values of fuzzy membership changes the performance of the application using it. Therefore, in the proposed approach, we try to find membership values using a lexical knowledge base. We also see a considerable potential in game theory to be used for various NLP tasks. Thus, in the proposed work, we used game-theory-based approach for WSD of Hindi text.

3 PRELIMINARIES

In this section, we give a preliminary view of the topics required for the readers for better understanding of the proposed work. First, we discuss Fuzzy Hindi WordNet. Next, we give a brief introduction of ConceptNet. Later, we discuss basic nuances of Cooperative Game Theory. In the last, we explain Shapley Value and its properties.

3.1 Fuzzy Hindi WordNet

Fuzzy Hindi WordNet is a fuzzy graph where nodes represent synsets and edges represent the fuzzy relationship between two synsets. Fuzzy Hindi WordNet was proposed by Jain and Lobiyal [2016]. The graph is a weighted graph where the weight of an edge depicts the strength of the relationship between two synsets and weights can vary between 0 and 1. The strength of 0 indicates that the synsets are not related to each other whereas the value of 1 depicts that the two synsets have a classical relationship with each other. Since classical logic is a specialization of fuzzy logic, thus, all relations in Hindi WordNet are included in Fuzzy Hindi WordNet too. The value of membership is unity. Synsets in Fuzzy Hindi WordNet are related to one another with several lexical and semantic relations. These relations help to organize the lexical knowledge base. There are in total 17 relations defined in Fuzzy Hindi WordNet. A few of them are listed below:

1. Fuzzy Association Relation.

Two synsets are related to each other by fuzzy association relation if the synsets have approximately similar meanings. A few examples of fuzzy association are (अलमारी (alamārī), रैक (raika), 0.7), (लड़की (laḍakī), औरत (aurata), 0.8) and so on. (*English Translations:* अलमारी – almirah, रैक – rack, लड़की – girl, औरत – woman.)

2. Fuzzy Hypernymy and Fuzzy Hyponymy.

Two synsets are related to each other by fuzzy hypernymy/hyponymy relation if they share the partial superset/subset relationship with each other. For example, (दुखी (dukhī), परेशान (parēśāna), 0.9), (सज्जन (sajjana), दयालु (dayālu), 0.8), and so on. (*English Translations:* दुखी – miserable, परेशान – annoyed, सज्जन – gentle, दयालु – kind.)

3. Fuzzy Meronymy and Fuzzy Holonymy.

The partial "part-whole" relationship between synsets is captured by fuzzy meronymy and fuzzy holonymy relations. For example, (फल (phala), पेड़ (pēḍa), 0.7), (रेडियो (rēiḍayō), कार (kāra), 0.8) and so on. (*English Translations:* फल – fruit, पेड़ – tree, रेडियो – radio, कार – car.)

4. Fuzzy Ability Link.

Fuzzy Ability Link specifies the inherited features of a nominal concept. For example, (मगरमच्छ (magaramaccha), तैरना (tairanā), 0.7), (मछली (machalī), तैरना (tairanā), 1.0) and so on. (*English Translations:* मगरमच्छ – crocodile, मछली – fish, तैरना – swim.)

3.2 ConceptNet

ConceptNet is a multilingual knowledge base. Words and phrases are related to each other via common-sense. The knowledge has been collected from a variety of resources, including crowd-sourced resources, such as Wiktionary and Open Mind Common Sense [Singh et al. 2002]; purpose-driven games, such as Verbosity [Ahn et al. 2006]; and expert-created resources, such as WordNet [Miller 1995, Fellbaum 1998]. ConceptNet is a network of labeled nodes and edges. The nodes,



Fig. 1. Single edge of ConceptNet [Jain and Lobiyal, 2020].

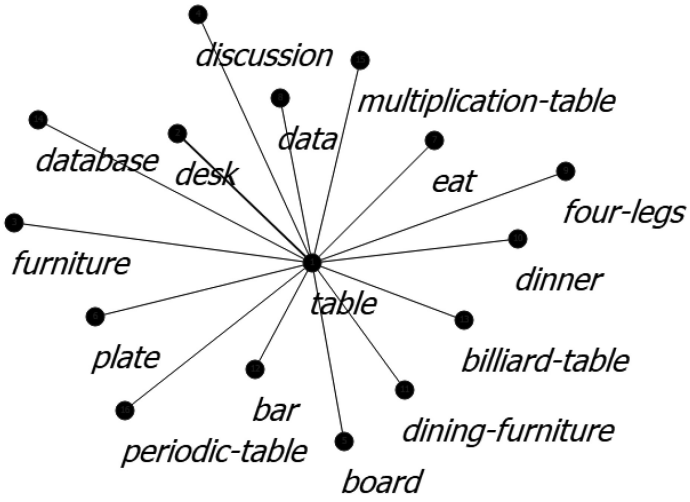


Fig. 2. Example subgraph of ConceptNet with "Table" as an ambiguous word [Jain and Lobiyal, 2020].

or concepts, are words, word senses, and short phrases in a variety of different languages. These concepts are connected with a particular relation, which is represented as an edge. The proposition expressed by a relation between two concepts is called an assertion. Every object in ConceptNet has a **Uniform Resource Identifier (URI)**, which is structured similar to a path. URIs are meaningful, and we can tell what the object is by looking at its URI [Jain and Lobiyal, 2020].

Figure 1 shows a single edge connecting two nodes where strings /c/en/car, /c/en/drive, and /r/UsedFor are URIs from ConceptNet. Figure 2 illustrates a subgraph of ConceptNet with *Table* as an ambiguous word. The *table* is connected to words belonging to different senses of its usage. The *table* is connected to *furniture, periodic table, eat,* and *discussion*, which belong to different senses of the word.

3.3 Cooperative Game Theory

Game Theory is concerned with the behavior of how the decisions of decision-makers are affected by each other [Von Neumann and Morgenstern, 1944]. It is more rational than psychological or sociological. The applicability of the game theory is enormous and varied. It can be applied to all interactive situations of sociology, computer science, economics, and so on.

Game theory can be broadly classified as *non-cooperative* and *cooperative game theory*. A non-cooperative game is a game with competition between individual players. In non-cooperative game theory, we try to predict each player's individual strategies and payoffs and to find Nash equilibria, whereas, in a cooperative game, competition is among different groups of players or coalitions with specific external enforcement of cooperative behavior. The goal of Cooperative Game Theory is to predict the different coalitions that may be formed, the actions that these coalitions take, and the resulting collective payoff. Every coalition is specified a value. A simple coalitional game is one in which the payoffs for different coalitions are either 0 or 1, i.e., a coalition is either "winning" or "losing".

In a cooperative game, a value is specified for every coalition. The coalition game consists of *N*, a finite set of players, also called the *grand coalition*. For every coalition, we have a function called the *characteristic function* denoted by $v{:}2^N \rightarrow \mathbb{R}$. This function is defined from a set of possible coalitions to a set of payments. The base condition for this function is $v(\emptyset) = 0$, i.e., an empty coalition has a value of 0. The function describes how much collective payoff a set of players can gain by forming a coalition, and the game is sometimes called a *value game* or a *profit game*. The characteristic function *v* has the following meaning: For a coalition of players *S*, *v(S)* gives the total expected sum of payoffs obtained by the cooperation of members of *S*. *v(S)* is also called *worth of coalition S*.

Properties for characteristic functions:

1. **Superadditivity.** Characteristic functions are often assumed to be superadditive [Owen 1995]. The superadditivity of characteristic function implies that sum of the coalitions' separate values should be greater than or equal to the union of disjoint coalitions.

$$v(A \cup B) \geq v(A) + v(B) \, whenever A, B \subseteq N satisfy A \cap B = \emptyset \tag{1}$$

2. **Monotonicity.** Larger coalitions gain more:

$$S \subseteq T \Rightarrow v(S) \leq v(T) \tag{2}$$

Once the grand coalition *N* is formed, the challenge is to allocate the payoff *v(N)* amongst the players. The allocation to each player is represented by a *solution concept* or a *payoff vector*, a vector $x \in \mathbb{R}^N$. Various solution concepts have been proposed by researchers. One of the most prevalent is the Shapley Value.

3.4 Shapley Value

The *Shapley value* is a payoff vector. It was introduced by Lloyd Shapley [Shapley 1953]. It distributes the total payoff of the grand coalition among the players of the cooperative game. The Shapley Value model is defined as the average marginal contribution by a player that takes over all possible ways to form the grand coalition. As the coalition begins, it starts to add more players one by one. This value ensures that all players are gaining as much as they should or even more, depending on their input.

The setup is as follows: Players cooperate with each other and form a coalition. In return, players obtain a gain from their cooperation. It happens that some players contribute more to the coalition than others. Also, players possess different bargaining power. The problem is to distribute the payoff of coalition amongst the players as per their contribution. This also depicts the importance of each player in the coalition. The Shapley value provides one possible solution to this problem. The Shapley value is one of the ways to distribute the total payoff of coalition amongst the players as per their contribution. According to the Shapley value, the amount that player *i* gets given in a coalition game *(v, N)* is

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} \left(v(S \cup \{i\}) - v(S)\right) \tag{3}$$

where *N* is the total number of players and the sum extends over all subsets *S* of *N* not containing player *i*. The formula works like this: One actor is added at a time in a coalition, and the fair compensation asked by the player is $v(S \cup \{i\}) - v(S)$. The payoff for each player is the average of its contribution over the possible different permutations in which the coalition can be formed. The Shapley value is characterized by the following desirable properties:

1. **Efficiency.** The total gain of the grand coalition is fully distributed amongst the players.
$$\sum_{i \in N} \phi_i(v) = v(N) \tag{4}$$
2. **Symmetry.** Two players i and j are considered equivalent if for every subset S of N, where both i and j are not in S, $v(S \cup \{i\}) = v(S \cup \{j\})$. Then, the individual payoff for two equivalent actors is also equal.
$$\phi_i(v) = \phi_j(v) \tag{5}$$
3. **Linearity.** If two coalition games described by gain functions v and w are combined, then the distributed gains should correspond to the gains derived from v and the gains derived from w: $\phi_i(v + w) = \phi_i(v) + \phi_i(w)$ for every i in N. Also, for any real number a, $\phi_i(av) = a\phi_i(v)$; for every i in N.
4. **Zero player (null player).** A player is a zero-player if its marginal contribution in a coalition is zero (0). The Shapley value $\phi_i(v)$ of a null player i in a game v is zero.
$$v(S \cup \{i\}) = v(S) \tag{6}$$

4 ASSIGNING MEMBERSHIP VALUE BETWEEN THE PAIR OF WORDS IN FUZZY HINDI WORDNET

In this section, we explain the approach used for assigning membership value to fuzzy relations in the Fuzzy Hindi WordNet. Membership value signifies the degree of relationship between nodes in a graph. High membership value means two nodes are more related to each other and vice-versa. We used ConceptNet to quantify the relationship. This is performed in the following two steps:

(a) Create a subgraph of ConceptNet with the two words as end nodes.
(b) Quantify the relationship between words.

4.1 Create ConceptNet Subgraph

We traverse the ConceptNet using Depth First Search (DFS). The end nodes of the graph are those two Hindi words among which we have to find the connection. We start creating ConceptNet subgraph with one of the end nodes and proceed till a path length threshold *(L)* is reached. To restrict the information dilution, we restricted the path length threshold to $L = 8$. We incorporated nodes of Hindi as well as English language in the paths.

4.2 Quantify the Relationship Between Words

Once we have a graph, the next step is to quantify the relationship. The relationship value should be between 0 and 1 to qualify as a membership value for a fuzzy relationship. To quantify the relationship, we used a combination of two graph connectivity measures viz. *Vertex Connectivity* and *Edge Connectivity*.

Vertex Connectivity, $\lambda(G)$, is the minimum number of nodes to be deleted to obtain a disconnected graph G. For example, a disconnected graph has the vertex connectivity 0, and a connected graph with an articulation point has the vertex connectivity 1. Edge Connectivity $\kappa(G)$, is the minimum number of edges to be deleted to obtain a disconnected graph G. The edge connectivity of a disconnected graph is 0, while that of a connected graph with a graph bridge is 1. The following inequality holds between the vertex connectivity and edge connectivity.

$$0 \leq \text{ Vertex connectivity } \lambda(G) \leq \text{ Edge Connectivity } \kappa(G)$$

In the proposed work, our aim is not to merely disconnect the graph, instead to disconnect the graph such that two end nodes are in different components. To compute the membership value, we

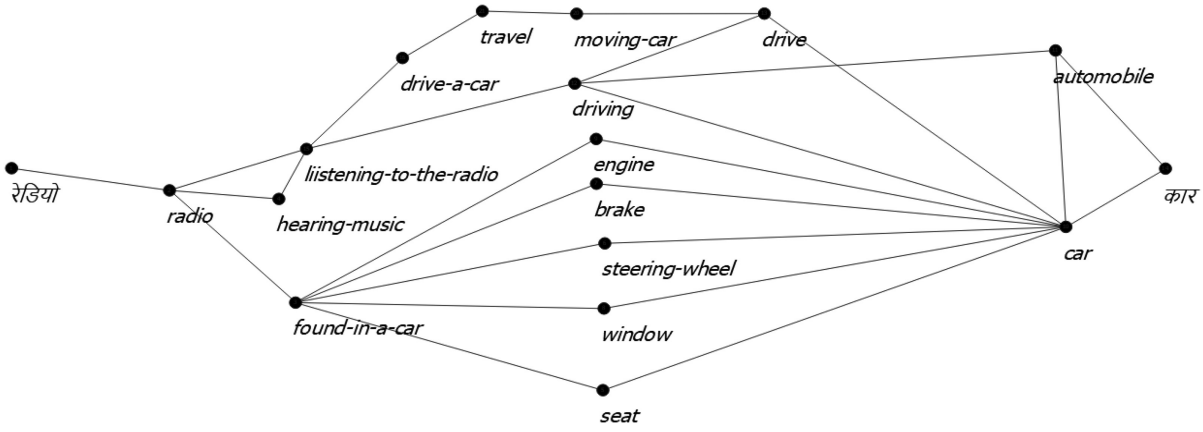


Fig. 3. Excerpt of ConceptNet graph showing the connection between रेडियो and कार.

create a combination of these metrics. We call it *Coupling Quotient* (τ). Coupling Quotient (τ) of a graph G, refers to the degree with which the two nodes are related to each other. It is determined by dividing vertex connectivity $\lambda(G)$ by edge connectivity $\kappa(G)$, as given in Equation (7).

$$\tau(G) = \frac{\lambda(G)}{\kappa(G)} \tag{7}$$

where $0 \leq \tau \leq 1$. For a disconnected graph, value for $\tau = 0$.

The value of coupling quotient (τ) is used as a membership value for the fuzzy relationship between the two words. A higher value of coupling quotient symbolizes that the information is concentrated in fewer vertices, and hence, there is lesser uncertainty. On the contrary, a lower value of τ indicates that information is diffused in the graph leading to higher uncertainty.

During initial experiments, we found that the coupling quotient was coming out to be unity (1) in numerous examples. To resolve this, we combine end nodes with their English synonyms in a single node. The graph connecting रेडियो (rēiḍayō) and कार (kāra) is illustrated in Figure 3. In Figure 3, it is visible that vertex connectivity of the graph is 1 and the vertex to be deleted is *radio*. Similarly, the edge connectivity of the graph in Figure 3 is 1 because of the bridge *radio* - रेडियो. The problem was resolved once we fused Hindi nodes and their English synonyms together as a single node. The updated graph for रेडियो (rēiḍayō) and कार (kāra) is shown in Figure 4. With the updated graph, vertex connectivity is 2 and edge connectivity is 3, thus the value of coupling quotient for the graph is $2/3 = 0.67$.

Table 1 shows some of the examples of the fuzzy membership values derived using the proposed algorithm. In the table, we have posted few examples of pair of nodes along with the fuzzy relation type that connects those words. Also, the fuzzy relationship value given by Jain and Lobiyal (2016) is given along with the fuzzy relationship value obtained using the proposed algorithm. From the table, it is evident that the fuzzy relationship values have increased or decreased among the pair of nodes.

5 WORD SENSE DISAMBIGUATION USING SHAPLEY VALUE

To perform WSD of Hindi text, we create a Fuzzy Hindi WordNet graph for an input sentence. The graph consists of word senses of all the ambiguous words in the sentence and the paths connecting these nodes using Fuzzy Hindi WordNet. The graph is established using Depth First Search (DFS). The computed coupling quotients are assigned as weights for fuzzy semantic relationships between the nodes of the graph.

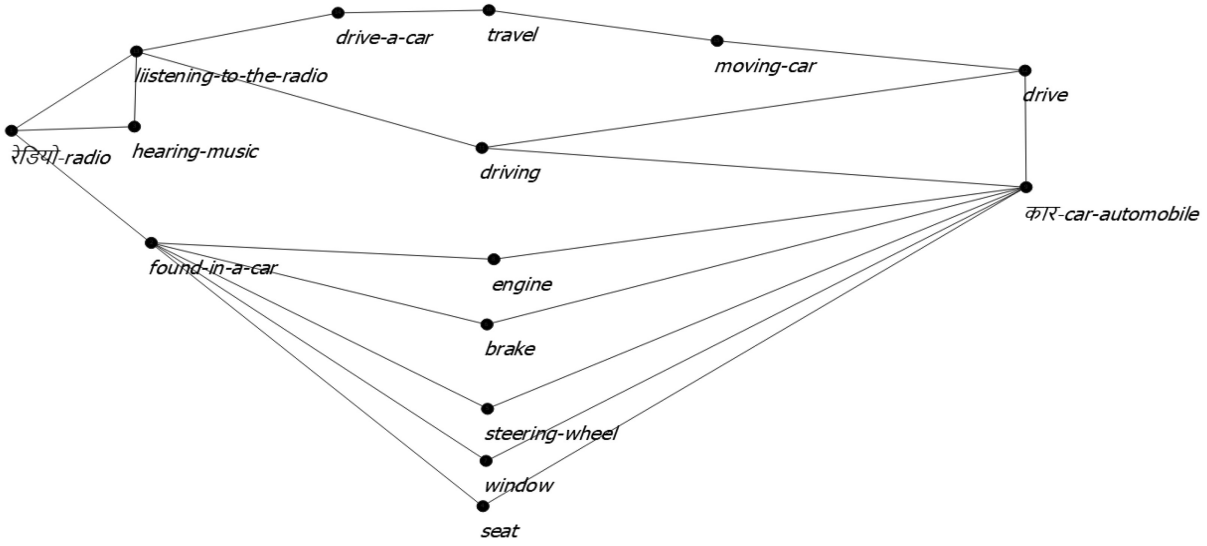


Fig. 4. Updated ConceptNet graph where Hindi end nodes are fused with their English synonyms into one node.

To compute Shapley value, we need to compute characteristic function *(v)* of coalitions. In the proposed work, we use *Fuzzy Group HITS* centrality measure. **Hyperlink Induced Topic Search (HITS)** was developed by Kleinberg [1999]. HITS was designed to rank web pages. It distinguishes web pages as Hubs and Authorities. A good hub represents a page that points to many other pages, and a good authority represents a page that is linked by many different hubs. In an undirected graph, the value of Hub and Authority is equal.

$$\text{Authority (v)} = \sum_{(u,v)\in E} Hub\,(u) \tag{8}$$

$$\text{Hub (v)} = \sum_{(u,v)\in E} Authority\,(u) \tag{9}$$

Jain and Lobiyal [2016] proposed Fuzzy HITS centrality measure for fuzzy graphs. The equations are given below.

$$\text{Authority}_{\text{f}}\,(\text{v}) = \sum_{(u,v)\in E} \mu_{uv} Hub\,(u) \tag{10}$$

$$\text{Hub}_{\text{f}}\,(\text{v}) = \sum_{(u,v)\in E} \mu_{uv} Authority\,(u) \tag{11}$$

where μ_{uv} is the strength of an edge between node u and node v.

In the proposed approach, we modified fuzzy HITS centrality measure to incorporate a group of words, called *Fuzzy Group HITS Centrality*. The modification was required to compute the value of the characteristic function for the coalitions. The modified formula for HITS is given in Equation (12). We chose HITS because this is one of the few centrality measures that assign a value of 0 to an isolated node. An isolated node does not affect any other nodes in the graph and is a null player.

$$\text{Authority}_{\text{fg}}\,(\text{S}) = \sum_{v\in S,\ (u,v)\in E} \mu_{uv} Hub\,(u) \tag{12}$$

$$\text{Hub}_{\text{fg}}\,(\text{S}) = \sum_{v\in S,\ (u,v)\in E} \mu_{uv} Authority\,(u) \tag{13}$$

$$v(S) = HITS_{fg}(S) \tag{14}$$

Table 1. Examples of Words Relationships and Associated Membership Values

S. No.	Pair of End nodes	Relation Type	Jain and Lobiyal (2016)	Proposed Method
1.	विद्यालय (vidyālaya, school) → संस्थान (saṁsthā, institute)	Fuzzy Association	0.8	0.71
2.	चूड़ी (cūḍī, thin bangle of glass) → कंगन (kangana, bangle)	Fuzzy associationism	0.7	1.00
3.	रेडियो (rēiḍayō, radio) → कार (kāra, car)	Fuzzy holonymy/meronymy	0.8	0.67
4.	फल (phala, fruit) → पेड़(pēḍa, tree)	Fuzzy holonymy/meronymy	0.7	0.80
5.	सोना (sōnā, sleep) → लेटना (lēṭanā, lie down)	Fuzzy Entailment	0.9	1.00
6.	चमकना (camakanā, to shine) → मेहनत करना (mēhanata karanā, to work hard)	Fuzzy Entailment	0.8	0.67
7.	दौड़ना (dauḍanā, to run) → चलना (calanā, to walk)	Fuzzy Troponymy	0.7	0.64
8.	पढ़ना (paḍhanā, to read) → सीखना (sīkhanā, to learn)	Fuzzy Troponymy	0.8	0.90
9.	मगरमच्छ (Magaramaccha, crocodile) → तैरना (tairanā, to swim)	Fuzzy Ability Link	0.7	0.87
10.	मछली (machalī, fish) → तैरना (tairanā, to swim)	Fuzzy Ability Link	1.0	1.0
11.	प्रोफेसर (pṛōphēsara, Professor) → पढ़ाना (paṛhānā, to teach)	Fuzzy Function Link	0.7	0.75

We compute the characteristics function for a coalition S as per Equation (14). We apply the Shapley value centrality measure on the graph of the input sentence to deduce the correct sense of ambiguous words. The original formula is given in Equation (3) in Section 3. We modified the formula to reduce the computation efforts. The modified formula is given in Equation (15). In this equation, we incorporated the fact that we cannot have a coalition containing different senses of a

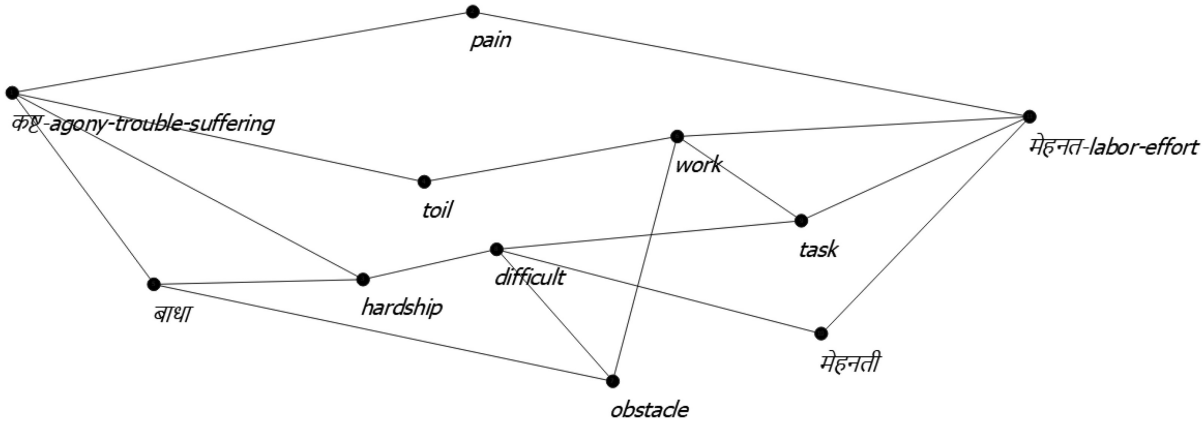


Fig. 5. Excerpt of a graph depicting the connection between कष्ट and *मेहनत* in ConceptNet.

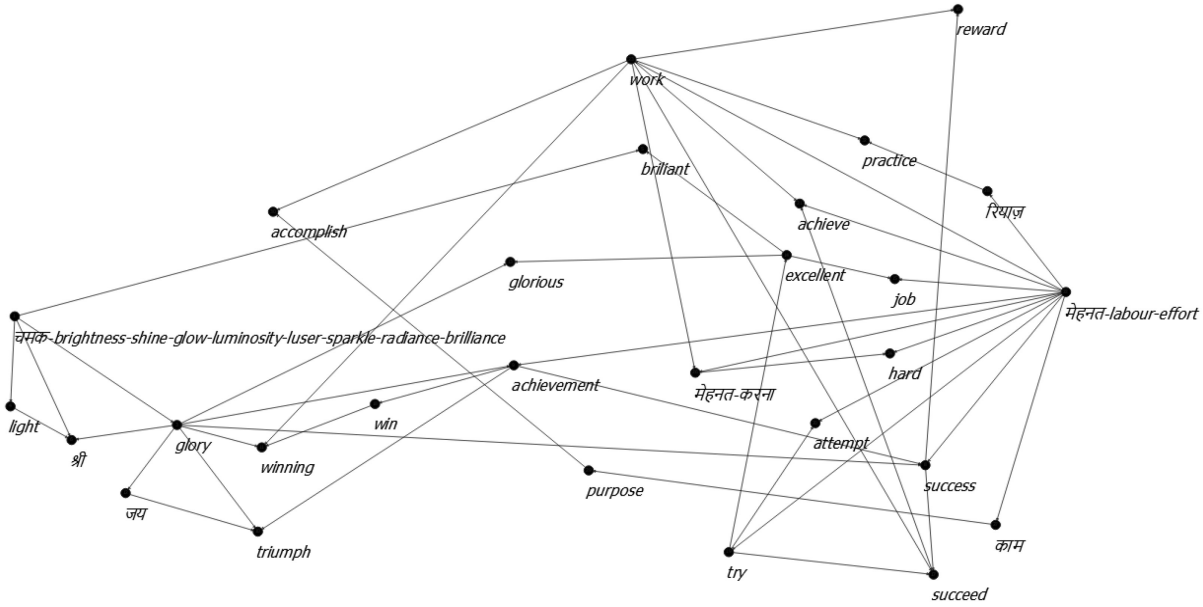


Fig. 6. Excerpt of a graph depicting the connection between **चमक** and **मेहनत** in ConceptNet.

word. The word sense with the highest Shapley Value is deduced as the correct word sense in the given context.

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} \left(v(S \cup \{i\}) - v(S)\right) (no.\ of\ senses\ of\ player\ i) \tag{15}$$

6 ILLUSTRATIVE EXAMPLES

To demonstrate our approach, we use the same example sentence as used by Jain and Lobiyal [2016]. Consider the Hindi sentence "कष्ट झेलकर ही चमकते है| (kaṣṭa jhēlakara hī camakte hai.). English translation of the sentence is "Glow only after suffering.". This sentence consists of three content words viz. "कष्ट *(kaṣṭa)*","झेलकर *(jhēlakara)*" and "चमकते *(camakte)*" The base words for these content words are कष्ट *(kaṣṭa)*, *झेलना (jhēlanā)* and चमक (*camak)* respectively.

Different paths connecting कष्ट *(kaṣṭa) and मेहनत* (mēhanata) are shown in Figure 5. The connectivity of चमक (*camak*) and *मेहनत (mēhanata)* via different paths is shown in Figure 6. For the sake of brevity, we show a part of the subgraph of ConceptNet.

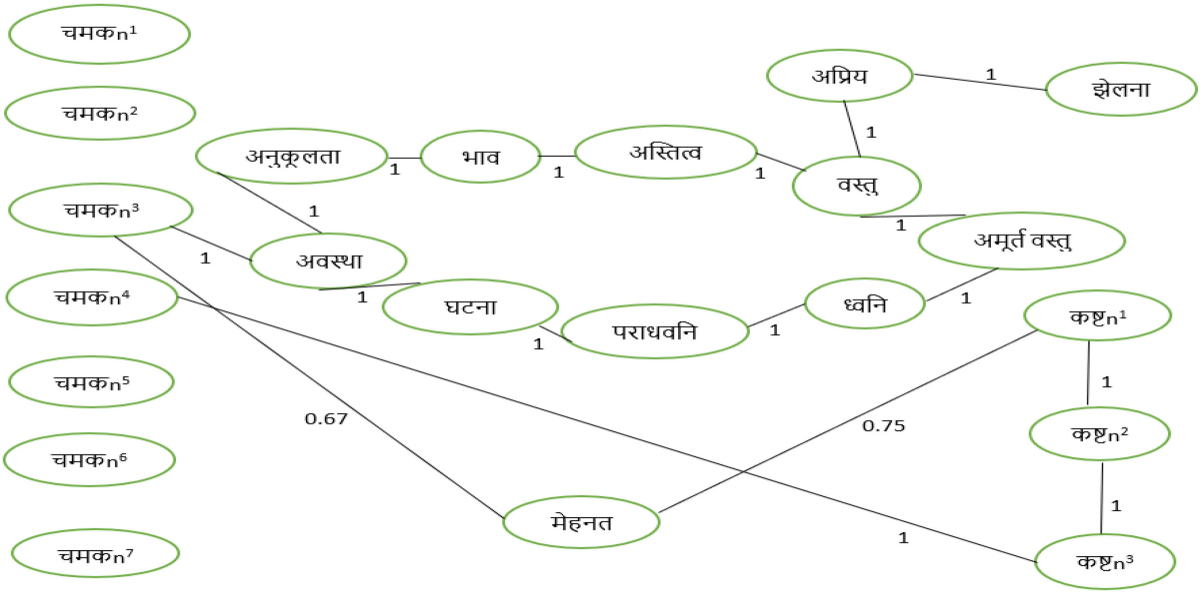


Fig. 7. FHWN graph with updated membership values for the example sentence.

In Figure 5, the vertex connectivity = 3 (pain, work and difficult), and edge connectivity = 4. Thus, coupling quotient = 3/4 = 0.75. Hence, the membership value of the fuzzy relationship between कष्ट and *मेहनत* is 0.75. In Figure 6, the vertex connectivity = 2 (glory and brilliant), and edge connectivity = 3. Thus, the coupling quotient = 2/3 = 0.67. Hence, the membership value of the fuzzy relationship between चमक and *मेहनत* is 0.67. We create an FHWN graph using the updated membership values. The graph is shown in Figure 7.

From Table 2, it is evident that different local centrality measures guide us towards different senses of the word. Also, the sense deduced for कष्ट is कष्ट $_n{}^3$ (incorrect) instead of कष्ट $_n{}^1$ (correct). However, the Shapley value helps us in identifying the correct sense of both the polysemous words कष्ट and चमक. Computations of characteristics function and consequent Shapley values are tabulated in Tables 3 and 4. From Table 3, we notice that the value of characteristic function (v) for isolated vertices in the graph is 0, e.g., चमक $_n{}^1$, चमक $_n{}^2$, and so on. Thus, they are declared as null or zero players. From Table 4, it is evident that चमक $_n{}^3$ and कष्ट $_n{}^1$ are the correctly deduced senses of the words in the given context.

Let us take another sentence "कुंजी का गत्ता फटा है" (kunjee ka gatta phata hai). (The English translation of the sentence is – "Cover of guide book is torn".) The open class words for this sentence are कुंजी (guide), गत्ता (cover) and फटा (torn). Hindi WordNet indicates फटना as the base form of the word फट, and thus we use फटना as the word to be disambiguated. गत्ता is monosemous, कुंजी has five senses, and फटना as noun is monosemous in the Hindi WordNet. The word कुंजी was not related with the word किताब (book) in the previous version of Hindi WordNet. Thus, the approach proposed by Jain and Lobiyal [2016] introduced a fuzzy association relationship between the two words with a strength of 0.7. However, these two words are related with one another with hypernymy relationship in the current version of Hindi WordNet. Thus, we take the strength of relationship between कुंजी and किताब as 1.0. गत्ता and किताब are not related in the current version of Hindi WordNet. The fuzzy membership value between the two words is computed to be 0.67 by the proposed algorithm. From Table 5, we notice that the two senses of word कुंजी viz. कुंजी $_n{}^2$, and कुंजी $_n{}^3$ have equal and highest fuzzy degree centrality value. Thus, the correct sense could not

Table 2. Values of Fuzzy Local Centrality Measures for the sentence "कष्ट झेलकर ही चमकते है। (kaṣṭa jhēlakara hī camakte hai.)"

Words	Degree	PageRank	HITS	Closeness
चमक$_n^1$	0.0	0.0057	0.0	0.0384
चमक$_n^2$	0.0	0.0057	0.0	0.0384
चमक$_n^3$	0.064	0.0302	**0.1902**	0.2674
चमक$_n^4$	**0.08**	**0.0360**	0.0062	**0.29**
चमक$_n^5$	0.0	0.0057	0.0	0.0384
चमक$_n^6$	0.0	0.0057	0.0	0.0384
चमक$_n^7$	0.0	0.0057	0.0	0.0384
झेलना	**0.04**	**0.0227**	**0.1113**	**0.1963**
कष्ट$_n^1$	0.072	0.0343	**0.0354**	0.2615
कष्ट$_n^2$	0.04	0.0205	0.0227	0.2381
कष्ट$_n^3$	**0.16**	**0.0694**	0.0134	**0.3424**

be deduced for the word कुंजी with the updated membership values. However, the Shapley Value deduced the correct sense.

Let us take one more example with more than one word that is ambiguous – "आग से घर जल गया" (*āg se ghar jal gaya*). (The English translation of the sentence is – "House got burnt with fire.".) The open class words for this sentence are आग (fire), घर (house) and जल (burn). Hindi Wordnet indicates जल, and जलना as two base forms of the word जल. We use जलना as the word to be disambiguated. आग has 3 senses as a noun, घर has 10 senses as a noun, and जलना has 8 senses as a verb in the Hindi WordNet. We generated a fuzzy Hindi WordNet graph for the sentence and computed Shapley Value for all the nodes. Using the proposed algorithm, we found correct senses for all the three ambiguous words of the sentence. The correct sense for आग is आग$_n^1$, घर is घर$_n^1$ and जलना is जलना$_v^3$.

7 EXPERIMENTAL SETUP AND RESULTS

We evaluated our proposed algorithm using three datasets. The first two datasets are the publicly available sense-marked corpus at IIT Bombay website.[3] Both the datasets are domain bound– one

[3] http://www.cfilt.iitb.ac.in/wsd/annotated_corpus/.

is of the health domain dataset and the other is the tourism domain dataset. Since, we were keen to evaluate the efficiency of the proposed algorithm on the general domain, we created a dataset using Hindi newspaper articles.

The evaluation metrics used for the experiments are given in 7.1. The datasets used to evaluate the performance of the proposed algorithm are described in 7.2. There was a need to extend the FHWN before performing experiments on the datasets, the process we followed is given in 7.3. We discuss the details and the performance of the proposed algorithm with different datasets in 7.4. The stability of Shapley Value centrality measure by varying the membership values was proved by undertaking various experiments as given in 7.5.

Table 3. Values of Characteristic function (v) for Various Coalitions for the Example Sentence

Coalition	Characteristic Function(v)	Coalition	Characteristic function (v)
{चमक$_n^1$}	0.0	{चमक$_n^1$, कष्ट$_n^1$}	0.2030
{चमक$_n^2$}	0.0	{चमक$_n^2$, कष्ट$_n^1$}	0.2030
{चमक$_n^3$}	0.1902	{चमक$_n^3$, कष्ट$_n^1$}	0.2256
{चमक$_n^4$}	0.0062	{चमक$_n^4$, कष्ट$_n^1$}	0.0416
{चमक$_n^5$}	0.0	{चमक$_n^5$, कष्ट$_n^1$}	0.2030
{चमक$_n^6$}	0.0	{चमक$_n^6$, कष्ट$_n^1$}	0.2030
{चमक$_n^7$}	0.0	{चमक$_n^7$, कष्ट$_n^1$}	0.2030
{झेलना}	0.1113	{चमक$_n^1$, कष्ट$_n^2$}	0.1622
{कष्ट$_n^1$}	0.0354	{चमक$_n^2$, कष्ट$_n^2$}	0.1622
{कष्ट$_n^2$}	0.0227	{चमक$_n^3$, कष्ट$_n^2$}	0.2129
{कष्ट$_n^3$}	0.0134	{चमक$_n^4$, कष्ट$_n^2$}	0.0289
{चमक$_n^1$, कष्ट$_n^3$}	0.1983	{चमक$_n^5$, कष्ट$_n^2$}	0.1622
{चमक$_n^2$, कष्ट$_n^3$}	0.1983	{चमक$_n^6$, कष्ट$_n^2$}	0.1622
{चमक$_n^3$, कष्ट$_n^3$}	0.2036	{चमक$_n^7$, कष्ट$_n^2$}	0.1622
{चमक$_n^4$, कष्ट$_n^3$}	0.2045	{चमक$_n^1$, झेलना}	0.0354
{चमक$_n^5$, कष्ट$_n^3$}	0.1983	{चमक$_n^2$, झेलना}	0.0354
{चमक$_n^6$, कष्ट$_n^3$}	0.1983	{चमक$_n^3$, झेलना}	0.3015
{चमक$_n^7$, कष्ट$_n^3$}	0.1983	{चमक$_n^4$, झेलना}	0.1175

(Continued)

Table 3. Continued

Coalition	Characteristic Function(v)	Coalition	Characteristic function (v)
{झेलना, कष्ट$_n^1$}	0.134	{चमक$_n^5$, झेलना}	0.0354
{झेलना, कष्ट$_n^2$}	0.1467	{चमक$_n^6$, झेलना}	0.0354
{झेलना, कष्ट$_n^3$}	0.1247	{चमक$_n^7$, झेलना}	0.0354
{चमक$_n^1$, कष्ट$_n^1$, झेलना}	0.134	{चमक$_n^1$, कष्ट$_n^3$, झेलना}	0.1247
{चमक$_n^2$, कष्ट$_n^1$, झेलना}	0.134	{चमक$_n^2$, कष्ट$_n^3$, झेलना}	0.1247
{चमक$_n^3$, कष्ट$_n^1$, झेलना}	0.3369	{चमक$_n^3$, कष्ट$_n^3$, झेलना}	0.3149
{चमक$_n^4$, कष्ट$_n^1$, झेलना}	0.1529	{चमक$_n^4$, कष्ट$_n^3$, झेलना}	0.1309
{चमक$_n^5$, कष्ट$_n^1$, झेलना}	0.134	{चमक$_n^5$, कष्ट$_n^3$, झेलना}	0.1247
{चमक$_n^6$, कष्ट$_n^1$, झेलना}	0.134	{चमक$_n^6$, कष्ट$_n^3$, झेलना}	0.1247
{चमक$_n^7$, कष्ट$_n^1$, झेलना}	0.134	{चमक$_n^7$, कष्ट$_n^3$, झेलना}	0.1247
{चमक$_n^1$, कष्ट$_n^2$, झेलना}	0.1467	{चमक$_n^5$, कष्ट$_n^2$, झेलना}	0.1467
{चमक$_n^2$, कष्ट$_n^2$, झेलना}	0.1467	{चमक$_n^6$, कष्ट$_n^2$, झेलना}	0.1467
{चमक$_n^3$, कष्ट$_n^2$, झेलना}	0.3242	{चमक$_n^7$, कष्ट$_n^2$, झेलना}	0.1467
{चमक$_n^4$, कष्ट$_n^2$, झेलना}	0.1402		

7.1 Evaluation Metrics

The performance of the proposed algorithm is evaluated using a standard parameter of F1-measure. F1-measure is a measure that conveys balance between the precision and the recall. It is determined by computing the weighted harmonic mean of precision and recall. Precision is the number of correct answers out of the number of answers provided by the algorithm. Recall computes the ratio of the number of correct answers and the total number of answers to be provided. F1-measure is computed according to Equations (16), (17), and (18). Since our algorithm provides an answer to all ambiguous words in the dataset, the values for precision, recall, and F1-measure are the same.

$$\text{F1} - \text{measure} = \frac{2*\text{Precision} * \text{Recall}}{(\text{Precision} + \text{Recall})} \tag{16}$$

$$\text{Precision} = \frac{\text{No. of correct answers}}{\text{No. of answers provided}} \tag{17}$$

$$\text{Recall} = \frac{\text{No. of correct answers}}{\text{No. of answers to be provided}} \tag{18}$$

Table 4. Shapley Value for Each Sense of Every Ambiguous Word for the Example Sentence

Word	Shapley Value	Word	Shapley Value
चमक$_n^1$	0.0	चमक$_n^7$	0.0
चमक$_n^2$	0.0	**झेलना**	0.0951
चमक$_n^3$	0.2028	**कष्ट$_n^1$**	0.0363
चमक$_n^4$	0.0059	कष्ट$_n^2$	0.0249
चमक$_n^5$	0.0	कष्ट$_n^3$	0.0137
चमक$_n^6$	0.0		

Table 5. Fuzzy Degree Centrality and Shapley Value for Each Sense of the Word **कुंजी** in the Example Sentence

Word Sense →	कुंजी$_n^1$	कुंजी$_n^2$	कुंजी$_n^3$	कुंजी$_n^4$	कुंजी$_n^5$
Fuzzy Degree Centrality	0.056	0.077	0.077	0.069	0.0
Shapley Value	0.0182	0.044	0.0529	0.0413	0.0

7.2 Description of Datasets

We experimented with the proposed algorithm using two standard-sense annotated datasets – Health and Tourism datasets, prepared by Khapra et al. [2010] and one developed dataset of general domain. The first dataset is based on the Health domain and the second dataset is based on the Tourism domain.

The data for Health domain was obtained from two doctors and the data for Tourism domain was collected from Indian Tourism websites written in English. The data was manually sense-annotated by two lexicographers who were adept in English. WordNet 2.1 was used as the sense inventory. A total of 42,046 words from the Health domain and a total of 134,095 words from the Tourism domain were manually sense marked. Some files were sense marked by both the lexicographers and the *Inter Tagger Agreement (ITA)* calculated from these files was 83% which is comparable to the 78% ITA reported on the SemCor corpus considering the domain-specific nature of the corpus. The datasets were translated into Hindi to create a standard publicly available Hindi sense-marked corpora by Khapra et al. [2010]. The datasets have evolved from the time of inception. In the currently available Health dataset, the corpus consists of 147 text files, each containing approximately 100 sentences. A total of 107,775 words are sense marked. The number of words in Health corpus in different POS categories is given in Table 6.

In the current version of the Tourism dataset, a total of 141,337 words are manually sense marked. The statistics of the Tourism dataset in terms of POS categories is given in Table 7. The newspaper dataset consists of 327 sentences with a total of 962 open class words to be sense marked. The number of words with different POS tags is given in Table 8. The words were sense marked by three Hindi lexicographers with an inter-annotator agreement of 79.2%.

Table 6. POS Category-Wise Description of Health Dataset

POS Category	Number of sense marked words
Noun	52,230
Verb	24,291
Adjective	22,699
Adverb	8,555

Table 7. **POS** Category-Wise Description of Tourism Dataset

POS Category	Number of sense marked words
Noun	72,932
Verb	26,086
Adjective	32,499
Adverb	9,820

Table 8. POS Category-Wise Description of Newspaper Dataset

POS category	Number of sense marked words
Noun	478
Verb	195
Adjective	216
Adverb	73

Fuzzy relationships between words of Health dataset were collected from the authors [Jain and Lobiyal 2016]. However, there was a need to add the fuzzy relationships between the words of sentences for the other two datasets—Tourism and Newspaper. Thus, our next task was to explore the fuzzy relationships between the words of the sentences of the two datasets and extend the FHWN.

7.3 Extension of Fuzzy Hindi WordNet (FHWN)

The fuzzy relationships between words of Tourism and Newspaper datasets were not included in the FHWN. Thus, we undertook this task of adding relationships. Rather than relying on the subjective opinion of handful experts, we used ConceptNet to assign the fuzzy relationships. We included fuzzy relationships with all those Hindi words that were at a maximum distance of 5 or below with one another in the ConceptNet. This restriction was required to check the growth of graph and the information dilution. For the Tourism and Newspaper datasets, a total of 1,059 new fuzzy relationships between the words were developed. The statistics about the number of fuzzy relationships of each type is given in Table 9.

Once the FHWN was extended with additional fuzzy relationships, next task was to assign the fuzzy membership values to these relationships. The algorithm discussed in Section 4 was used to assign the membership values. The results obtained after WSD of Hindi text on the three datasets is given in 7.4.

7.4 Results and Discussion

The performance of the proposed algorithm was compared with performances of other algorithms on the three datasets. The results for the Health, Tourism, and Newspaper datasets are tabulated in Table 10. From Table 10, it is evident that the performance of the proposed algorithm surpasses

Table 9. Number of fuzzy Relationships Created using ConceptNet for the two Datasets

S. No.	Fuzzy Relationship	Number of Fuzzy Relations
1	Fuzzy Association Relation	367
2	Fuzzy Hypernymy and Fuzzy Hyponymy	333
3	Fuzzy Meronymy and Fuzzy Holonymy	108
4	Fuzzy Entailment	32
5	Fuzzy Troponymy	29
6	Fuzzy Causative	14
7	Fuzzy Ability Link	47
8	Fuzzy Capability Link	35
9	Fuzzy Function Link	19
10	Fuzzy Attribute	79

Table 10. Comparison of the Proposed Algorithm with Existing Approaches

Dataset →	Health		Tourism		Newspaper	
Algorithm	All	Polysemous	All	Polysemous	All	Polysemous
Jain and Lobiyal [2016]	0.6183	0.4931	0.5629	0.4126	0.5307	0.3961
Kumari and Lobiyal [2019]	0.6217	0.5138	0.5591	0.3995	0.5411	0.4022
Tripathi et al. [2020]	0.5811	0.4307	0.4963	0.3880	0.3971	0.3168
Proposed approach with Hindi WordNet	0.5417	0.4300	0.5192	0.3989	0.4905	0.3726
Proposed approach with Fuzzy Hindi WordNet	0.6601	0.5328	0.5880	0.4593	0.5448	0.4274

Performance of the Proposed Approach using Hindi WordNet and Fuzzy Hindi WordNet.

the performance of the state-of-the-art algorithms. Table 10 also contains the results obtained using Shapley Value centrality measures with Hindi WordNet sentence graph instead of Fuzzy Hindi WordNet. The performance of Shapley Value on the Hindi WordNet graph was worse in comparison to the performance with Fuzzy Hindi WordNet as shown in Table 10.

We were able to achieve better results than other state-of-the-art approaches because of the following two reasons:

(a) **Enhanced Knowledge Base.** We used FHWN as the knowledge base for WSD of Hindi text. FHWN is an extension of Hindi WordNet with added relationships between words that are related with one another with some degree. The degree of relationship lies between 0 to 1, both inclusive. With this extended knowledge base, we had better insight about the word relationships.
(b) **Centrality Measure.** We experimented with various graph centrality measures that have been used in the literature for WSD. However, we noticed that the results vary with the change in fuzzy relationship strength. Thus, we used game-theory-based Shapley Value centrality measure. There is minimum impact on the Shapley Value with the change in fuzzy relationship strength.

Despite good results, there is a high probability for improvement for the proposed algorithm. We could identify a few factors that restricted the performance at different steps of the algorithm. The first task for WSD of Hindi text using the proposed algorithm is to establish new fuzzy relationships between words of the sentence under consideration. This itself is a challenging task. We performed this task using ConceptNet to bring out new fuzzy relationships between the words efficiently. However, there is no method to categorize these relationships without human

Table 11. Comparison of the Performance of Different Centrality Measures when Membership Values Were Incremented by 0.1

Centrality Measure	F1-Measure
Degree	0.5926
PageRank	0.6135
HITS	0.5310
Closeness	0.5437
Shapley Value	0.6618

Table 12. Comparison of the Performance of Different Centrality Measures when Membership Values Were Decremented by 0.1

Centrality Measure	F1-Measure
Degree	0.6395
PageRank	0.5928
HITS	0.5359
Closeness	0.5441
Shapley	0.6573

intervention yet. Though it does not affect the performance of our proposed algorithm, in future, we will try to bring out a novel method to tag fuzzy relationships automatically. Other point of concern for the proposed algorithm is that the list of fuzzy relationships is not exhaustive. A more voluminous graph could help us get better results. The assignment of membership values for fuzzy relationships are dependent on the availability of Hindi concepts in ConceptNet. The absence of Hindi words, e.g., विकल्प (option), टेंशन (tension), आधुनिकता (modernity), and so on in ConceptNet also impacted the performance of the proposed algorithm.

7.5 Stability of Shapley Value

To confirm the stability of Shapley value against other fuzzy graph centrality measures, we performed few more experiments on the Health dataset. In the first experiment, we incremented the membership values of fuzzy relationships by *0.1*. For example, membership value between मेहनत and चमक was increased from 0.67 to 0.77 and between मेहनत and कष्ट from 0.75 to 0.85. If the old membership value is $\geq$0.9, then the value is incremented to 1.0. The results obtained from various fuzzy centrality measures and the Shapley value on the Health dataset with the updated membership values are given in Table 11.

We performed another experiment by decrementing the membership value by 0.1. If the value of membership value is $\leq$0.1, the membership value is reduced to 0.0, signifying that the words are no more related to each other. The results with the updated membership values are tabulated in Table 12.

From Tables 11 and 12, we notice that the performance of fuzzy graph centrality measures varies with change in membership values of fuzzy relationships. Thus, without valid membership values, these centrality measures may provide varying results. The centrality measure that was profoundly impacted is Degree centrality. Also, from Tables 11 and 12, we find that the performance of Shapley value improved marginally with updated membership values. Since the change is minimal, it may be neglected.

Table 13. Comparison of the Performance of Different Centrality Measures as a Characteristic Function (v) for Shapley Value

Centrality measure as characteristic function (v)	F1-measure
Degree	0.6493
HITS	0.6601

We also tried different centrality measures to be used as a characteristic function for Shapley Value. One of the prime conditions to choose a centrality value for characteristic function is that it should provide a value of 0.0 for an isolated vertex. As seen in Table 2, PageRank and Closeness centrality do not provide 0.0 value for an isolated vertex. Thus, these centrality measures could not be used. Hence, we were left with only two choices viz. Degree and HITS, to be used as a characteristic function. We eventually opted for HITS because, with HITS, our algorithm gave better performance on the given dataset. Results are tabulated in Table 13.

8 CONCLUSION AND FUTURE WORKS

Fuzzy Hindi WordNet connects words with one another with varying degrees (between 0 to 1) of relationships called fuzzy relations. Therefore, the relations are closer to the human use of language. However, the assignment of strength to the relationship between words is a complex and challenging task. The proposed algorithm successfully demonstrates that ConceptNet is a useful knowledge base to assign strengths to the relationships between words. We also highlighted the limitations of local and global fuzzy centrality measures that restrict their usage with fuzzy graphs. Those limitations were overcome by using the GameTheory-based Shapley value centrality measure. The findings of our research show that Shapley Value correctly deduced the correct senses of the words. Further, we observed that the impact of changing fuzzy values on the word sense disambiguation was reduced, if not wholly eliminated, to a great extent by the proposed algorithm. We tested the proposed algorithm's efficacy on three datasets, two standard domain-specific datasets, and one general dataset. We obtained good results on all of the three datasets.

BERT language model has proven to be an effective tool for various NLP tasks. We plan to explore its utility for WSD tasks for Indian languages. Further, the exploration of game theory for WSD will be interesting future work.

REFERENCES

E. Agirre, O. Lopez De Lacalle, A. Soroa, and I. Fakultatea. 2009. Knowledge-based WSD and specific domains: Performing better than generic supervised WSD. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence.* (Pasadena, California), 1501–1506.

L. V. Ahn, M. Keida, and M. Blum. 2006. Verbosity: A game for collecting common sense facts. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems.* 75–8.

S. Banerjee and T. Pedersen. 2002. An adapted Lesk algorithm for word sense disambiguation using Wordnet. In *Lecture Notes in Computer Science*, vol. 2276. Springer Verlag, Berlin, 136–145.

P. Basile, A. Caputo, and G. Semeraro. 2014. An enhanced Lesk word sense disambiguation algorithm through a distributional semantic model. In *Proceedings of COLING 2014, 25th International Conference on Computational Linguistics.* 1591–1600.

P. Bhattacharya. 2010. IndoWordNet. In *Proceedings of the Lexical Resources Engineering Conference* 2010. Malta.

S. Bhingardive and P. Bhattacharya. 2016. Word sense disambiguation using IndoWordNet. In *WordNet in Indian Languages.* N. Dash, P. Bhattacharyya, and J. Pawar J. (eds), 243–260.

M. Carpuat and D. Wu. 2005. Word sense disambiguation vs. statistical machine translation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL'05).* 387–394.

M. Carpuat and D. Wu. 2007. Improving statistical machine translation using word sense disambiguation. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP/CoNLL'07).* 61–72.

Y. S. Chan, H. T. Ng, and D. Chiang. 2007. Word sense disambiguation improves statistical machine translation. In *Proceedings of Annual Meeting-Association for Computational Linguistics (ACL'07)*.

D. S. Chaplot, P. Bhattacharyya, and A. Paranjape. 2015. Unsupervised word sense disambiguation using Markov random field and dependency parser. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*. (Austin, Texas). AAAI, 2217–2223.

J. Chen and J. Liu. 2011. Combining WordNet and ConceptNet for word sense disambiguation. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*. 686–94.

Z. Dong, Q. Dong, and C. Hao. 2010. HowNet and its computation of meaning. In *Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations (COLING'10)*. (Beijing, China, Aug. 2010). 53–56.

C. Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. Cambridge Press.

C. B. S. Gautam and D. K. Sharma. 2016. Hindi word sense disambiguation using Lesk approach on Bigram and Trigram words. In *Proceedings of International Conference on Advances in Information Communication Technology and Computing. (AICTC)*.

H. Gonen and Y. Goldberg. 2016. Semi-supervised preposition-sense disambiguation using multilingual data. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics*. (Osaka, Japan, Dec. 2016). 2718–2729.

C. Havasi, R. Speer, and J. Pustejovsky. 2010. Coarse word sense disambiguation using common sense. Common sense knowledge. In *Proceedings of the AAAI Fall Symposium: Commonsense Knowledge*.

M. Hoey. 2005. *Lexical Priming. A New Theory of Words and Language*. Routledge, London, U.K.

C. Hung and S.-J. Chen. 2016. Word sense disambiguation-based sentiment lexicon for sentiment analysis. *Knowledge Based Systems*. 110, 224–232.

A. Jain and D. K. Lobiyal. 2014. A new approach for unsupervised word sense disambiguation in Hindi language using graph connectivity measures. *International Journal of Artificial Intelligence and Soft Computing*, 4, 4 (2014), 318–334.

A. Jain and D. K. Lobiyal. 2015. Unsupervised Hindi word sense disambiguation based on network agglomeration. In *Proceedings of the 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*. (New Delhi, India).

A. Jain and D. K. Lobiyal. 2016. Fuzzy Hindi WordNet and word sense disambiguation using fuzzy graph centrality measures. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*. 15, 2 (2016), 8.1–8.36.

G. Jain and D. K. Lobiyal. 2020. Word sense disambiguation using implicit information. *Natural Language Engineering (NLE)*. 26, 4 (2020), 413–432.

M. Khapra, A. Kulkarni, S. Sohoney, and P. Bhattacharyya. 2010. All words domain adapted WSD: Finding a middle ground between supervision and unsupervision. In *Proceedings of the Conference of Association of Computational Linguistics (ACL'10)* (Uppsala, Sweden, July 2010).

J. Kleinberg. 1999. *Hubs, Authorities, and Communities*. Cornell University.

A. Kumari and D. K. Lobiyal. 2020. Word2vec's distributed word representation for Hindi word sense disambiguation. In *Proceedings of the International Conference on Distributed Computing and Internet Technology (ICDCIT'20)*. D. Hung and M. D'Souza (eds.) *Lecture Notes in Computer Science*. 11969. Springer, Cham, December 2019.

M. Lesk. 1986. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *SIGDOC'86: Proceedings of the 5th Annual International Conference on Systems Documentation*. 24–26.

S. L. Manion and R. Sainudiin. 2014. An iterative Sudoku style approach to subgraph-based word sense disambiguation. In *Proceedings of the 3rd Joint Conference on Lexical and Computational Semantics (*SEM'14)*. 40–50.

D. McCarthy, R. Koeling, J. Weeds, and J. Carroll. 2007. Unsupervised acquisition of predominant word senses. *Computational Linguistics*, 33, 4 (Dec 2007), 553–590.

R. Mihalcea. 2005. Unsupervised large-vocabulary word sense disambiguation with graph-based algorithms for sequence data labeling. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing* Oct 2005. (Vancouver, B. C. Canada). 411–418.

G. Miller. 1995. WordNet: A lexical database for English. *Communications of the ACM*, 38, 39–41.

A. Moro, A. Raganato, and R. Navigli. 2014. Entity linking meets word sense disambiguation: A unified approach. *Transactions of the Association for Computational Linguistics*, 2 (2014), 231–244.

D. Narayan, D. Chakrabarti, P. Pande, and P. Bhattacharyya. 2002. An experience in building the Indo WordNet - A WordNet for Hindi. In *Proceedings of the 1st International Conference on Global WordNet*. (Mysore, India).

R. Navigli and M. Lapata. 2007. Graph connectivity measures for unsupervised word sense disambiguation. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*. (Hyderabad, India).

R. Navigli and M. Lapata. 2010. An experimental study of graph connectivity for unsupervised word sense disambiguation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32, 4 (2010), 678–692.

G. Owen. 1995. *Game Theory* (3rd ed.). Academic Press, San Diego, California, ISBN 978-0-12-531151-9.

M. Pace-Sigge. 2013. The concept of lexical priming in the context of language use. *ICAME Journal No.* 37, 149–173.

S. Patwardhan, S. Banerjee, and T. Pedersen. 2003. Using measures of semantic relatedness for word sense disambiguation. In *Proceedings of CICLing'03*. 241–257.

L. N. Pina and R. Johansson. 2016. Embedding sense for efficient graph-based word sense disambiguation. In *Proceedings of 2016 Workshop on Graph-Based Methods for Natural Language Processing (NAACL-HLT)*. 1–5.

R. Rapp and M. Zock. 2014. The CogALex-IV shared task on the lexical access problem. In *Proceedings of the 4th Workshop on Cognitive Aspects of the Lexicon*. 1–14.

V. Rentoumi, G. Giannakopoulos, V. Karkaletsis, and G. A. Vouros. 2009. Sentiment analysis of figurative language using a word sense disambiguation approach. In *Proceedings of RANLP*. 370–375.

S. Rothe and H. Schütz. 2015. AutoExtend: Extending word embeddings to embeddings for synsets and lexemes. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* (Volume 1: Long Papers). Association for Computational Linguistics. Stroudsburg, PA. 1793–1803.

S. Rothe and H. Schütze. 2017. AutoExtend: Combining word embeddings with semantic resources. *Computational Linguistics*. 43, 3 (2017), 593–617.

T. Ruas, W. Grosky, and A. Aizawa. 2019. Multi-sense embeddings through a word sense disambiguation process. *Expert Systems with Applications*, 136 (2019), 288–303.

G. Sajini and J. S. Kallimani. 2017. Recognition and disambiguation of polysemy words in entered Hindi documents. In *Proceedings of the International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques*. 873–880.

R. Sanghal, S. Paul, and P. K. Mayee. 2013. PurposeNet: A knowledge base organized around purpose. In *Proceedings of the 20th International Conference on Conceptual Structures (ICCS'13)*, (Mumbai, India, Jan. 10–12). 2013, 29–30.

F. C. Schadd and N. Roos. 2015. Word-Sense disambiguation for ontology mapping: Concept disambiguation using virtual documents and information retrieval techniques. *Journal on Data Semantics*, 4 (2015), 167–186.

L. Shapley. 1953. A value for *n*-person games. In *Contributions to the Theory of Games. Annals of Mathematical Studies*. 28. Princeton University Press, Princeton, N.J., 307–317.

P. Singh, T. Lin, E. T. Mueller, G. Lim, T. Perkins, and W. L. Zhu. 2002. Open mind common sense: Knowledge acquisition from general public. In *Proceedings of On the Move to Meaningful Internet Systems*. 1223–1237.

S. Singh and T. J. Siddiqui. 2014. Role of semantic relations in Hindi word sense disambiguation. *Procedia Computer Science*, 46 (2014), 240–248.

R. Sinha and R. Mihalcea. 2007. Unsupervised graph-based word sense disambiguation using measures of word semantic similarity. In *Proceedings of the International Conference on Semantic Computing*. Washington D. C.

C. Sumanth and D. Inkpen. 2015. How much does word sense disambiguation help in sentiment analysis of micropost data? In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. 115–121.

K. Taghipour and H. T. Ng. 2015. Semi-supervised word sense disambiguation using word embeddings in general and specific domains. In *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL* (Colorado, May 31–June 5, 2015), 314–323.

D. K. Tayal, L. Ahuja, and S. Chhabra. 2015. Word sense disambiguation in Hindi language using hyperspace analogue to language and fuzzy C-Means clustering. In *Proceedings of 12th International Conference on Natural Language Processing*. D. S. Sharma, R. Sangal, and E. Sherly (eds.), 49–58.

P. Tripathi, P. Mukherjee, M. Hendre, M. Godse, and B. Chakraborty. 2020. Word sense disambiguation in Hindi language using score based on modified Lesk algorithm. *International Journal of Computing and Digital Systems*.

R. Tripodi and M. Pelillo. 2017. A game theoretic approach to word sense disambiguation. *Computational Linguistics*, 43 (2017), 31–70.

L. Vanderwende, G. Kacmarcik, H. Suzuki, and A. Menezes. 2005. MindNet: An automatically-created lexical resource. In *Proceedings of HLT/EMNLP Interactive Demonstrations,* Oct 2005. (Vancouver, B.C., Canada), 8–9.

D. Vickrey, L. Biewald, M. Teyssier, and D. Koller. 2005. Word-sense disambiguation for machine translation. *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*. 771–778.

J. Von Neumann and O. Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, N.J.

T. Wang, J. Rao, and Q. Hu. 2014. Supervised word sense disambiguation using semantic diffusion kernel. *Engineering Applications of Artificial Intelligence*, 27 (Jan. 2014), 167–174.

D. Yuan, J. Richardson, R. Doherty, C. Evans, and E. Altendorf. 2016. Semi-supervised word sense disambiguation using neural modes. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics*. (Osaka, Japan, Dec. 2016), 1374–1385.

Z. Zhong and H. T. Ng. 2010. It makes sense: A wide-coverage word sense disambiguation system for free text. In *Proceedings of the ACL 2010 System Demonstrations*. 78–83.

Received May 2019; revised March 2020; accepted September 2021