SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(96th Edition)

DECEMBER 2020

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

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

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

This issue contains the information published during **December, 2020**. The arrangement of the contents is alphabetical. The full text of the article which is either subscribed by the university or available in the web is provided in this bulletin.

Central Library

CONTENTS

1. A Deep Language-independent Network to analyze the impact of COVID-19 on the World via Sentiment Analysis, ***6.Ashima Yadav*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

2. A New Reconstruction of Numerical Fluxes for Conservation Laws using Fuzzy Operators, ***6.Ruchika Lochab*** and ***3.Vivek Kumar***, Mathematics, DTU

3. A Novel Trench FinFET as biosensor for early detection of dengue fever, ***8.Aditya M*** and ***3.Rishu Chaujar***, Physics, DTU

4. A robust digital ECG signal watermarking and compression using biorthogonal wavelet transform, Ashish Kumar, Ramana Ranganatham, Samayveer Singh, Rama Komaragiri and ***3.Manjeet Kumar***, Electronics, DTU

5. A Theoretical Comparison between CH3NH3PbI3 and CH3NH3SnI3 based Solar Cells, Gagandeep, ***3.Mukhtiyar Singh***, Ramesh Kumar, Vinamrita Singh and Sunita Srivastava, Physics, DTU

6. An experimental assessment of bio-mixed-phase change material in a lightweight compartment and its potential as latent energy storage in north-east climate of India, Shubham Kumar Jaiswal and ***6.1.Devesh Kumar***, Mechanical, DTU

7. Assessing brand love, brand sacredness and brand fidelity towards halal brands, ***6.1.Richa Joshi*** and Prerna Garg, Management, DTU

8. Cluster Based Deep Contextual Reinforcement Learning for top-k Recommendations, ***8.Anubha Kabra***, ***8.Anu Agarwal*** and ***3.Anil Singh Parihar***, CSE, DTU

9. CONCENTRATED SOLAR POWER FOR MID-DAY MEAL PROGRAM IN DELHI, ***8.Rajat Malik***, ***8.Raviratna Subir***, ***8.Rishabh Goswami*** and ***3.J.P. Kesari***, Mechanical, DTU

10. Crack Detection of Structures using Deep Learning Framework, ***8.Arunish Kumar***, ***8.Amit Kumar***, ***8.Avinash Kumar Jha*** and ***3.Ashutosh Trivedi***, Civil, DTU

11. Emerging Biosensors for Point of Care Cancer Detection, Varinder Garg, Harish Kumar, Surinder Rana, Bikramjeet Singh Kalsi, ***8.Sidharth Bhasin***, Mehak Sood, Ramandeep Kaur, ***8.Siddhant Rajkumar***, Biotechnology, DTU

12. Enzyme-Accelerated Acid Hydrolysis of Untanned Proteinaceous Wastes from Tanning Industry, Md. NUR-E-ALAM, Nasifa AKTER, Kanish FATEMA, Md. Abul Kashem AZAD, ***7.Shimul CHAKMA*** and Md. Anwar Arfien KHAN, DTU

13. Estimation and Experimental Validation of Mean-Field Homogenised Effective Properties of Composite, ***6.A. Sharma***, ***3.V. Rastogi*** and ***3.A.K. Agrawal***, Mechanical, Production, Industrial and Automobile, DTU

14. Experimental and computational analysis of mechanical properties of RGD840 material manufactured through PolyJet process, Nagendra Kumar Maurya, ***3.Vikas Rastogi*** and ***3.Pushpendra Singh***, Mechanical, DTU

15. Hoax news-inspector: a real-time prediction of fake news using content resemblance over web search results for authenticating the credibility of news articles, ***6.Deepika Varshney*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

16. Impact of IRDA Guidelines on Consumer Confidence in Life Insurance Market: Then and Now, ***3.Shikha Khera*** and Divya, DSM, DTU

17. KeyGames: A Game Theoretic Approach to Automatic Keyphrase Extraction, ***8.1.Arnav Saxena***, ***8.1.Mudit Mangal*** and ***3.Goonjan Jain***, Mathematics, DTU

18. Leave against Medical Advice in Children: Rural Indian Perspective, Debadyuti Datta, Tanushree Mondal, Mihir Sarkar, Soumita Chel, Sankar Kumar Das, Arnab Nandy, ***8.Divyoshanu M Ivan*** and Rakesh Mondal, Biotechnology, DTU

19. Multi-functional circuit using two Current Feedback Operational Amplifiers, ***8.Yashwant Singh***, ***8.Suyash Jain***, ***8.Shubham Snehill***, ***3.Ram Bhagat*** and ***3.DC Meena***, Electrical, DTU

20. Numerical Simulation of n-TiO2/p-CIGS Solar Cell, Gagandeep, ***3.Mukhtiyar Singh***, Ramesh Kumar and Sonia Kumari, Physics, DTU

21. Outsourcing and reverse supply chain performance: a triple bottom line approach, ***3.Saurabh Agrawal*** and Rajesh Kumar Singh, DSM, DTU

22. RECENT ADVANCEMENTS IN SOLAR FURNACE TECHNOLOGY, ***8.Sameer Kapur***, ***8.Satyam Shandilya***, ***8.Vishal Tiwari*** and ***3.J.P. Kesari***, Mechanical, DTU

23. Recent developments in meta -benziporphodimethene: A new porphyrin analogue, ***6.Deepali Ahluwalia***, ***3.Anil Kumar*** and ***6.SudhirGopalrao Warkar***, Applied Chemistry, DTU

24. Rehoboam at the NTCIR-15 SHINRA2020-ML Task, Tushar Abhishek, ***8.Ayush Agarwal***, Anubhav Sharma, Vasudeva Varma and Manish Gupta, DTU

25. Remote Mid IR Photoacoustic Spectroscopy for the Detection of Explosive Materials, ***6.Mohit Mann***, ***3.A.S. Rao*** and Ramesh C Sharma, Physics, DTU

26. Research on Understanding the Effect of Deep Learning on User Preferences, ***6.Garima Gupta*** and ***3.Rahul Katarya***, CSE, DTU

27. SketchFormer: transformer-based approach for sketch recognition using vector images, **3.*Anil Singh Parihar***, ***8.Gaurav Jain***, ***8.Shivang Chopra*** and ***8.Suransh Chopra***, CSE, DTU

28. Small-Signal Modeling of Mutually Coupled Switched Reluctance Motor, Siddharth Mehta, ***8.1.Prerit Pramod***, Iqbal Husain and Md Ashfanoor Kabir, Electrical, DTU

29. Soil Nailing for Highway Construction, Archita Goyal and ***3.A. K. Shrivastava***, Civil, DTU

30. Speed and Current Controller Modelling for Close Loop Control of Separately Excited DC Motor for Constant Flux and Flux Weakening Operation, ***8.Varij Kumar*** and ***3.Saurabh Mishra***, Electrical, DTU

31. Stochastic Filtering in Electromagnetics, ***6.Rahul Bansal***, ***3.Sudipta Majumdar*** and Harish Parthasarthy, Electronics, DTU

32. Study of Estimating Dynamic State Jacobian Matrix and Dynamic System State Matrix Based on PMU, ***7.Annu Ahlawat*** and ***3.S.T Nagarajan***, Electrical, DTU

33. Study of Trapping and Intermixing of Delhi Drains for Rejuvenation of the River Yamuna, ***7.Prateek Shukla***, ***3.SK Singh***, ***3.Anunay Gour***, Environmental, DTU

34. Synthetic data augmentation for surface defect detection and classification using deep learning, Saksham Jain, Gautam Seth, Arpit Paruthi, Umang Soni and ***3.Girish Kumar***, DTU

35. THE FAMILY OF SZÁSZ-DURRMEYER TYPE OPERATORS INVOLVING CHARLIER POLYNOMIALS, ***3.NAOKANT DEO*** AND ***6.RAM PRATAP***, MATHEMATICS, DTU

36. THE STATUS OF WOMEN PARTICIPATION IN ENGINEERING AND TECHNOLOGY PROFESSION AND THEIR CHALLENGES IN WORKING INSTITUTIONS, ***6.Addissie Melak*** and ***3.Seema Singh***, Humanities, DTU

37. Time varying dynamics of globalization effect in India, ***6.Shikha Gupta*** and ***3.Nand Kumar***, Humanities, DTU

38. Transfer Learning by Deep Tuning of Pre-trained Networks for Pulmonary Nodule Detection, ***8.Dhaarna Sethi***, ***8.Kriti Arora***, and ***3.Seba Susan***, IT, DTU

39. Twitter data in Emotional Analysis – A study, ***8.Lokesh Singh***, ***8.Prabhat Gupta***, ***3.Rahul Katarya*** and ***8.Pragyat Jayvant***, CSE, DTU

40. UNIVERSITIES FOR THE FUTURE JOBS AND HUMAN EXCELLENCE, ***1.1.P B SHARMA*** AND SANJNA VIJ, DTU

41. VDDDA Based Multifunction Voltage-Mode MISO Filter, ***7.Dhruv Goel*** and ***3.Deva Nand***, Electronics, DTU

42. Voltage and Frequency Control of Self Excited Induction Generator Integrated with PV System, ***8.Sombir*** and ***3.Madhusudan Singh***, Electrical, DTU

43. VolTAGE: Volatility Forecasting via Text Audio Fusion with Graph Convolution Networks for Earnings Calls, Ramit Sawhney, ***8.Piyush Khanna***, Arshiya Aggarwal, Taru Jain, Puneet Mathur and Rajiv Ratn Shah, CSE, DTU

1. *Vice Chancellor*

2. *Pro Vice Chancellor*

3. *Faculty*

4. *Teaching-cum-Research Fellow*

5. *Alumni*

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.*Asst. Librarian*

6.1. *Ex Research Scholar*

7.1. *Ex PG Scholar*

8.1. *Ex Undergraduate Student*

A Deep Language-independent Network to analyze the impact of COVID-19 on the World via Sentiment Analysis

Ashima Yadav, Dinesh Kumar Vishwakarma, *Senior Member, IEEE*

Abstract—Towards the end of 2019, Wuhan experienced an outbreak of novel coronavirus, which soon spread all over the world, resulting in a deadly pandemic that infected millions of people around the globe. The government and public health agencies followed many strategies to counter the fatal virus. However, the virus severely affected the social and economic lives of the people. In this paper, we extract and study the opinion of people from the top five worst affected countries by the virus, namely USA, Brazil, India, Russia, and South Africa. We propose a deep language-independent Multilevel Attention-based Conv-BiGRU network (MACBiG-Net), which includes embedding layer, word-level encoded attention, and sentence-level encoded attention mechanism to extract the positive, negative, and neutral sentiments. The embedding layer encodes the sentence sequence into a real-valued vector. The word-level and sentence-level encoding is performed by a 1D Conv-BiGRU based mechanism, followed by word-level and sentence-level attention, respectively. We further develop a COVID-19 Sentiment Dataset by crawling the tweets from Twitter. Extensive experiments on our proposed dataset demonstrate the effectiveness of the proposed MACBiG-Net. Also, attention-weights visualization and in-depth results analysis shows that the proposed network has effectively captured the sentiments of the people.

Index Terms—Attention, Machine learning, Natural Language Processing, Neural nets, Sentiment analysis.

1 Introduction

THE outbreak of COVID-19 pandemic[1] (popularly referred to as coronavirus), caused by severe acute respiratory syndrome coronavirus 2 (SARS- CoV- 2), has badly hit the entire world. According to the World Health Organization (WHO)[2], this pandemic has affected 7,941,791 people with a death toll of 43,47,96 worldwide. Thus, the WHO declared a Public Health Emergency of International Concern on January 30, 2020. Nearly every country of the world has reported active cases, following which the WHO has advised them to follow strict COVID-19 mitigation guidelines like social distancing, contact tracing, non-essential business closure, quarantine, ban on travel and mass gatherings, etc. [1] [2]. This creates a strong need to assess the attitude and behavior of the people related to the COVID-19 pandemic.

Sentiment analysis is a popular branch in Natural Language Processing which extracts the opinions or views of the people embedded in user-generated data like online reviews, comments, and multimodal data [3] [4]. The field has shown tremendous success in areas like Business review analysis [5] [6], Financial market prediction [7], Recommendation system [8], Multimodal analysis [9]. In the medical domain, the sentiments reflect a patient's health status, which can be good (positive), bad (negative), or normal (neutral) [10]. They can be used for analyzing the drug and medication's effectiveness or confirming the presence or absence of disease [11]. Recently, deep learning models have demonstrated significant improvements in many text processing tasks [12] [13]. They avoid the traditional handcrafted feature engineering by automatically learning the semantic features from the training data itself.

Hence, in this paper, we propose a Multilevel Attention-based Conv-BiGRU Network (MACBiG-Net) for classifying the opinions of the people from the countries which were worse-affected by the COVID-19 pandemic. Firstly, the words are encoded into a real-valued vector with the help of the Glove embedding matrix. Then word-level encoding is performed by 1D CNN, where different convolution kernels help in learning the local features of the text, followed by BiGRU to capture the long term dependencies of a word in both left and right direction. After this, we apply the attention network to extract the most informative word. Finally, we follow the same combination of CNN- BiGRU for encoding the sentence vector and again using attention mechanism to compute the essential weights of the sentences for the entire document classification. In this way, MACBiG-Net can efficiently classify the text into positive, negative, and neutral sentiments. The significant contribution of our work is outlined as follows:

- We propose a deep Multilevel Attention-based Conv-BiGRU Network (MACBiG-Net) for classifying the opinions of the people from the top five worst affected countries by the coronavirus, namely USA, Brazil, India, Russia, and South Africa.
- We also develop the COVID-19 Sentiment dataset by

Ashima Yadav and Dinesh Kumar Vishwakarma are with Biometric Research Laboratory, Department of Information Technology, Delhi Technological University, Delhi, India. E-mail: ashimayadavdtu@gmail.com, dvishwakarma@gmail.com.

[1]https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

[2]https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200815-covid-19-sitrep-208.pdf?sfvrsn=9dc4e959_2 (As on 15 August, 2020)

crawling and downloading the tweets from the famous microblogging site, Twitter, and labelling them according to the sentiments belonging to the positive, negative, and neutral sentiment classes. We download and analyze the tweets of the people belonging to the USA, Brazil, India, Russia, and South Africa, starting from January 1, 2020, to June 7, 2020.

- The proposed MACBiG-Net is language-independent as it can classify the tweets belonging to multiple languages like Hindi, Japanese, Arabic, Spanish, Urdu, etc. effectively by encoding the words and sentences, and assigning larger weights to relevant contexts for the document classification.
- We conduct rigorous experiments on the proposed COVID-19 Sentiment dataset. Experimental results demonstrate that the proposed network show significant improvement over the popular baseline methods for sentiment classification in terms of Precision, Recall, F1 score, Accuracy, Receiver operating characteristic (ROC), and Area under the curve (AUC) evaluation metrics. We also visualize the attention weights of the network for getting in-depth knowledge about the attention mechanism.
- Finally, we draw several crucial observations by visualizing the country-wise sentiments of the people from January to June. We also analyze the popular positive and negative sentiment topics, words, and hashtags that were generated during the pandemic to see how the pandemic affected the lives of people all over the world. This analysis can serve as the feedback to the government agencies regarding the mitigation plans taken by them. Further, it may also guide the future planning of the public health agencies in case of any such outbreak.

The remainder of the paper is organized as follows: Section 2 discusses the Related Work for sentiment analysis in the health-care domain. Section 3 focuses on the proposed methodology by describing the proposed MACBiG-Net. Section 4 presents the experimental results by discussing quantitative and qualitative analysis. Section 5 concludes our paper and also focuses on future work.

2 Related Work

The sentiment analysis is a vast field having ample applications, as discussed above. Since we could not find any previous work which had focused on sentiments expressed during the COVID-19 pandemic, hence we discuss prominent works related to sentiments in the health-care domain only. The facets of sentiments in medicine are broadly categorized into [10]: capturing patient's health status (good/bad/no change), diagnosing the certainty or severity of a disease, and the effectiveness of the treatment or drug.

The early work in this area focuses on in-built lexicons or machine learning-based approaches for extracting the sentiments. Yang et al. [14] proposed a framework based on the weighted scheme Latent Dirichlet Allocation (LDA) to analyze the health-related information regarding breast cancer. The medical topics were clustered together to perform sentiment analysis using AFINN lexicon. Each topic was associated with positive and negative sentiment terms. The framework can benefit other users to seek advice or take necessary precautions in case they experience similar symptoms. Sabra et al. [15] proposed a framework based on semantic extraction and sentiment assessment of the risk factors to predict the diagnosis of Venous thromboembolism disorder. The clinical narratives were parsed to extract the risk factors with the help of MetaMap. These risk factors were stored in a dictionary, and semantic-rich rules were derived from weighting the risk factors according to the severity. Finally, the polarity of the adjectives and adverbs were extracted with the help of SentiWordNet lexicon. The framework achieved 54.5% precision and 85.7% recall to identify patient's likelihood to develop VTE early and start the treatment as soon as possible.

Several clinical trials are necessary before releasing any drug to the market. Still, discovering all the side effects of the drugs is a challenging task. Moh et al. [16] developed an architecture to study the side effects of five crucial drugs by capturing the tweets from Twitter. The sentiment scores were extracted using several lexicons like SentiWordNet, AFINN, etc. Similarly, Jiménez-Zafra et al. [17] applied machine learning classifiers and lexicon-based approaches to compute the sentiment polarities for the physicians and the drugs. The reviews related to physicians contain informal language, while discussions about drugs contain the combination of colloquial language and specific terms like adverse effects or name of the drug. Korkontzelos et al. [11] identified the adverse drug reactions on social media platforms by integrating the sentiment feature. Experimental results show that sentiment features improved the F-score from 72.14% to 73.22%.

Recently, deep based approaches are being explored to extract the health-care sentiments expressed by the people. Limsopatham et al. [18] applied CNN and word embeddings on social media messages to medical concepts. Chen et al. [19] extracted the emotion features based on the LSTM network for screening the users on perinatal depression. The results show that the method improves the early screening process for perinatal depression. Grisstte et al. . [20] applied the Bi-LSTM model for identifying the adverse drug reactions from social media text. Talpada et al. [21] studied the impact of Telemedicine for heart attacks and epilepsy. They compared the lexical and semantic methods with the deep based methods. The lexical and semantic methods were based on VADER and TextBlob, whereas LSTM was employed for deep based classification. Experimental results signify that sentiments can be used to obtain the demographic distribution of Telemedicine for hearth attacks and epilepsy. Thus, sentiment in the context of medicine shows the impact of a disease or drug on the life of a person over a period of time. This can provide feedback to assess the effectiveness of the treatment or measures taken to combat a disease or a crucial event.

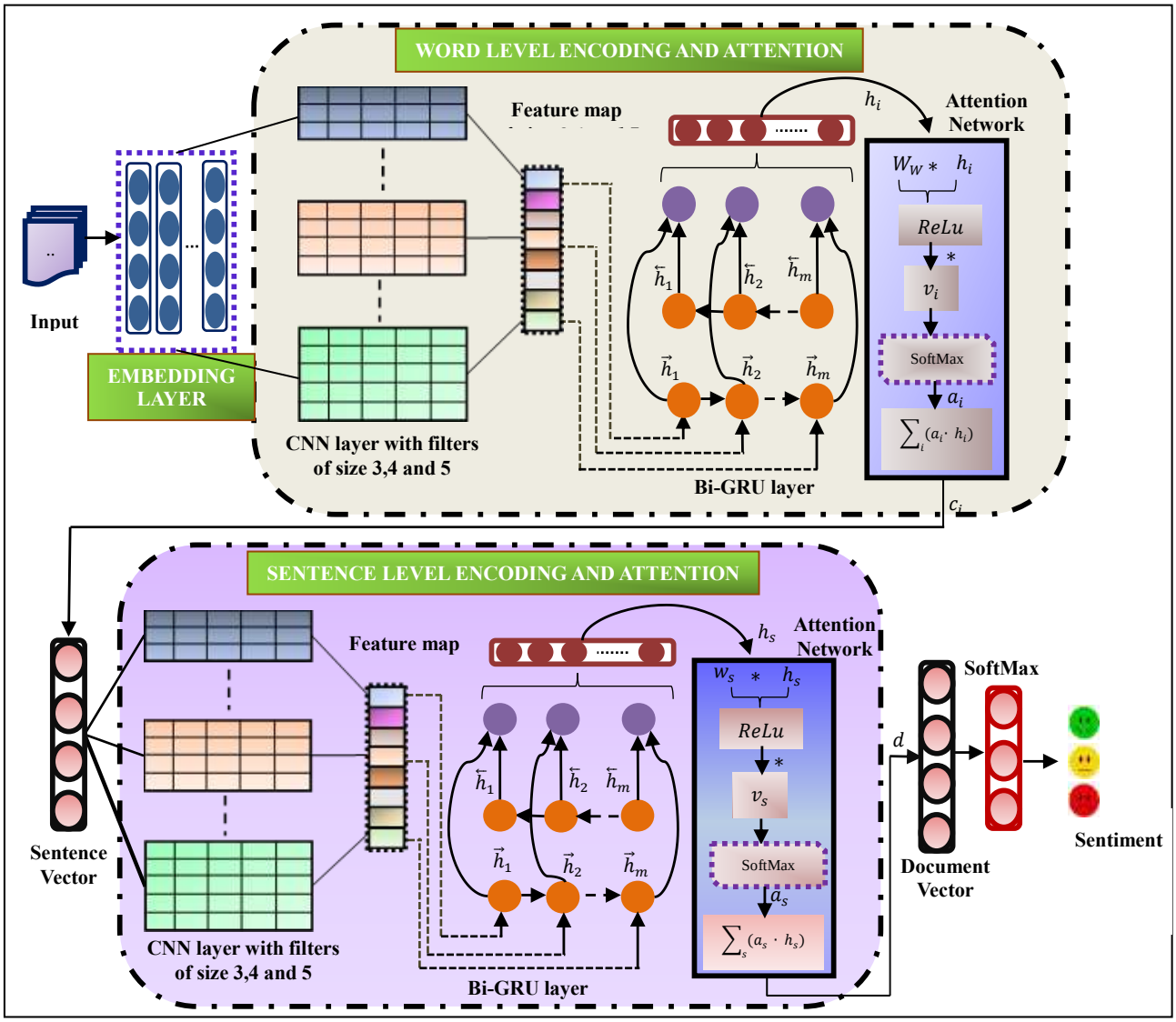


Fig. 1. Block diagram of the proposed Multilevel Attention-based Conv-BiGRU Network (MACBiG-Net)

3 Proposed Methodology

In this section, we explain the proposed framework for sentiment classification. It describes the data preprocessing process where we follow a two-step data cleaning approach, followed by the proposed Multilevel Attention-based Conv-BiGRU network (MACBiG-Net) to learn the semantic information of a sentence by encoding the essential words and sentences in a sequence

3.1 Data preprocessing

We scrapped the data from one of the most popular microblogging site Twitter for all the five countries based on their location. The complete data scraping, cleaning, and labeling process are discussed in Section 4. The final dataset comprises of people's views belonging to three output classes: positive, negative, and neutral. Since data processing is one of the most crucial tasks for handling the unstructured textual data, hence we meticulously designed a two-step data cleaning approach. The first approach is followed before translating the data into the English language so that the translation process is not affected by special characters or non-ASCII codes in the sentence. This includes removing hyperlink, special characters/symbols, retweet headers, twitter usernames, non-ASCII codes, NaN (Not a Number) character, and whitespaces. Finally, all the uppercase words are converted into lowercase by using the NumPy library functions.

In the second approach, the labeled dataset goes through stemming and lemmatization to reduce the word to its root form. Porter Stemmer and WordNet Lemmatizer are used for stemming and lemmatization, respectively. We also removed the stop words that frequently occur in the text and convey no meaningful information, followed by hashtags and punctuations removal. The cleaned data is then passed to the Multilevel Attention-based Conv-GRU network, as discussed in the subsequent section for the final classification.

3.2 Multilevel Attention-based Conv-BiGRU Network (MACBiG-Net)

The section describes the proposed MACBiG-Net, as shown in Fig. 1. The network includes three main steps: Embedding Layer, which encodes the input words into the low-dimensional vectors, Word-level Encoding and

Table I
Layer description of Step 2: Word-level Encoding and Attention

Layer Name	Input Shape	Output Shape	# Parameters	Connected to
Input Layer	(None, 200)	(None, 200)	0	-
Embedding	(None, 200)	(None, 200, 100)	1835200	Input Layer
Conv1D_1	(None, 200, 100)	(None, 198, 128)	38528	Embedding
Conv1D_2	(None, 200, 100)	(None, 197, 128)	51328	Embedding
Conv1D_3	(None, 200, 100)	(None, 196, 128)	64128	Embedding
MaxPooling1D_1	(None, 198, 128)	(None, 66, 128)	0	Conv1D_1
MaxPooling1D_2	(None, 197, 128)	(None, 65, 128)	0	Conv1D_2
MaxPooling1D_3	(None, 196, 128)	(None, 65, 128)	0	Conv1D_3
Concatenate	(None, 66, 128), (None, 65, 128), (None, 65, 128)	(None, 196, 128)	0	MaxPooling1D_1, MaxPooling1D_2, MaxPooling1D_3
MaxPooling1D_4	(None, 196, 128)	(None, 65, 128)	0	Concatenate
Bidirectional_GRU	(None, 65, 128)	(None, 65, 200)	183200	MaxPooling1D_4
TimeDistributed (Dense)	(None, 65, 200)	(None, 65, 100)	20100	Bidirectional_GRU
Hierarchical_attention_Network	(None, 65, 100)	(None, 100)	10200	TimeDistributed (Dense)

Attention, in which the words of a sentence are encoded by 1D Conv-BiGRU based representation to obtain the sentence representation, followed by word-level attention which computes the essential weights for the sentence vector, and Sentence level encoding and attention, where the sentences of a document are again encoded by 1D Conv-BiGRU based representation to obtain the document representation, followed by sentence-level attention to get the document vector for the classification.

3.2.1 Embedding Layer

Each word in the sequence of the sentence is converted into a real-valued vector. Formally, let $\{r_1, r_2, \ldots\ldots, r_n\}$ be the sequence of n words in a sentence. We first embed each word w_i into a real-valued vector through a word embedding matrix $E \in R^{d_r * |V|}$, where d_r is the word embedding size, and V is the vocabulary size. We use the Glove[3] embedding matrix to get the final sequence of vector $\{e_1, e_2, \ldots\ldots, e_n\}$, which serves as an input for the next layer.

3.2.2 Word-level Encoding and attention

Our proposed architecture is inspired by [22], which intends to capture the structure of words from the sentence and the structure of sentences from a document. The word encoding is performed by 1D convolution, followed by the Bi-GRU mechanism. We first describe the convolution process as follows:

The local features are extracted by using a 1D CNN network with different kernel sizes to generate a feature map. The different convolution kernels help in learning the various local characteristics of the text. In general, for e_i^{th} embedding corresponding to r_i^{th} word, the concatenation operation $\oplus$ is expressed as follows:

$$e_{1:n} = e_1 \oplus e_2 \oplus \ldots\ldots\ldots \oplus e_n \quad (1)$$

The convolution operation is applied on the window of j words to generate a new feature value m_i as shown below:

$$m_i = f\,(W \cdot e_{i:i+j-1} + b) \quad (2)$$

Where, b = bias and f = Rectified Linear unit activation function (ReLu). This filter is applied to different possible window of words in a sentence which generates the following feature map:

$$M = [m_1, m_2, \ldots\ldots\ldots, m_{n-j+1}] \quad (3)$$

We apply the filter of sizes 3, 4, and 5 to obtain multiple features followed by max-pooling layer to capture the maximum feature value. The concatenated output of all three filter sizes is passed into the BiGRU based layer, which contains the hidden states to store crucial information. The bidirectional nature of GRU is based on the fact that words in a sentence are not only related to the previous words but also to the following words. We have chosen GRU over LSTM due to its efficiency in training, which makes it computationally cheaper than LSTM. The forward GRU with $\vec{h}_i = \overrightarrow{GRU}\left(e_i, \vec{h}_{i-1}\right)$is concatenated with backward GRU, $\overleftarrow{h}_i = \overleftarrow{GRU}\,(e_i, \overleftarrow{h}_{i+1})$to obtain the output of BiGRU at step w as $h_i = \left[\vec{h}_i, \overleftarrow{h}_i\right]$. Thus, the BiGRU layer fetches the contextual information of every sentence centered around the word w_i. This makes sense as the orientation of sentiment needs to consider the past and future context information in a sentence. Thus, the word is encoded by the Conv-BiGRU block by storing the local and contextual features.

The attention mechanism is based on the idea that when users read a document word by word in each sentence, then they would pay attention to the most informative word or sentence. Hence, we model both the word level and sentence level attention differently. In word-level attention, the word annotation h_i is fed into the network to get its vector representation v_i. We apply the *ReLu* activation function over *tanh* as specified in [22], as *ReLu* converges quickly and results in cheaper computation. This is described in the (4):

$$v_i = ReLu\,(W_w \cdot h_i + b_w) \quad (4)$$

[3] https://nlp.stanford.edu/projects/glove/

Table II
Layer description of Step 3: Sentence-level Encoding and Attention

Layer Name	Input Shape	Output Shape	# Parameters	Connected to
Input Layer	(None, 15, 200)	(None, 15, 200)	0	-
TimeDistributed (Model)	(None, 15, 200)	(None, 15, 100)	2202684	Input Layer
Conv1D_1	(None, 15, 100)	(None, 13, 128)	38528	TimeDistributed (Model)
Conv1D_2	(None, 15, 100)	(None, 12, 128)	51328	TimeDistributed (Model)
Conv1D_3	(None, 15, 100)	(None, 11, 128)	64128	TimeDistributed (Model)
MaxPooling1D_1	(None, 13, 128)	(None, 4, 128)	0	Conv1D_1
MaxPooling1D_2	(None, 12, 128)	(None, 4, 128)	0	Conv1D_2
MaxPooling1D_3	(None, 11, 128)	(None, 3, 128)	0	Conv1D_3
Concatenate	(None, 4, 128), (None, 4, 128), (None, 3, 128)	(None, 11, 128)	0	MaxPooling1D_1, MaxPooling1D_2, MaxPooling1D_3
MaxPooling1D_4	(None, 11, 128)	(None, 3, 128)	0	Concatenate
Bidirectional_GRU	(None, 3, 128)	(None, 3, 200)	183200	MaxPooling1D_4
TimeDistributed (Dense)	(None, 3, 200)	(None, 3, 100)	20100	Bidirectional_GRU
Hierarchical_attention_Network	(None, 3, 100)	(None, 100)	10200	TimeDistributed (Dense)
Dropout	(None, 100)	(None, 100)	0	Hierarchical_attention_Network
Dense	(None, 100)	(None, 3)	303	Dropout

The resulting annotation v_i is multiplied (dot product) by the word context vector v_{wt} which is learned during the training. The softmax function is applied to obtain attention weights a_i as shown in (5) below:

$$a_i = \frac{\exp(v_{wt} \cdot v_i)}{\sum_i \exp(v_{wt} \cdot v_i)} \quad (5)$$

Finally, the attention weights a_i is concatenated with the word annotations to obtain the sentence vector c_i as described in (6) below:

$$c_i = \sum_i (a_i \cdot h_i) \quad (6)$$

The layer details of the word-level encoding and attention step are shown in Table I.

3.2.3 Sentence level encoding and attention

As discussed, selecting important sentences in the document becomes another crucial task for the classification. Hence, we follow the same procedure for generating the document representation of the sentences and, as discussed in Section 3.2.2, by encoding the sentences with relevant context and computing the crucial weights of these contexts for the document classification.

The sentence vector c_i obtained in Section 3.2.2 is passed into the 1D CNN layer and Bi-GRU layer to encode the sentence and obtain information h_s of the neighboring sentences centered around sentence s. The sentence context vector v_{st} is used to compute the sentence level attention a_s, which further yields the document vector d that summarizes all the information of the sentences in a document. Mathematically, this is represented as below in (7)-(8).

$$v_s = ReLu\,(W_s \cdot h_s + b_s) \quad (7)$$

$$a_s = \frac{\exp(v_{st} \cdot v_s)}{\sum_s \exp(v_{st} \cdot v_s)} \quad (8)$$

$$d = \sum_s (a_s \cdot h_s) \quad (9)$$

The layer details of the Sentence level encoding and attention step are shown in Table II. The learned vector d of [1* 100] dimension is passed to the dropout layer with a 0.5 rate to handle the overfitting problem in deep neural network architectures. This is followed by a dense layer and softmax activation function. The entire network is trained end to end for obtaining the best weights by monitoring the validation accuracy of the model. We use categorical cross-entropy loss, which is defined in (10) as follows:

$$Loss = -\sum_{j=1}^{3} (y_j \cdot \log \hat{y}_j) \quad (10)$$

Where, y_j is the target value corresponding to the model output $\hat{y}_j$ for the j^{th} sample. We have used the L2-regularization parameter, which penalizes the larger weights in the network for avoiding the high variance problem. This makes the objective function of the network as follows:

$$Cost\ function = Loss + (\lambda/2m * \sum ||w||^2) \quad (11)$$

Here, λ is the regularization parameter and one of the hyperparameters of the network, which is tuned according to the validation set. Our model achieves optimized results with λ = 0.001.

4 Experimental Analysis

This section discusses the data collection and labeling process in Section 4.1, implementation details in Section 4.2. We evaluate the effectiveness of the network in Section 4.3 and perform baseline comparison in Section 4.4. Finally, we visualize our classification results and extract the sentiment topics in Section 4.5.

4.1 Dataset Collection and Labeling

In order to analyze the impact of the novel coronavirus, we extracted the Tweets from January 1, 2020 to June 7, 2020 of top five most affected countries, namely, USA (52,03,206 cases), Brazil (32,24,876 cases), India (25,26,192 cases), Russia (9,17,884 cases), and South Africa (5,79,140 cases) as per

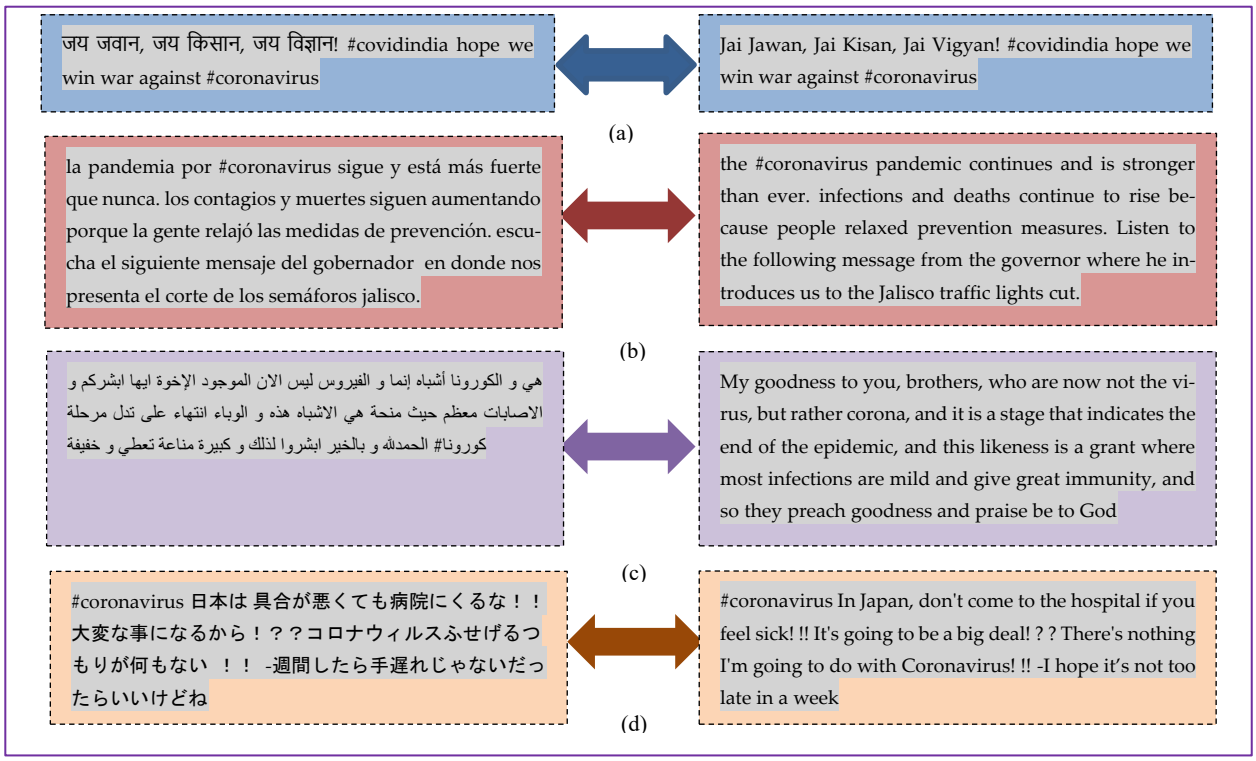


Fig. 2. Sample tweets in different languages from COVID-19 Sentiment Dataset (a) Hindi to English (b) Spanish to English (c) Arabic to English (d) Japanese to English.

data available from WHO[4]. We used some popular hashtags like #coronavirus, #covid19, and #COVID_19. The scrapped tweets were preprocessed as discussed in Section 3.1 and translated into English by the Google translator. Some sample tweets in a different language and their English translation is shown in Fig. 2.

Table III
Country-wise details of COVID -19 Sentiment Dataset

Country	Sentiment			Total
	Positive	Negative	Neutral	
USA	383	321	431	1135
Brazil	267	318	444	1029
India	302	164	299	765
Russia	147	168	249	564
South Africa	200	193	232	625
Total	**1299**	**1164**	**1655**	**4118**

We used the Textblob[5] library to weakly label the processed tweets with the sentiment score of [-1, 1]. The tweets with +1 sentiment scores were labeled as positive, tweets with -1 score were labeled as negative, and finally, tweets with 0 sentiment scores were labeled as neutral. However, we observe that in many cases, the Textblob library fails to classify the data into correct sentiment labels. E.g., Consider the following tweet, "*My dad tested positive for the corona*" was labeled as positive sentiment, though it belongs to the negative sentiment category. Hence, the shortlisted tweets were again manually annotated (strong labels) by three humans with positive (+1), negative (-1), and neutral (0) sentiment category. Finally, 4118 tweets were annotated with an average value of Cohen Kappa inter-annotator agreement as 0.85. The country-wise dataset details corresponding to each sentiment category are shown in Table III.

4.2 Implementation Details

The proposed MACBiG-Net is implemented on Python 3 using the popular Keras framework. The experiments were performed on Windows 10 machine with 128GB RAM using NVIDIA Titan RTX GPUs. The embedding dimension is 100, maximum length of all sequences is 200, and Punkt sentence Tokenizer is used for dividing the text into a list of sentences. Adam optimizer with learning rate = 0.0001, default parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$ is set for training. The entire network is trained end to end with 64 batches for 100 epochs. Dropout and Regularization are used for avoiding overfitting. We have used a ten-fold cross-validation testing strategy with 80% data for training, 5% data for validation, and 15% data for testing. The validation accuracy is monitored during training the network. The model achieving the highest validation accuracy is selected for the testing phase.

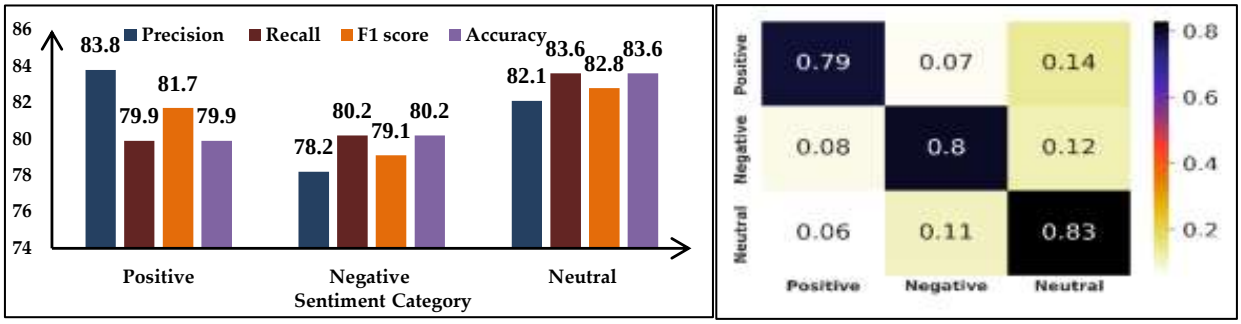
4.3 Classification Results and Observations

We compare the sentiment prediction performance of the MACBiG-Net in terms of precision, recall, F1 score, and accuracy corresponding to different sentiment categories, as shown in Fig. 3 (a). Since we used a ten-fold cross-validation strategy, the final results are reported by averaging the results across each of the test fold.

The accuracy of positive, negative, and neutral sentiment category is 79.9%, 80.2%, and 83.6%, respectively.

[4] https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200815-covid-19-sitrep-208.pdf?sfvrsn=9dc4e959_2 (As on 15 August, 2020)

[5] https://textblob.readthedocs.io/en/dev/



(a) (b)

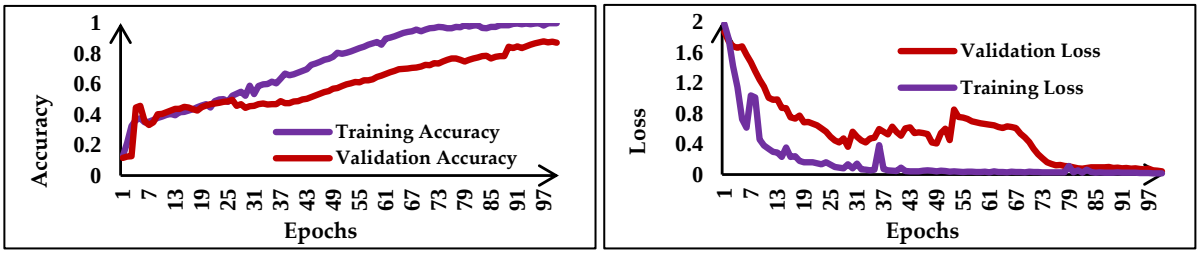
Fig. 3. (a) Sentiment classification results on COVID-19 Sentiment Dataset (%) (b) Confusion Matrix.

The average accuracy of the network is 81.5%. Further, the confusion matrix in Fig. 3 (b) gives an in-depth analysis of the classification results. It can be seen that a major amount of misclassification occurs when a positive sample is incorrectly classified as neutral. Yet, we can see that the network can discriminate well across each of the sentiment categories. These results clearly explain the effectiveness of MACBiG-Net.

We also provide accuracy and loss curves for training and validation set in Fig. 4 (a) and (b), respectively. As we can see, the training accuracy is 100%, which means the model is trained completely, and the validation accuracy shows how effectively the model can perform on the unseen samples. Similarly, decreasing loss curves with each epoch confirm the adequate learning of the model and validate its internal classification performance.

The foremost aim of our work is to analyze the sentiments of people expressed globally on the COVID-19 pandemic. This can be visualized in Fig. 5, where we plot the positive (+1), negative (-1), and neutral (0) sentiments of people over time located in the top five worst affected countries by the pandemic. As discussed, we retrieved the tweets from January 1, 2020 to June 7, 2020. However, in many countries, sentiments were generated in the month of late February or March. This may be because the first confirmed case in that country started in late February (Eg: *The virus was first confirmed in Brazil on February 25, 2020* [6]), or no specific sentiments were generated during January and February. A five-day window represents the time component in Fig. 5 along the x-axis. This means that each bar represents the sentiment expressed in these five days. However, each day, several sentiments were being generated. Hence, we took the sentiment that was voiced the maximum number of times by the people on any particular day as the final sentiment for that day. We summarize our observations as follows:

- In the USA, significant sentiments emerged from March onwards. This may be because Trump declared a national emergency[7] on March 13, and by the end of March, the cases were reported in all 50 US States[8]. In early March, positive sentiment was recorded. However, by the end of March till mid-April, negative sentiment was prevalent. Similarly, by the end of May, we observed mostly negative sentiments. One of the primary reasons was the breaking news of George Floyd, which created a state of panic and terror in the USA from May 25 onwards. People started tweeting that COVID-19 will continue to spread as citizens were protesting and moving in groups, without following any preventive measures. Overall, it was observed that 43.6% of sentiments were neutral, 37.9% were positive, and 18.3% were negative.
- In India, positive sentiments were observed majorly in March and April. This may be because people were hopeful in the initial stages to overcome the virus. Further, Prime Minister (PM) Narender Singh Modi came up with some ideas to boost the support of all the frontline workers. On March 22, the PM requested every Indian to clap their hands or ring some bells from their house to boost the morale of frontline workers. Similarly, on April 5, the PM urged everyone to light candles or *diyas* for nine minutes as a mark to fight against the deadly virus. Several people followed these tasks and happily enjoyed them. Towards the end of May, we see the people expressed negative sentiments. Since, by the end of May, a large number of active cases were



(a) (b)

Fig. 4. Tracing internal performance of MACBiG-Net with (a) Accuracy curve for train and validation set (b) Loss curves for train and validation set.

[6] https://www.gov.br/saude/pt-br

[7] https://edition.cnn.com/2020/03/13/politics/donald-trump-emergency/index.html

[8] https://archive.vn/20200428073138/https://banningca.gov/DocumentCenter/View/7139/CDC_COVID19-Weekly-Key_Messages_03292020_FINAL

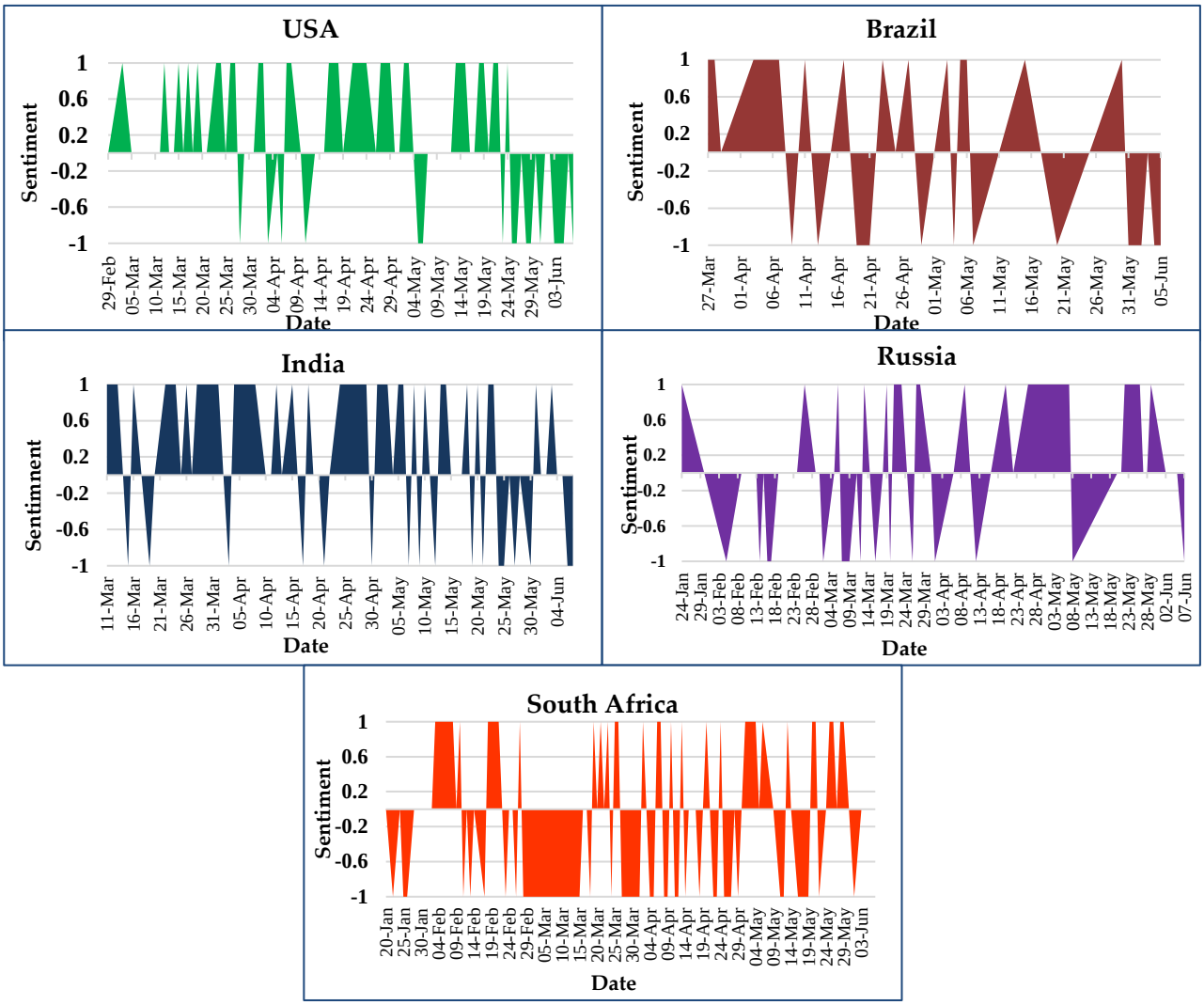


Fig. 5. Visualizing the sentiments of people over time in top five worst affected countries by COVID-19 pandemic.

being reported each day. Overall, 50.6% of positive sentiments, 28.3% of neutral sentiments, and 20.9% of negative sentiments were observed in India.

- Russia reported its first positive case on January 31, 2020. People have majorly shown neutral and negative sentiments during the mid of February. However, positive sentiments were expressed in April. Overall, 39.2% of neutral sentiment, 33.9% of positive sentiment, and 26.7% of negative sentiment were observed in Russia.
- In South Africa, majorly negative sentiments were being observed from late February till mid of March. People finally realized that the deadly virus had hit their country as the number of cases started rising sharply. However, in April, we can see some positive sentiments as citizens started motivating everyone by asking them to wear masks, read books, involve in daily exercise, and posting positive ideas by their former President Nelson Mandela. Overall, South Africa expressed 39.6% of negative sentiments, 32.6% of positive sentiments, and 27.7% of neutral sentiments.
- Finally, we conclude that overall, 38.3% of positive sentiments, 32.6% of neutral sentiment, and 28.3% of negative sentiments were expressed by these top five affected countries.

4.4 Baseline Comparison

We use several popular baseline methods for comparison with the proposed MACBiG-Net. The details are discussed as follows:

Table IV
Comparative Results of different methods on COVID-19 Sentiment Dataset (%)

Method	Accuracy	Precision	Recall	F1 score
LSTM [23]	72.6	73.3	73.6	73.4
Bi-LSTM [24]	73.2	74.9	72.0	73.4
CNN [25]	75.8	73.1	75.2	74.1
CNN-RNN [26]	77.7	78.6	74.2	76.3
HAN [22]	79.4	77.4	78.9	78.1
Self-attention-based LSTM [27]	70.9	68.2	66.7	67.4
RNN-Capsule [28]	71.5	68.5	70.8	69.6
BERT [29]	76.6	75.2	74.8	74.9
MACBiG-Net (Ours)	**81.5**	**81.3**	**81.2**	**81.2**

- **LSTM** [23]**:** The classical long-short memory network is applied. Word vectors are initialized by Glove. The hidden layer size is 300 units. Adam optimizer with learning rate of 0.001 is used.
- **Bi-LSTM** [24]**:** This represents standard bi-directional LSTM. We follow the same settings as described above for LSTM.
- **CNN** [25]**:** Convolutional neural network with multiple filters of size [3,4,5], each having 100 filters is applied. Adam optimizer is used with a learning rate of 0.001.
- **CNN-RNN** [26]**:** We successively stack the CNN layer with max-pooling layer followed by RNN units. For RNN, we experimented with LSTM and GRU variants. The GRU based integration shows the highest performance.
- **HAN** [22]**:** The hierarchical attention network uses two LSTM layer followed by attention mechanism at word-level and sentence-level for composing the final text representation. The HAN is the simplest form of our proposed approach as it does not encounter localized features.
- **Self-attention based LSTM** [27]**:** The self-attention mechanism is used with Bi-LSTM, which learns the contribution of each hidden state by providing a set of summarized weight vectors for each hidden state. A learning rate of 0.06 is used.
- **RNN-Capsule** [28]**:** We apply capsule networks where the input instance representation is taken from the LSTM unit. The dimension of hidden vectors is set to 256. The attention mechanism is used to construct the capsule representation inside another capsule. All word vectors are initialized with Glove. Adam optimizer is used with a learning rate of 0.001.
- **BERT** [29]: We use a pre-trained Bidirectional Encoder Representation from Transformers (BERT) model with the help of an online released TensorFlow[9] library. The transformer blocks are set to 12, hidden size to 768, and self-attention head to 12.

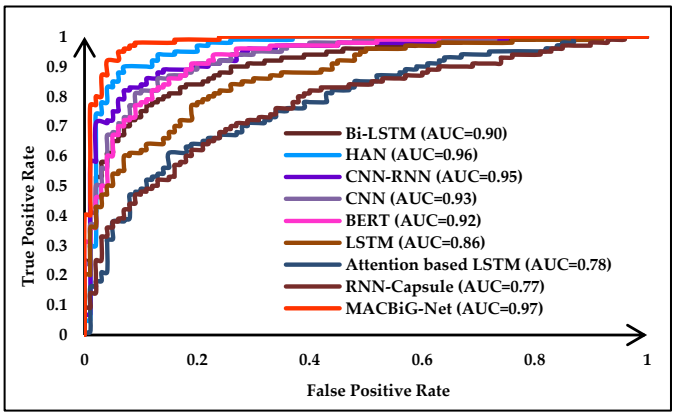


Fig. 6. Comparison of MACBiG-Net with baseline methods in terms of ROC curve and Area under the curve (AUC).

As evident from Table IV, the proposed network has outperformed the baseline methods and recent works in terms of sentiment classification by achieving 81.5% average accuracy. However, the HAN [22] and CNN-RNN [26] provides good enough results for the classification. This serves as the primary motivation of our proposed MACBiG-Net, which inherits the basic idea of both the methods and achieves nearly 2% and 4% higher accuracy compared to [22] and [26], respectively. This can also be visualized from the ROC curves and area under the curve (AUC), shown in Fig. 6, where we plot the ROC curves for the baseline methods. The AUC values help in comparing the ROC curves in a better way. From the AUC values, it is clear that MACBiG-Net has consistently outperformed all the previous methods for sentiment classification.

4.5 Qualitative Analysis

In this section, qualitative analysis is performed to visualize the word-level attention and extract the positive and negative sentiment topics that were generated during the pandemic.

4.5.1 Visualizing Word-level Attention

In this section, qualitative analysis is performed on MACBiG-Net, to investigate the role of word-level attention for the sentiment classification. We randomly picked four tweets from our dataset, two of them were positive, and two were negative. Thus, in Fig. 7, (a) – (b) belongs to the positive sentiment class (shown in Red color), and (c) – (d) belongs to the negative class (shown in Blue color). The color signifies the attention weights of each word. The darker the color in the respective category, the higher amount of attention is given to the word.

We can see that attention focuses on the prominent part of the sentence, by giving darker color to the most important word. E.g., In review sentences (a) - (b), positive sentiment is conveyed by the crucial words like *'brilliant'*, *'happy'*, *'healthy'*, *'safety'*, *'good'*, which have received more score by the attention mechanism for better sentiment prediction. Similarly, in reviews (c)-(d), negative sentiment is reflected by words like *'scary'*, *'pandemic'*, *'death'*, which again have received more attention score. The words which convey no informative meaning (like stop words), were given very low weights by highlighting them with light colors. This shows that the attention weights focus on those words which have more relation with the output sentiment category. Thus, the visualization shows how the attention weights are changing along with the words in a sentence.

4.5.2 Visualizing positive and negative sentiment topics

To get an in-depth analysis on the opinion of the people, we apply LDA (Latent Dirichlet Allocation) based topic modeling to determine the set of words (topics) that contribute to generate the positive and negative sentiments. This will help in finding the hidden thematic structure of the document and drawing essential conclusions from them. We use LDAvis [30] tool for visualization, which further helps in finding the prevalent topics in a document, relation between the topics, and the conditional distribution of a given term over the set of topics. The optimal number of topics was determined by building many LDA models with a different number of topics (k) and selecting

[9] https://github.com/tensorflow/tensor2tensor

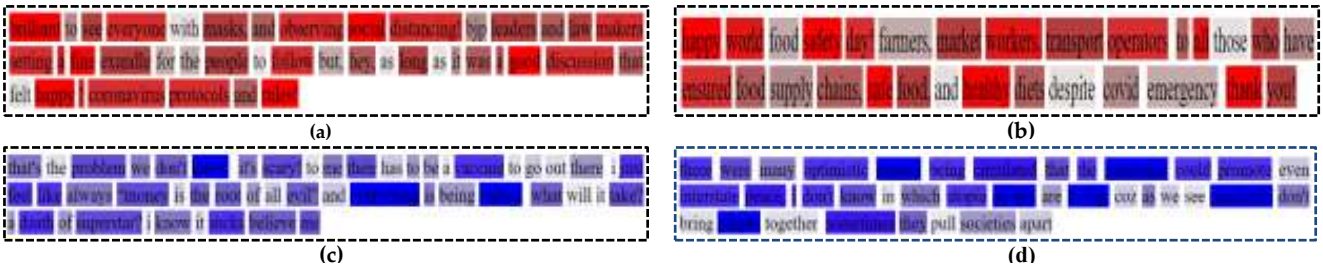


Fig. 7. Visualization the word-level attention weights from proposed dataset: (a)-(b) Positive Sentiment (c)-(d) Negative Sentiment.

the one with the highest coherence score [31]. This gives us k =10. The layout of LDAvis is shown in Fig. 8 (a)-(b) for positive and negative sentiment labels, respectively.

If we consider (a), the entire layout is divided into two panels: the left panel shows the ten topics in the form of circles where the similar topics are plotted close to each other, the right panel shows the bargraphs which represent the terms or words that are most useful for understanding the currently selected topic (from the left panel). The length of the gray bars depicts the entire document-wise frequencies of each term, and the red bars depict the topic-wise frequencies of each term. Thus, for a given topic, we can compare the length of the red and gray bars to understand the relevance of the topic. Further, selecting a term on the right panel tells the conditional distribution of the selected term over the topics (on the left panel). In this way, topic modeling helps in explaining the underlying information in a document. Hence, we summarize our observations as follows:

- The prevalent positive words related to COVID-19 on twitter are: 'good', 'stay safe', 'happy', 'best', 'thank you', 'support', 'doctors', 'mask', 'together', 'proud' and the popular negative words linked to COVID-19 on twitter are: 'virus', 'deaths', 'bad', 'trump', 'help', 'lost', 'fake', 'died', 'outbreak', 'infected', 'hate', 'worst', 'crisis', 'fight'.
- Many positive tweets were directed towards the frontline workers where users were motivating and boosting the morale of the health care workers by appreciating them, donating them masks, dedicating some songs, providing free accommodation to them, and donating money to restaurants for distributing meals to essential workers.
- In some countries like Brazil, health care workers celebrate the discharge of the first COVID-19 patient from the hospital. In India, many citizens were delighted to see everyone unified with candles or flashlights in their hands from their balconies.
- Finally, positive sentiments were directed towards maintaining a healthy diet, reuniting with families after travel restrictions were lifted, boosting the immune system, asking everyone to stay safe, binge-watching television series, encouraging everyone to take precautions, wishing and motivating the users on special occasions by posting positive messages on occasions like Birthdays, Mothers Day, Earth Day, World Bicycle Day, Memorial Day (celebrated in the US), World Labor Day, International Nurses Day, and staying together to beat the deadly virus.
- During the pandemic, several social-economic factors also contributed to generate negative sentiments in people all around the world. In the US, the negative tweets were mainly focused on the racial injustice related to the death of George Floyd, which resulted in massive protests and violence even during the time of the pandemic. People fear that these mass demonstrations may occur in an uptick in the COVID-19 cases.
- In India, Mumbai (one of the worst-hit cities by COVID-19), was severely affected by cyclone winds. In West Bengal, Cyclone Amphan caused extensive damage. The poor, jobless migrants went traveling back to their home by walking thousands of kilometers due to lockdown restrictions in India.
- In general, many people also lost their jobs. The unemployment rate was increased during the pandemic. People were bored and anxious during home-quarantine. The upsurge in infected cases also created fear and stress among people. Further, lots of fake news[10] was generated in social media, creating panic about the disease.
- The popular positive hashtags are: #salutetocoronafighters, #inthistogether, #lovemycountry, #withme, #thankyouindia, #indiafightscorona, #flattenthecurve, #fightthevirus, and #frontlineheroes and negative hashtags are: #DictatorTrump, #BlackLivesMatter, #GeorgeFloyd, #AmericaInCrisis, #Riot2020, #migrantlivesmatter, #pandemia, #chinavirus, #wuhanvirus, and #fuck_coronavirus.

5 Conclusion

This paper is aimed to analyze the impact of the novel coronavirus on the people by extracting and analyzing the sentiments of the top five most affected countries with the virus. We propose MACBiG-Net, which aims to extract the positive, negative, and neutral sentiment using word-level encoded attention and sentence level encoded attention. This is achieved by developing and labeling the COVID-19 Sentiment Dataset, which contains 4118 labeled tweets crawled from Twitter. Further, we visualize the sentiments of the people from January 31 to June 7 and summarize essential observations. We conclude that during the pandemic, lots of sentiments were being generated, and people were posting their views or opinions on several topics like frontline workers, upsurge in active cases, travel restrictions, and about the virus itself. Much fake news was also generated during this time, which was misleading the

[10] https://timesofindia.indiatimes.com/blogs/toi-editorials/social-media-menace-beware-of-fake-news-going-viral-faster-than-covid-19-it-will-cost-lives-literally/

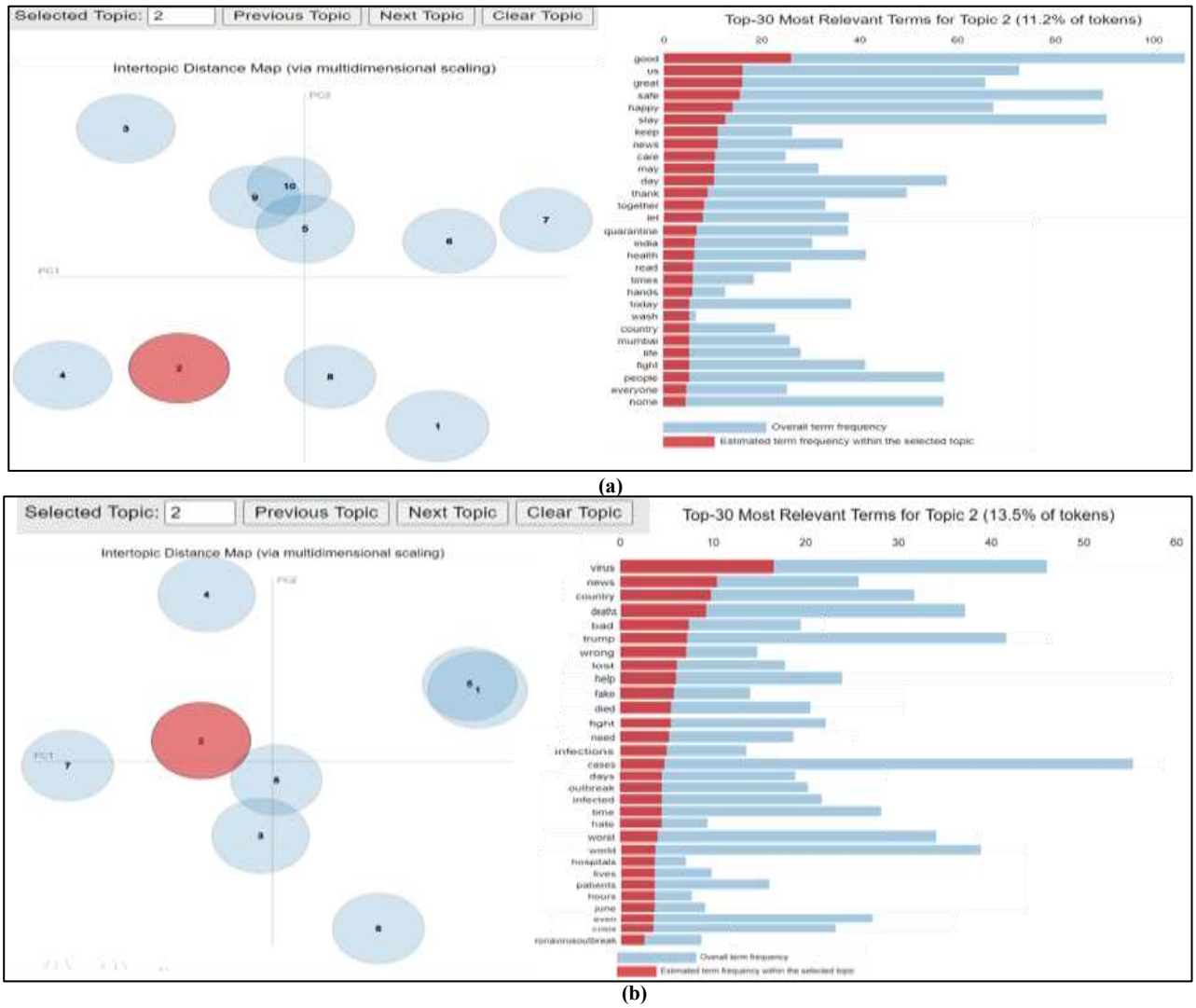


Fig. 8. Identifying the hidden information in document through LDAvis visualization for (a) Positive and (b) Negative sentiment

people. However, despite being stressed, people appreciated the efforts of all frontline workers and motivated each other to follow all precautions.

In the future, we aim to expand our study to analyze the impact of the virus on the rest of the world by doing in-depth research on several strategies like home-quarantine, work from home, social distancing undertaken to fight the deadly virus. We also aim to study the sentiments generated by other popular nations in the world.

REFERENCES

[1] R. M. Anderson, H. Heesterbeek, D. Klinkenberg and T. D. Hollingsworth, "How will country-based mitigation measures influence the course of the COVID-19 epidemic?," *The Lancet,* vol. 395, no. 10228, pp. 931-934, 2020.

[2] J. Hellewell, S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russell, J. D. Munday, A. J. Kucharski and W. J. Edmunds, "Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts," *The Lancet Global Health,* vol. 8, no. 4, pp. e488-e496, 2020.

[3] A. Yadav and D. K. Vishwakarma, "A comparative study on bio-inspired algorithms for sentiment analysis," *Cluster Computing,* p. 1–21, 2020.

[4] M. Tsytsarau and T. Palpanas, "Managing Diverse Sentiments at Large Scale," *IEEE Transactions on Knowledge and Data Engineering,* vol. 28, no. 11, pp. 3028-3040, 2016.

[5] S. M. Mohammad, P. Sobhani and S. Kiritchenko, "Stance and sentiment in tweets," *ACM Transactions on Internet Technology (TOIT),* vol. 17, no. 3, pp. 1-23, 2017.

[6] W. Zhao, Z. Guan, L. Chen, X. He, D. Cai, B. Wang and Q. Wang, "Weakly-supervised deep embedding for product review sentiment analysis," *IEEE Transactions on Knowledge and Data Engineering,* vol. 30, no. 1, pp. 185-197, 2017.

[7] T.-c. Huang , R. N. Zaeem and K. S. Barber, "It Is an Equal Failing to Trust Everybody and to Trust Nobody: Stock Price Prediction Using Trust Filters and Enhanced User Sentiment on Twitter," *ACM Transactions on Internet Technology (TOIT),* vol. 19, no. 4, pp. 1-20, 2019.

[8] C. Huang, W. Jiang, J. Wu and G. Wang, "Personalized Review Recommendation based on Users' Aspect Sentiment," *ACM Transactions on Internet Technology (TOIT),* vol. 20, no. 4, pp. 1-26, 2020.

[9] A. Yadav and D. K. Vishwakarma, "A deep learning architecture of RA-DLNet for visual sentiment analysis," *Multimedia Systems,* vol. 26, p. 431–451, 2020.

[10] K. Denecke and Y. Deng, "Sentiment analysis in medical settings: New opportunities and challenges," *Artificial Intelligence in Medicine,* vol. 64, pp. 17-27, 2015.

[11] I. Korkontzelos, A. Nikfarjam, M. Shardlow, A. Sarker, S. Ananiadou and G. H. Gonzalez, "Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts," *Journal of Biomedical Informatics,* vol. 62, pp. 148-158, 2016.

[12] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: a review," *Artificial Intelligence Review,* vol. 53, no. 6, pp. 4335-4385, 2019.

[13] Y. Wang, Q. Chen, M. Ahmed, Z. Li, W. Pan and H. Liu, "Joint Inference for Aspect-level Sentiment Analysis by Deep Neural Networks and Linguistic Hints," *IEEE Transactions on Knowledge and Data Engineering,* pp. 1-12, 2019.

[14] F.-C. Yang, A. J. Lee and &. S.-C. Kuo, "Mining Health Social Media with Sentiment Analysis," *Journal of medical systems,* vol. 40, no. 11, pp. 1-8, 2016.

[15] S. Sabra, K. M. Malik and M. Alobaidi, "Prediction of venous thromboembolism using semantic and sentiment analyses of clinical narratives," *Computers in Biology and Medicine,* vol. 94, pp. 1-10, 2018.

[16] M. Moh, T. Moh, . Y. Peng and L. Wu, "On adverse drug event extractions using twitter sentiment analysis," *Network Modeling Analysis in Health Informatics and Bioinformatics,* vol. 6, no. 1, pp. 1-12, 2017.

[17] S. M. Jiménez-Zafra, M. T. Martín-Valdivia, M. D. Molina-González and L. A. Ureña-López, "How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for medical domain," *Artificial intelligence in medicine,* vol. 93, pp. 50-57, 2019.

[18] N. Limsopatham and N. Collier, "Normalising Medical Concepts in Social Media Texts by Learning Semantic Representation," *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics,* p. 1014–1023, 2016.

[19] Y. Chen, B. Zhou, W. Zhang, G. Gong and G. Sun, "Sentiment Analysis Based on Deep Learning and Its Application in Screening for Perinatal Depression," *IEEE Third International Conference on Data Science in Cyberspace,* pp. 451-456, 2018.

[20] H. Grisstte and E. Nfaoui, "Daily life patients Sentiment Analysis model based on well-encoded embedding vocabulary for related-medication text," *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining,* pp. 921-928, 2019.

[21] H. Talpada, M. N. Halgamuge and N. T. Q. Vinh, "An Analysis on Use of Deep Learning and Lexical-Semantic Based Sentiment Analysis Method on Twitter Data to Understand the Demographic Trend of Telemedicine," *11th International Conference on Knowledge and Systems Engineering (KSE), Vietnam,* pp. 1-9, 2019.

[22] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola and E. Hovy, "Hierarchical attention networks for document classification," *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, San Diego,* pp. 1480-1489, 2016.

[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation,* vol. 9, no. 8, pp. 1735-1780, 1997.

[24] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural networks,* vol. 18, no. 5-6, pp. 602-610, 2005.

[25] Y. Kim, "Convolutional neural networks for sentence classification," *Empirical Methods in Natural Language Processing,* pp. 1746-1751, 2014.

[26] S. Lai, L. Xu, K. Liu and J. Zhao, "Recurrent convolutional neural networks for text classification," *Twenty-ninth AAAI conference on artificial intelligence,* p. 2267–2273, 2015.

[27] Z. Lin, M. Feng, C. Nogueira, M. Yu, B. Xiang, B. Zhou and Y. Bengio, "A structured self-attentive sentence embedding," *International Conference on Learning Representations,* 2017.

[28] Y. Wang, A. Sun, J. Han, Y. Liu and X. Zhu, "Sentiment Analysis by Capsules," *Proceedings of the 2018 world wide web conference,* pp. 1165-1174, 2018.

[29] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT),* p. 4171–4186, 2019.

[30] C. Sievert and K. S. Shirley, "LDAvis: A method for visualizing and interpreting topics," *Proceedings of the workshop on interactive language learning, visualization, and interfaces,* pp. 63-70, 2014.

[31] K. Stevens, P. Kegelmeyer, D. Andrzejewski and D. Buttler, "Exploring topic coherence over many models and many topics," *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning,* pp. 952-961, 2012.



Ashima Yadav received B.Sc.(Hons) in Computer Science from University of Delhi, New Delhi, India in 2013, and M.C.A. from Guru Gobind Singh Indraprastha University, New Delhi, India in the year 2016. She is currently working towards the Ph.D. degree from the Department of Information Technology, Delhi Technological University, New Delhi, India. Her current research interest includes deep learning, natural language processing, machine learning, Emotion Recognition, and sentiment analysis.



Dinesh Kumar Vishwakarma (M'16, SM'19) received the B.Tech. degree from Dr. Ram Manohar Lohia Avadh University, Faizabad, India, in 2002, the M.Tech. degree from the Motilal Nehru National Institute of Technology, Allahabad, India, in 2005, and the Ph.D. degree degree in the field of Computer Vision and Machine Learning from Delhi Technological University University (Formerly Delhi College of Engineering), New Delhi, India, in 2016. He is currently an Associate Professor with the Department of Information Technology, Delhi Technological University, New Delhi. His current research interests include Computer Vision, Machine Learning, Deep Learning, Sentiment Analysis, Fake News and Rumour Analysis, Crowd Behaviour Analysis, Person Re-Identification, Human Action and Activity Recognition. He is a reviewer of various Journals/Transactions of IEEE, Elsevier, and Springer. He has been awarded with "Research Excellence Award" by Delhi Technological University, Delhi, India in 2018, 2019 and 2020.

A Novel Trench FinFET as biosensor for early detection of dengue fever.

Aditya M and Rishu Chaujar*
* Delhi Technological University
E-mail: adityam98@gmail.com,chaujar.rishu@dtu.ac.in

***Abstract*— In this work we propose and simulate a novel Trenched FinFET for the detection of biomolecules, which may be used as a basis for the diagnosis of diseases such as dengue fever based on their marker molecules. The device is based on the dielectric properties of these molecules and all comparisons have been made with air as a reference. A shift in the threshold voltage (Vth) is observed in the presence of the analyte molecule and increases from 0.392 V to 0.556 V with the increase in its dielectric constant. Similarly, it is also found that the change in the Switching ratio (Ion/Ioff) is also significant and follows an increasing trend from 26.74 to 2.63 x 103 with the dielectric constant. The saturation current of the device increases from 3.64 x 10-4A to 4.32 x 10-4A with increase in the dielectric constant of the analyte.**

***Keywords*— Sensor, Biosensor, Sensitivity, FinFET, Dielectric constant**

1. Introduction

Dengue is a mosquito-borne viral disease that has rapidly spread all over the world in recent years. The dengue virus is transmitted by female mosquitoes mainly of the species Aedes. Symptoms of dengue can vary from subclinical illness to extreme flu-like symptoms in those infected. Many people experience extreme dengue, associated with serious bleeding, organ damage and/or plasma leakage. Severe dengue has a higher risk of death when not managed appropriately.

The ability to determine a patient's health, onset and development of illness, and track treatment through a non-invasive approach is a goal to be accomplished in healthcare. To achieve this goal, the prerequisites are: unique biomarkers suggesting a stable or diseased state [1]. Biosensor technologies offer the ability to meet those requirements through a merger of nanotechnology, chemistry and medical science [2]. FET biosensors, due to their good scalability, high sensitivity, rapid electrical detection, low power consumption, direct electrical readings and affordability compared to conventional methods, have attracted considerable attention in recent years [5]. The change in electrical parameters the introduction of biomolecules on the device, are responsible for bio-sensing action [6]. APTES((3-Aminopropyl) triethoxysilane) functions as a facilitator to immobilize the NS1 antigen on the surface in the detection process of dengue virus. The proposed biosensor can be used as APTES based sensor [7].

In this work we present a novel Trench-FinFET with its fin modified to include a hollow trench for the detection of proteins and potentially other biomolecules that show variation in properties for the detection of dengue.

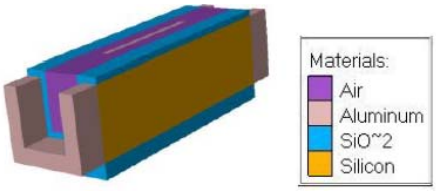


Fig. 1. 3D view of the simulated Trench-FinFET.

Table 1. Device Parameters

Parameter	Value
Channel Length (L_g)	40 *nm*
Fin Height (H_f)	23 *nm*
Fin Width (W_f)	20 *nm*
Thickness Oxide (t_{ox})	2 *nm*
Trench Width	5 *nm*
Trench Depth	18 *nm*
Source and Drain Length	20 *nm*
Source and Drain Doping Concentration	10^{21} cm^{-3}
Gate Work function	4.26 *eV*

2. Device Structure

The Device structure as shown in Fig 1 with parameters given in Table 1 was simulated using DEVEDIT 3D by SILVACO [8]. The

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on January 05,2021 at 06:59:29 UTC from IEEE Xplore. Restrictions apply.

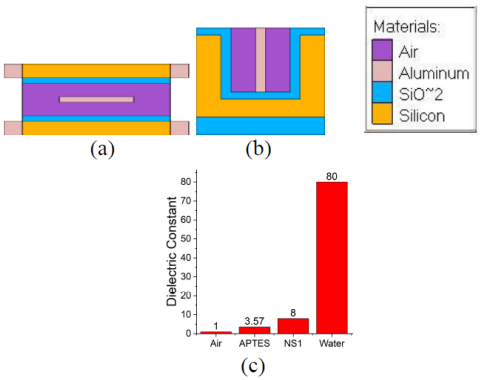


Fig. 2. (a) Lateral (b) Transverse section of the device.

simulated device is a Trench-FinFET with a fin that is trenched in the middle to accommodate biomolecules in it. The trench-like structure of the fin allows the target biomolecules that need to be detected to lodge themselves inside the cavity on the fin thus allowing them to be detected. The device simulated here has an 80 nm long fin that is covered on the surface externally by 2 nm thick SiO_2. The gate length was taken as 40 nm and an Aluminium gate with a work function of 4.26 eV was used. The width of the Silicon fin was 26 nm and the height was 23 nm. The dimensions of the

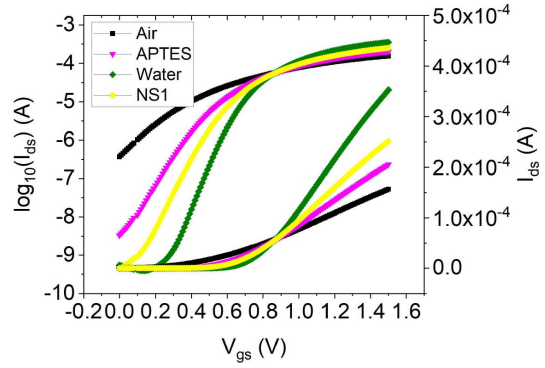


Fig. 3. (a) Transfer characteristics of the device for different molecules at $V_{ds}=0.5V$. **(b)** g_m of the device for different molecules at $V_{ds}=0.5V$.

trench was 80 nm x 18 nm x 5 nm as shown in Fig 2. The fin was doped with a uniform n-type doping of 10^{15} cm^{-3}. The source and drain were doped with n-type doping of 10^{21} cm^{-3}. The characteristics of the device were simulated using SILVACO ATLAS Device Simulator [9]. The models used in the device simulation were Concentration Dependent SHR Recombination Model, Concentration Dependent Arora Mobility Model and Fermi-Dirac Statistics. [9].

3. Methodology

The analysis methodology used in this work is the variation of the dielectric constant inside the trench and studying the variation of the electrical properties. This methodology is used to simulate the presence of biomolecules inside the cavity thus enabling the device's action as a biosensor. It is known that the dielectric constant of water is 80. APTES functions as a facilitator to immobilize NS1 antigen on the surface in the detection process of dengue virus thus causing a change in the electrical characteristics [7]. NS1 is a protein marker for the dengue virus and the dielectric constant of the proteins are known to be in the range of 6-10 thus an average value of 8 is assumed in this work [7].

4. Results and Discussion

A. Transfer Characteristics

The I_{ds} *vs* V_{gs} curves for the molecules were simulated at $V_{ds}=0.5V$ and the shift in the V_{th} from air is calculated and are shown in Fig 3. The increase in the Drain Current (I_{ds}) is attributed to the fact that the Electric Field applied to the fin by the gate is able to permeate more into the fin due to the presence of the biomolecule simulated by the dielectric material in the trench. The increase in the dielectric constant of the cavity causes the effective thickness of the oxide to decrease thus increasing the capacitance between the gate and the fin.

B. Output Characteristics

The Ids vs Vds curves for the device were simulated for the molecules at Vgs=1.5V as shown in Fig 4. The switching ratio and the sensitivity were calculated from the curves. The sensitivity of the device is defined as shown in Equation (1).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on January 05,2021 at 06:59:29 UTC from IEEE Xplore. Restrictions apply.

$$S = \frac{I_{on(k>1)}}{I_{on(k=1)}} \qquad (1)$$

It is seen that the Drain Current (Ids) increases with the increase in the dielectric constant (k), this is due to the increase in the effective Electric Field inside the fin that causes better channel formation. The sensitivity of the device also increases as the dielectric constant increases. This is due to the increase in the Drain Current with increase in dielectric constant (k).

C. Electric Field Density

The Electric Field Distributions are shown in Fig 6. It can be seen that the Electric Field in the trench increases with the increase in dielectric constant keeping the applied gate voltage (Vgs) constant. This is due to the decrease in the effective width of the oxide when a biomolecule is inside the hollow cavity. This causes the effective thickness of the oxide to decrease thus increasing the capacitance between the gate and the fin. This in turn increases the Electric Field in the fin thus resulting in better channel formation.

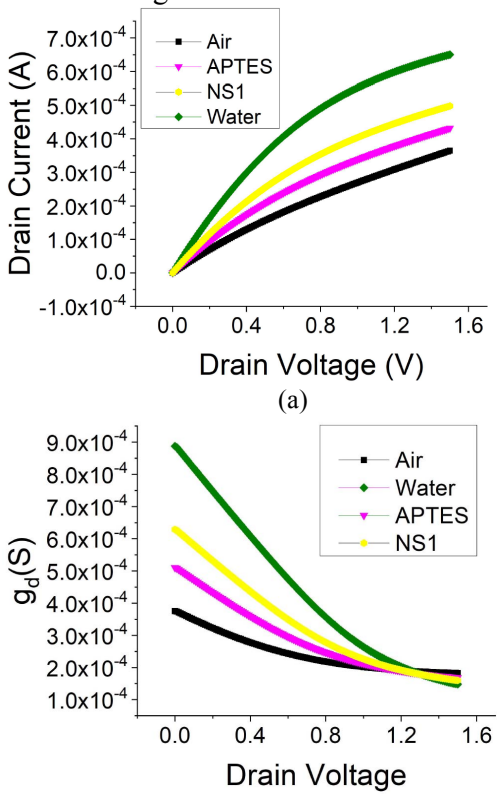


Fig. 4. (a) Output characteristics of the device at different molecules at V_{gs}=*1.5V*. **(b)** g_d of the device for different molecules at V_{gs}=*1.5V*.

Table 2. Extracted Parameters

Medium	V_{th}(V)	ΔV_{th}(V)	Sensitivity	SR
Air	0.392	0	1	26.74
APTES	0.556	0.164	1.18	2635.54
NS1	0.616	0.224	1.25	14666.66
Water	0.679	0.287	1.78	15966.85

D. Electric Field Density

The Electron Current Density in the Z direction is shown in Fig 6. It can be observed that the current density increases with the increase in the dielectric constant inside the hollow cavity. This increase can be attributed to the increase in the Electric Field that can permeate into the fin with the dielectric constant as shown in Fig 5.

E. Sensitivity and ΔV_{th}

The Sensitivity defined in Equation (1) and Threshold Voltage shift in the Device are shown in Table 2. The Threshold Voltage shift due to the presence of the molecule in the trench is calculated with respect to air. It is seen that the Shift increases with a

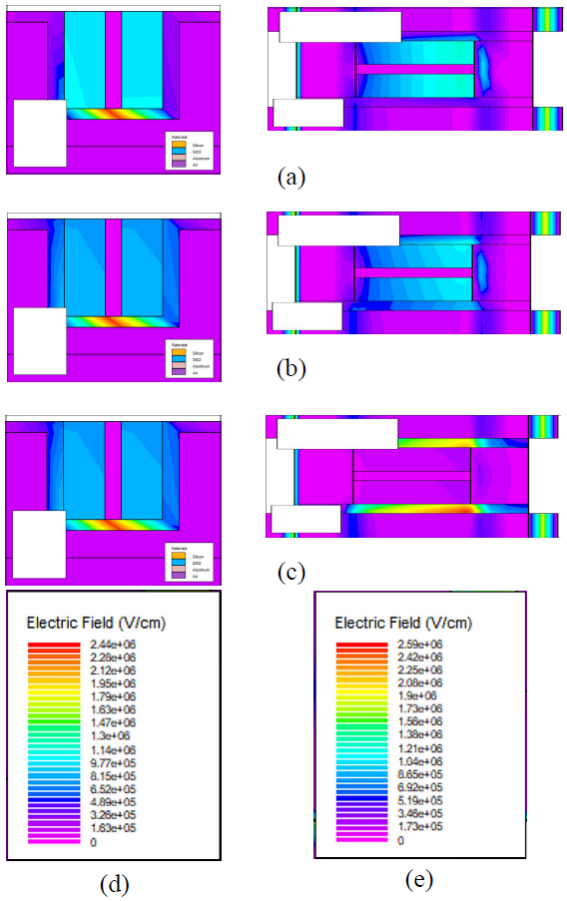


Fig. 5. Electric Field in **(a)**Air **(b)**APTES **(c)**water **(d)**Transverse E scale **(e)**Longitudinal E scale.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on January 05,2021 at 06:59:29 UTC from IEEE Xplore. Restrictions apply.

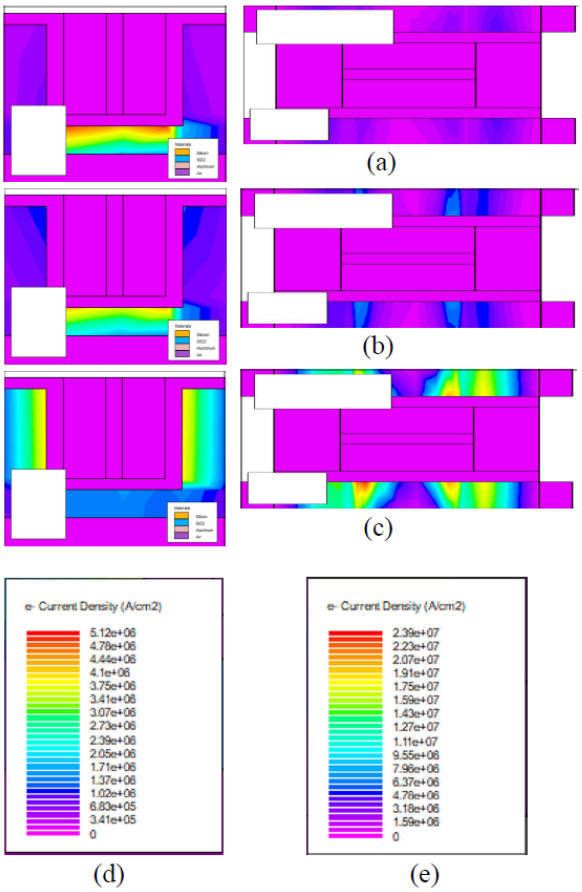


Fig. 6. Electron Current Density in **(a)**Air **(b)**APTES **(c)**water **(d)**Transverse e- scale **(e)**Longitudinal e- scale

increase in the dielectric constant of the molecule in the trench, this can be attributed to the increase in the Electric field.

$$SR = \frac{I_{on}}{I_{off}} \qquad (2)$$

The Switching ratio defined in Equation (2) increases due to the rapid increase in the I_{on} of the device with the increase in the dielectric constant of the molecules.

F. Conclusion

We have simulated a novel Trench-FinFET for detection of biomolecules that can be used for early diagnosis of dengue. The study involves the response of the device when the trench is filled with APTES. The threshold voltage shift (ΔV_{th}) follows an increasing trend with dielectric constant and maximum ΔV_{th} of *0.556 V* for APTES. The ON-State Drain current of the device also increases from 3.64 x 10-4A in air to 4.32 x 10^{-4}A for APTES. The switching ratio of the device also increases from 26.74 in air to 2.63 x 10^4 in APTES. The Electric field and Current Density in the fin also increases with dielectric constant. The sensitivity of the device increases to 1.28 for APTES. These changes in the electrical response of the device are identified to be primarily caused by the changes in the physical parameters that are observed in the device due to the presence of APTES and form the basis of it acting as a biosensor.

Acknowledgments.

This work was supported by Delhi Technological University. The author is grateful for the support provided from the Microelectronics Lab for this research.

References

[1]E. Comini, C. Baratto, G. Faglia, M. Ferroni, A.Ponzoni, D. Zappa, and G. Sberveglieri, "Metal oxidenanowire chemical and biochemical sensors," Journalof Materials Research, vol. 28, no. 21, pp. 2911–2931,2013.

[2]J. Lee, J. Jang, B. Choi, J. Yoon, J.-Y Kim,Y-KChoi, D.M. Kim, D.H Kim, S.-J. Choi: A highlyresponsive silicon nanowire/amplifier MOSFEThybrid biosensor. Sci. Rep. **5**, 12286 (2015)

[3]Bergveld, P.: The development and application ofFET-based biosensors. Biosensors **2**(1), 15–33 (1986)

[4]Limonta, D., Falcón, V., Torres, G. *et al.* Denguevirus identification by transmission electronmicroscopy and molecular methods in fatal denguehemorrhagic fever. *Infection* **40,** 689–694 (2012).https://doi.org/10.1007/s15010-012-0260-7

[5]Sang, S., Wang, Y., Feng, Q., Wei, Y., Ji, J.,Zhang, W.: Progress of new label-free techniques forbiosensors: a review. Crit. Rev. Biotechnol. **36**(3),465–481 (2016)

[6]Im, H., Huang, X.-J., Gu, B., Choi, Y.-K.: Adielectric-modulated field-effect transistor forbiosensing. Nat. Nanotechnol. **2**(7), 430–434 (2007)

[7] D.Wasik, ,A.Mulchandani, , M.Yates, (2018). Salivary Detection of Dengue Virus NS1 Protein with a Label-Free Immunosensor for Early Dengue Diagnosis. Sensors. 18. 2641. 10.3390/s18082641.

[8]PH Drive, S. Clara , Atlas User's Manual, DeviceSimulation Software; 2017. p. 1– 1715.

[9]PH Drive, S. Clara , DEVEDIT User's Manual,Device Structure Editing Software; 2017. p. 1– 1715.

[10]A. Kumar, M.M.Tripathi,R. Chaujar, "Sub-30nmIn2O5Sn gate electrode recessed channelMOSFET: A biosensor for early stage diagnostics."*Vacuum*. Vol. 164, pp. 46-52, Jun 201910.1016/.vacuum.2019.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on January 05,2021 at 06:59:29 UTC from IEEE Xplore. Restrictions apply.

Research on Biomedical Engineering
https://doi.org/10.1007/s42600-020-00108-1

ORIGINAL ARTICLE

A robust digital ECG signal watermarking and compression using biorthogonal wavelet transform

Ashish Kumar[1] · Ramana Ranganatham[2] · Samayveer Singh[3] · Rama Komaragiri[2] · Manjeet Kumar[4]

Received: 7 April 2020 / Accepted: 19 October 2020
© Sociedade Brasileira de Engenharia Biomedica 2020

Abstract

Purpose With the advent of the technological era, the complexity of data security and privacy has gained utmost importance. People are becoming more conscious about the security and confidentiality of their crucial information.

Methods The present study is first of its kind effort in for safeguarding an individual's electrocardiogram (ECG) data. ECG data not only does hold medical significance but also is of crucial and critical importance for purposes related to biometric information. Hence, a deciding factor for recognizing and validating personal information. The study progresses and proposes a new technique in ECG watermarking compression.

Results This method ensures high reliability of data and information with confidentiality. The efficiency of the proposed research is evaluated in terms of compression ratio (CR), compressed signal-to-noise-ratio (CSNR), root-mean-square error (RMSE), maximum absolute error, percentage root-mean-square difference (PRD), and quality score.

Conclusion After a comprehensive evaluation, the final results show that the proposed work is robust and feasible.

Keywords Electrocardiogram · Biorthogonal wavelet transform · ECG watermarking

Introduction

Protecting personal data (information) on the Internet seems like becoming a thing of past. Soon, individuals will be struggling to maintain their personal medical information on the Internet. The trend shows (Ibaida and Khalil, 2013) ECG as one of the most vulnerable data set is widely put to use. This increasing use of data on the Internet is also exerting necessitating pressure on the World Wide Web. To safeguard and protect the data, use of watermarks to ECG data is the focus of the present research study. An ECG is the most straightforward test among the plethora of heart activity monitoring, which lists the health of the heart. ECG signal helps take diagnostic and preventive measures. Such data is generated in large quantities on day to day basis. Hence, data storage with privacy and accuracy is one big challenge for today's information technology (IT) professionals engaged in the development of medical software applications. Many studies are being conducted to secure the ECG data. To continuously monitor a subject remotely, lossless compression of ECG data along with watermarking is gaining importance, which is the motivation for our research.

Various studies have been proposed in the past few years that individually perform either ECG compression or ECG watermarking. However, there is only a single study available that includes both ECG compression and watermarking (Tseng et al. 2014). Digital watermarking techniques are classified into spatial domain based, frequency domain based, or

✉ Manjeet Kumar
manjeetchhillar@gmail.com

Ashish Kumar
akumar.1june@gmail.com

Ramana Ranganatham
ramana.ranganatham@bennett.edu.in

Samayveer Singh
samayveersingh@gmail.com

Rama Komaragiri
rama.komaragiri@bennett.edu.in

[1] School of Electronics Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India

[2] Department of Electronics and Communication Engineering, Bennett University, Greater Noida, U.P., India

[3] Department of Computer Science and Engineering, National Institute of Technology, Jalandhar, Punjab, India

[4] Department of Electronics and Communication Engineering, Delhi Technological University (DTU), Delhi, India

Published online: 31 October 2020

both spatial and frequency domain based (Swanson et al. 1998). Frequency domain watermarking techniques, namely, discrete Fourier transform (DFT), discrete cosine transform (DCT), and discrete wavelet transform (DWT), are preferred over spatial domain watermarking techniques because they provide better resistance against attacks like contrast adjustment, lossy compression, and brightness variation to name a few (Mukherjee et al. 2004; Bender et al. 1996). The primary concern with spatial domain techniques is that they show a lack of resistance against geometric attacks, whereas DFT watermarking techniques provide better resistance against geometric attacks. DFT watermarking techniques require a change in transform coefficients which affect the entire ECG signal (Bender et al. 1996; Muharemagic and Furht 2006). ECG signals exhibit slowly changing trends or oscillations punctuated with transients. As ECG signals consist of abrupt changes, DWT, which is a well-localized function in time and frequency, is used. However, DWT watermarking techniques are not robust against the geometric distortion attacks, like rotation, scaling, and translation. Odd-even embedding is a powerful data embedding technique explored for digital watermarking in different image and signal processing applications. Odd-even embedding technique embeds the secret data into the signal by changing the secret data value at most by '1'. Hybrid watermarking techniques reported in Mohammed et al. (2012); Liu and Tan (2002); Ganic and Eskicioglu (2004); Lai and Tsai (2010); Ansari et al. (2016); Ali and Ahn (2014); and Mishra et al. (2014) fail to provide good robustness with the better imperceptibility, and further, the information of the original signal is not retained. Hence, there is a need for a watermarking technique which can overcome the problems discussed above. A robust watermarking algorithm compared to the previously proposed approaches based on odd-even embedding and third-level decomposition of DWT is proposed in the present work.

Medical regulatory boards do not approve the application of lossy data compression techniques in ECG signal compression; hence, lossless data compression techniques with a high compression ratio are required. In the literature (Manikandan and Dandapat 2014; Deepu 2014; Kumar et al. 2013; Elgendi et al. 2017; Deepu 2016a), compression ratio (CR), space saving, compression score, signal-to-noise ratio (SNR), compression speed (CS), and memory usage are used as performance parameters on which the lossless ECG data compression technique is evaluated. Based on the performance parameters, discrete wavelet transform–based ECG data compression technique is found to be most suitable (Kim et al. 2006).

Wavelets have been around for more than a hundred years, but it is the last decade that wavelets have become quite popular in signal analysis. A wavelet is a mathematical function whose initial and final values are identically equal to zero. They are generally used to extract the time-frequency information from a signal through the use of continuous wavelet transform (CWT). In the discrete time domain, DWT is commonly used for denoising. Due to the highly correlated set of coefficients provided by DWT, high compression ratios can be achieved. The DWT is commonly implemented using a multirate filter bank that divides a signal into various subbands allowing the signal to be decomposed until the required frequency band is obtained. In order to reduce the redundancy in the decomposed signals, as shown in Fig. 1 (Kumar et al. 2018), the filter bank is made to be critically sampled by down-sampling the signal after every stage. The DWT is the basis of visual watermarking algorithms, as emphasized in the literature (Deepu et al. 2016a, 2016b; Zheng et al. 2009; Zigel et al. 2000; Patra et al. 2009; Chou et al. 2004; Chen et al. 2010). The outcome of DWT-based watermarking algorithms is efficient and vigorous, which can survive diverse and varied attacks. The low frequency embedded watermark signals can resist and severe attacks comprising of lowpass characteristics like lossy compression, filtering, and geometric distortions. Though the survival is made, the watermark signals are vulnerable to the moderation of the signal histogram, in the form of gamma correction and histogram equalization. When embedding the watermark signal in mid and high frequencies, the signal is further exposed to attacks such as lowpass filtering and lossy compression, thereby leading to severe geometric deformation of the signal. Though DWT-based watermarking is vulnerable to attacks, at the same time it provides distinct advantages against pounces such as noise addition and nonlinear deformations of the signal. Therefore, it is highly imperative for the designers to incorporate all sub-bands for embedding the watermarks, thus making the outlive for the watermark robust against such attacks.

Methods

To safeguard the ECG data, an even-odd embedding based ECG watermarking combined with DWT-based ECG compression is proposed in this paper. The proposed watermarking process allows information to be stored and extracted from the ECG signal without any loss. Due to this ability of the algorithm to conceal the data to be watermarked without modifying the ECG signal drastically, it becomes difficult to identify and extract the watermark without using the correct extraction algorithm. Thus, the proposed process allows the ECG signal to be tracked and identified to prevent misuse. There are many ECG databases available to evaluate the performance of ECG detection algorithms. Some of the standard ECG databases include MIT-BIH arrhythmia database, PTB database, QT database, Fantasia database, and long-term ST (LTST) database. Their main characteristics are summarized in Table 1.

The performance of the proposed method is evaluated using compression ratio (CR), compressed signal-to-noise

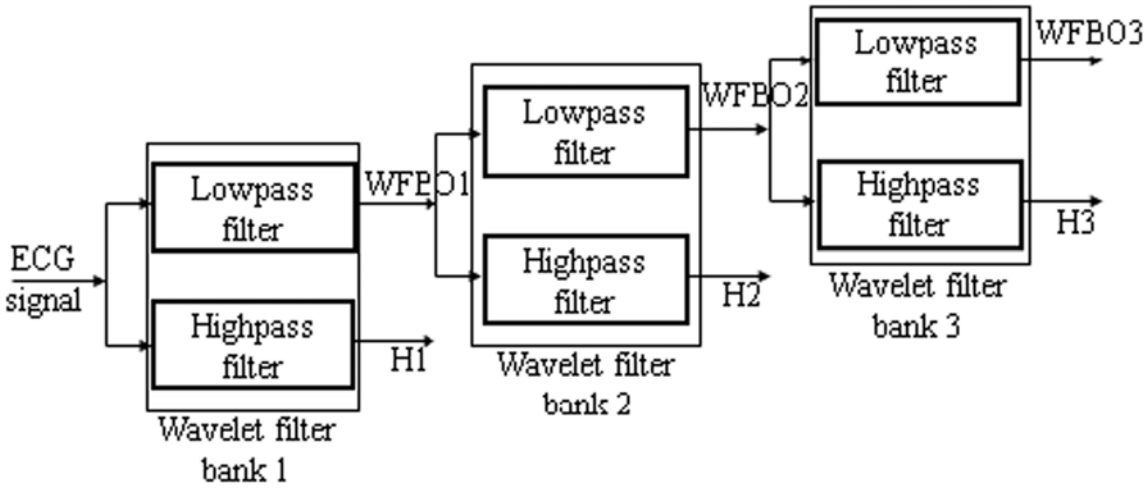


Fig. 1 Schematic representation of wavelet filter bank

ratio (CSNR), root-mean-square error (RMSE) (Das and Ari 2013, Kumar et al. 2019a, 2019b), maximum absolute error (MAE), percentage root-mean-square difference (PRD), and quality score (QS) which are defined in Eqs. (1) to (6).

$$CR = \frac{\text{Data size of original ECG signal}}{\text{Data size of compressed ECG signal}} \tag{1}$$

$$CSNR = 10\log\left[\frac{\sum_M \left(x_p - \overline{X}\right)^2}{\sum_M \left(x_p - x_r\right)^2}\right] \tag{2}$$

$$RMSE = \sqrt{\sum_M \frac{x_d - x_p}{M}} \tag{3}$$

$$MAE = \max\left(x_p - x_r\right) \tag{4}$$

$$PRD = 100\sqrt{\left\{\frac{\sum_M \left[x_p - x_d\right]^2}{\sum_M \left[x_p\right]^2}\right.} \tag{5}$$

$$QS = \frac{CR}{PRD} \tag{6}$$

Here, x_p is the original ECG signal, x_d is the denoised ECG signal, x_r is the reconstructed ECG signal, M is the number of samples, and $\overline{X}$ is a mean value of x_r.

The proposed method is explained with the help of a pseudo code as described below:

Generating location map: Traverse the signal and set the *i*th element of the location map to '*1*' if the *i*th element of the signal is odd and '*0*' if the *i*th element of the signal is even.

Pseudo code for embedding watermark

Step 1: Load the ECG signal and watermark data.
Step 2 If the size of the watermark is greater than 10% of the length of the signal, then split the watermark into multiple parts such that each part is less than 10% of the length of the signal, else continue.
Step 3 Generate a location map where every odd-valued element of the signal is represented by '0' and even-valued elements are represented by '1'.
Step 4 Convert the watermark from a character array into a set of 8-bit binary numbers and append the bits to create a payload bit stream.
Step 5 Traverse through the payload,

If the value of the i^{th} payload element is 1
If the value of the i^{th} location map element is 1
Decrease the signal value by 1.
Else
Increase the signal value by 1.

Table 1 Characteristics of the ECG databases

ECG database	Number of recordings	Sampling frequency (Hz)	Duration (min)	Status
MITDB	48	360	30	Open
QTDB	105	250	15	Open
LTST	86	1260	120	Open
PTB	549	1000	1.55	Open
Fantasia	40	250	120	Open

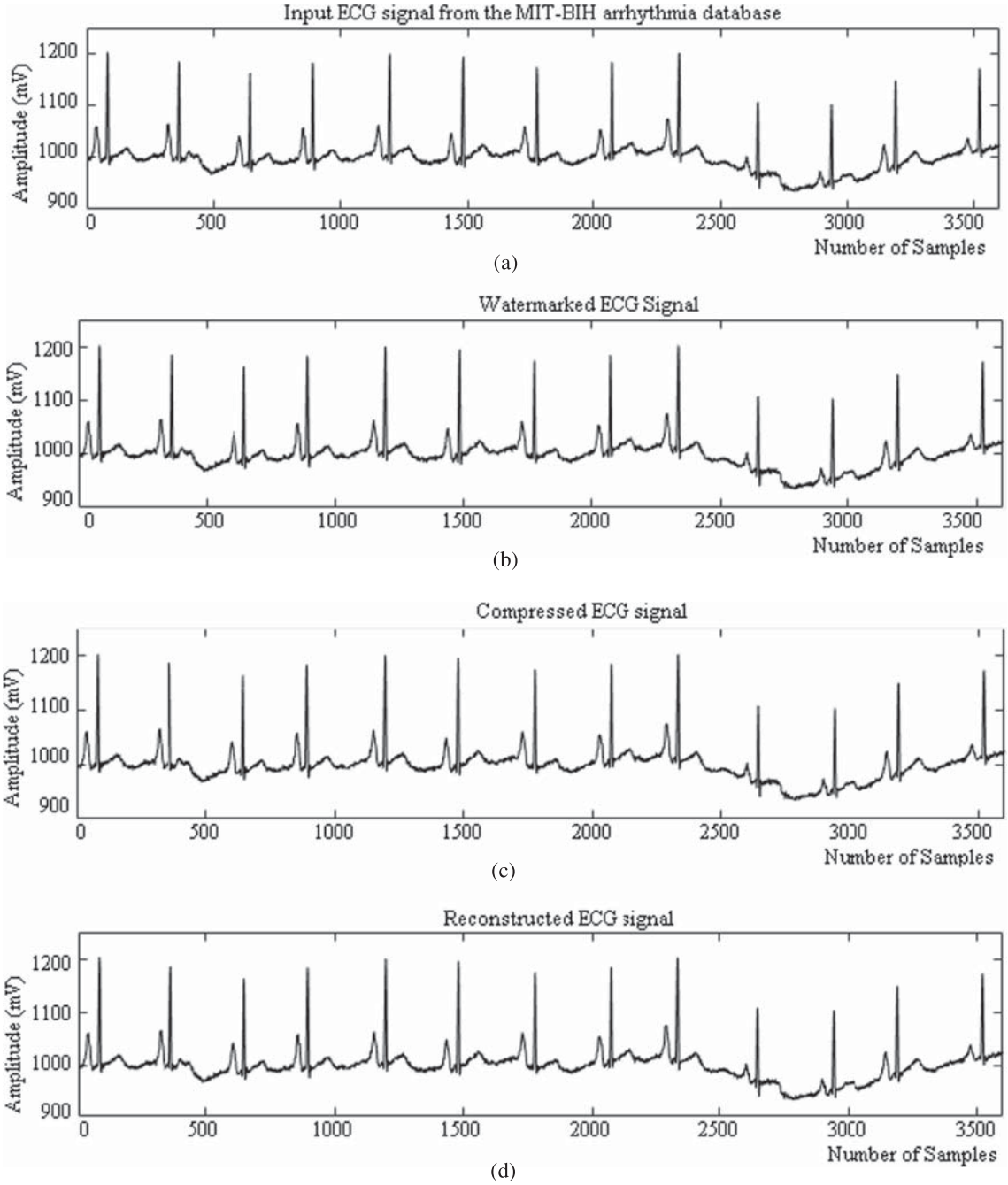


Fig. 2 Performance results of the proposed method. **a** Input ECG signal, **b** watermarked ECG signal, **c** compressed ECG signal, **d** reconstructed ECG signal

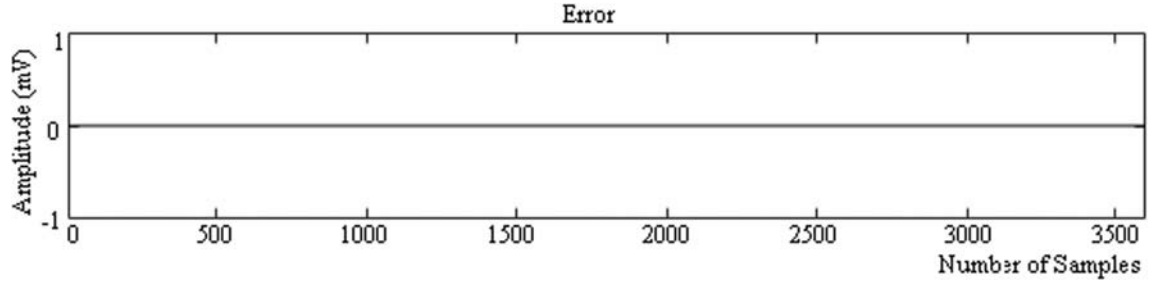


Fig. 3 Maximum absolute error of the proposed method

Table 2 Performance of the proposed technique using different ECG databases

ECG record number	Average PRD *(%)*	Average CR	Average CSNR (dB)	Compression time (ms)	Average MAE
108 m.mat	0	43.00	80.04	229.20	0
111 m.mat	0.001	39.76	74.31	248.66	0.028
121 m.mat	0	41.05	80.91	237.91	0.018
123 m.mat	0	39.47	72.20	261.31	0
205 m.mat	0.002	40.12	74.22	219.27	0
221 m.mat	0	44.59	76.02	229.20	0.027
233 m.mat	0.003	39.62	78.13	233.96	0

End if.
Else
Continue to the next element without modifying the signal.
End if.

Step 6: Compress the generated map using biorthogonal discrete wavelet transform. Perform 3rd-level biorthogonal wavelet transform decomposition which gives sub-bands LL3, LH3, HL3, and HH3 (approximation, Horizontal, Vertical, and diagonal, respectively).

Step 7 Convert the compressed map into binary and split it into groups of 8-bits.

Step 8 Convert the 8-bit groups into integers and append the numbers to the watermarked signal.

Step 9 Append the number of 8-bit groups created to the watermarked signal.

Pseudo code for extracting the watermark

Step 1 Load the watermarked signal.

Step 2 Extract the overhead.

Step 2.1 Extract and remove the last element of the watermarked signal (L_o).

Step 2.2 Extract and remove the last L_o elements of the watermarked signal (O_i).

Step 3 Convert every element in O_i into an 8-bit binary number and concatenate the numbers into a single vector.

Step 4 Decompress the binary vector using the biorthogonal discrete wavelet transform to obtain the encoded location map.

Step 5 Scan the signal after removing all overheads and generate a location map.

Step 6 Compute the exclusive 'OR' of the extracted and generated location maps.

Step 7 Split the resulting bit map into 8-bit groups.

Step 8 Convert the 8-bit groups into integers and append the integers into a single vector.

Step 9 Convert the integer vector into characters to form a character array containing the watermark.

Performance results

In this section, the proposed method is implemented and simulated using Matlab® R2019a. ECG signals from different databases, like MIT-BIH database, PTB diagnostic ECG database, Fantasia database, LTST database, and QT database, are used to evaluate the performance of the proposed design (Kumar et al. 2018, 2019a, 2019b). All the ECG records of each dataset described earlier have been utilized to evaluate the performance of the proposed method.

Simulation results in Fig. 2 and Fig. 3 represent the output performance of the proposed ECG signal watermarking and compression scheme. Record number 102 from the MIT-BIH

Table 3 Performance of the proposed technique using different ECG databases

ECG database	Number of beats	Average PRD (%)	Average CR	Average CSNR (dB)	Compression time (ms)	Average MAE
MIT-BIH	109,949	0	35.37	78.70	289.01	0
PTB	25,769	0	2954	72.65	211.52	0
Fantasia	19,445	0.002	30.27	78.38	198.93	0.032
STDB	10,173	0.006	32.89	70.32	153.29	0.002
QTDB	35,682	0.003	29.02	79.91	196.24	0
Average Value	201,018	0.002	31.418	76.01	209.79	0.0068

Table 4 Comparison of various performance evaluation parameters of the proposed method with the existing literature

Reference	Number of beats	Average PRD (%)	Average CR	Average CSNR (dB)	Compression time (ms)	Average MAE
(Tseng et al. 2014)	201,018	0.013	18.76	36.54	389.77	0.113
(Sanivarapu et al. 2020)	201,018	0.940	-	66.43	-	0.158
(Kaur et al. 2010)	201,018	0.643	-	56.88	-	0.232
(Ibaida et al. 2011)	201,018	0.106	30.08	58.62	353.29	0.082
(Mahmoud and Jusak 2018)	201,018	0.093	30.52	70.01	296.24	0.019
Proposed Work	201,018	0.002	31.41	76.01	209.79	0.0068

arrhythmia database, when used as an input ECG signal, is shown in Fig. 2a. After watermarking, and compression, the ECG signal 102 from the MIT-BIH arrhythmia database is as shown in Fig. 2 b and c, respectively. The reconstructed ECG signal is depicted in Fig. 2d. It is clear from Fig. 2d that the reconstructed ECG signal is a replica of the input ECG signal which states that the proposed scheme is a lossless data compression scheme. Further, to prove the lossless nature of the proposed scheme, error between the input ECG signal and reconstructed ECG signal is calculated and is shown in Fig. 3. Figure 3 shows zero error between the input and reconstructed ECG signals.

Table 2 shows the performance of the proposed technique using few ECG signals from MIT-BIH arrhythmia database. Table 3 shows the performance of the proposed technique using different ECG databases. The proposed technique achieves a higher average CR, higher average CSNR, lower average MAE, lower average PRD, and lower average compression time of 31.418, 76.01, 0.0068, *0.002*, and 209.79 *ms*, respectively. The average compression time of *209.79 ms* for the proposed method is much smaller compared to the existing literature which has an average compression time of *357.11 ms*.

Discussion

Comparison of the various performance evaluation parameters (i.e., average CR, average PRD, and average CSNR) of the proposed method with the existing literature is shown in Table 4.

From Table 4, it has been seen that a higher average CR, average CSNR, and the lowest average PRD values have been achieved by the proposed method when compared to the existing literature. SNR is a way of describing how much information a signal contains compared to the noise present in a signal. The higher the SNR, the more significant is the difference between the useful signal and unwanted noise. A higher SNR provides information about the quantity of quality information compared to the irrelevant information. Besides that, the lower PRD value represents the lossless nature of the proposed method. Many ECG compression algorithms described in the literature show a high compression performance with a small number of QRS complexes. The main drawback of these algorithms is that the high compression performance of the algorithm is verified using only a few ECG samples. Also, the performance of ECG compression algorithms presented in the literature is not assessed for some statistical parameters like computational complexity, space-saving, parameter choice, and compression gain. Some of the ECG compression algorithms presented in the literature are not verified using ECG data from the standard database. However, the effectiveness of the proposed algorithm in terms of the industrial implementation needs to be checked, which is the major drawback.

Conclusion

The present study successfully concludes with designing and execution of wavelet analysis of ECG watermark and compression algorithm. This technology is useful in safeguarding ECG transmission data, thus necessarily leading to enhanced security of ECG data with a reduction in transmission volume. This helps in augmenting the shape of ECG. To achieve the security assurance of the embedded watermark in the signal wavelet decomposition, lowest frequency coefficients have been used to embed the watermark. For a uniform comparison, MIT-BIH arrhythmia database is utilized. The proposed method provides higher CR, higher CSNR, lower RMSE, PRD, and compression time when compared to the existing literature.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Ali M, Ahn CW. An optimized watermarking technique based on self-adaptive DE in DWT–SVD transform domain. Signal Process. 2014;94:545–56.

Ansari IA, Pant M, Ahn CW. Robust and false positive free watermarking in IWT domain using SVD and ABC. Eng Appl Artif Intell. 2016;49:114–25.

Bender W, Gruhl D, Morimoto N, Lu A. Techniques for data hiding. IBM Syst J. 1996;35(3.4):313–36.

Chen ST, Huang HN, Chen CJ, Wu GD. Energy-proportion based scheme for audio watermarking. IET Signal Process. 2010;4(5): 576–87.

Chou HH, Chen YJ, Shiau YC, Kuo TS. A high performance compression algorithm for ECG with irregular periods. In: IEEE International Workshop on Biomedical Circuits and Systems, 2004. 2004; (pp. S2–4). IEEE.

Das MK, Ari S. Analysis of ECG signal denoising method based on S-transform. Irbm. 2013;34(6):362–70.

Deepu CJ, Lian Y. A joint QRS detection and data compression scheme for wearable sensors. IEEE Trans Biomed Eng. 2014;62(1):165–75.

Deepu CJ, Heng CH, Lian Y. A hybrid data compression scheme for power reduction in wireless sensors for IoT. IEEE Trans Biomed Circuits Syst. 2016a;11(2):245–54.

Deepu CJ, Zhang X, Heng CH, Lian Y. A 3-lead ECG-on-chip with QRS detection and lossless compression for wireless sensors. IEEE Trans Circuits Syst II: Express Briefs. 2016b;63(12):1151–5.

Elgendi M, Mohamed A, Ward R. Efficient ECG compression and QRS detection for e-health applications. Sci Rep. 2017;7(1):1–6.

Ganic E, Eskicioglu AM. Robust DWT-SVD domain image watermarking: embedding data in all frequencies. InProceedings of the 2004 Workshop on Multimedia and Security 2004 Sep 20 (pp. 166–174).

Ibaida A, Khalil I. Wavelet-based ECG steganography for protecting patient confidential information in point-of-care systems. IEEE Trans Biomed Eng. 2013;60(12):3322–30.

Ibaida A, Khalil I, Van Schyndel R. A low complexity high capacity ECG signal watermark for wearable sensor-net health monitoring system. In2011 Computing in Cardiology 2011 Sep 18 (pp. 393–396). IEEE.

Kaur S, Singhal R, Farooq O, Ahuja BS. Digital watermarking of ecg data for secure wireless communication. In 2010 International Conference on Recent Trends in Information, Telecommunication and Computing 2010 Mar 12 (pp. 140–144). IEEE.

Kim BS, Yoo SK, Lee MH. Wavelet-based low-delay ECG compression algorithm for continuous ECG transmission. IEEE Trans Inf Technol Biomed. 2006;10(1):77–83.

Kumar R, Kumar A, Pandey RK. Beta wavelet based ECG signal compression using lossless encoding with modified thresholding. Comput Electr Eng. 2013;39(1):130–40.

Kumar A, Komaragiri R, Kumar M. From pacemaker to wearable: techniques for ECG detection systems. J Med Syst. 2018;42(2):34.

Kumar A, Komaragiri R, Kumar M. Time–frequency localization using three-tap biorthogonal wavelet filter bank for electrocardiogram compressions. Biomed Eng Lett. 2019a;9(3):407–11.

Kumar A, Komaragiri R, Kumar M. Design of efficient fractional operator for ECG signal detection in implantable cardiac pacemaker systems. Int J Circuit Theory Appl. 2019b;47(9):1459–76.

Lai CC, Tsai CC. Digital image watermarking using discrete wavelet transform and singular value decomposition. IEEE Trans Instrum Meas. 2010;59(11):3060–3.

Liu R, Tan T. An SVD-based watermarking scheme for protecting rightful ownership. IEEE Trans Multimedia. 2002;4(1):121–8.

Mahmoud SS, Jusak J. Wavelets in ECG security application. Wavelet Theory Appl. 2018;43.

Manikandan MS, Dandapat S. Wavelet-based electrocardiogram signal compression methods and their performances: a prospective review. Biomed Signal Process Control. 2014;14:73–107.

Mishra A, Agarwal C, Sharma A, Bedi P. Optimized gray-scale image watermarking using DWT–SVD and firefly algorithm. Expert Syst Appl. 2014;41(17):7858–67.

Mohammed GN, Yasin A, Zeki AM. Digital image watermarking, analysis of current methods. In 2012 International Conference on Advanced Computer Science Applications and Technologies (ACSAT) 2012 Nov 26 (pp. 324–329). IEEE.

Muharemagic E, Furht B. Survey of watermarking techniques and applications. Multimed Watermark Tech Appl. 2006;3(91):130.

Mukherjee DP, Maitra S, Acton ST. Spatial domain digital watermarking of multimedia objects for buyer authentication. IEEE Trans Multimedia. 2004;6(1):1–5.

Patra D, Das MK, Pradhan S. Integration Of Fcm, Pca and neural networks for classification of Ecg arrhythmias. IAENG Int J Comput Sci. 2009;36(3).

Sanivarapu PV, Rajesh KN, Reddy NR, Reddy NC. Patient data hiding into ECG signal using watermarking in transform domain. Phys Eng Sci Med. 2020;43(1):213–26.

Swanson MD, Kobayashi M, Tewfik AH. Multimedia data-embedding and watermarking technologies. Proc IEEE. 1998;86(6):1064–87.

Tseng KK, He X, Kung WM, Chen ST, Liao M, Huang HN. Wavelet-based watermarking and compression for ECG signals with verification evaluation. Sensors. 2014;14(2):3721–36.

Zheng KM, Tang LW, Qian X. Watermarking technology for electrocardiogram signal certification. Comput Eng Appl. 2009;45(20):231–3.

Zigel Y, Cohen A, Katz A. ECG signal compression using analysis by synthesis coding. IEEE Trans Biomed Eng. 2000;47(10):1308–16.

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

AIP Conference Proceedings

A theoretical comparison between $CH_3NH_3PbI_3$ and $CH_3NH_3SnI_3$ based solar cells

Cite as: AIP Conference Proceedings **2265**, 030609 (2020); https://doi.org/10.1063/5.0017201
Published Online: 05 November 2020

Gagandeep, Mukhtiyar Singh, Ramesh Kumar, Vinamrita Singh, and Sunita Srivastava

View Online Export Citation

Meet the Next Generation of Quantum Analyzers
And Join the Launch Event on November 17th

Register now
Zurich Instruments

AIP Publishing

© 2020 Author(s).

A Theoretical Comparison between CH3NH3PbI3 and CH3NH3SnI3 based Solar Cells

Gagandeep[1, a)], Mukhtiyar Singh[2], Ramesh Kumar[1, b)],Vinamrita Singh[3] and Sunita Srivastava[1,4)]

[1]*Department of Physics, Guru Jambheshwar University of Science & Technology, Hisar, India*
[2]*Department of Applied Physics, Delhi Technological University, Delhi, India*
[3]*Department of Applied Sciences and Humanities, Ambedkar Institute of Advanced Communication Technologies & Research, Delhi, India*
[4]*Department of Physics, Punjab University Chandigarh, India*

Corresponding author: a) gagankundu22@gmail.com
b) rameshkumar.bibiyan@gmail.com

Abstract. The perovskite materials based photovoltaic cells play a crucial role in the conversion of sunlight into electric energy. The commonly used perovskite materials are $CH_3NH_3PbI_3$ and $CH_3NH_3SnI_3$. In the present manuscript, a theoretical comparison is done by simulating the structures ITO/TiO_2/$CH_3NH_3PbI_3$/p-Gr/Ag and ITO/TiO_2/$CH_3NH_3SnI_3$/p-Gr/Ag structures in AFORS-HET software under the illumination condition of AM 1.5G. Here, TiO_2 and p-Graphene (p-Gr) play the role of electron transporting material and hole transporting material respectively with ITO and Ag as front and back contacts. The thickness of layers is directly related to the cost of the cells. Therefore, we optimized the $CH_3NH_3PbI_3$ and $CH_3NH_3SnI_3$ layers thickness to observe the performance of the device. The power conversion efficiency of 10.62% is obtained for $CH_3NH_3SnI_3$based solar cell at optimized thickness of 800 nm as compared to 9.74% efficiency for $CH_3NH_3PbI_3$based solar cell at an optimized thickness of 700 nm. The textured nature of front surface also enhances the efficiency 10.56% and 12.38% for $CH_3NH_3PbI_3$ and $CH_3NH_3SnI_3$ respectively.

INTRODUCTION

The largely employed perovskite material is methyl ammonium lead halide perovskite i.e. $CH_3NH_3PbI_3$ for the perovskite solar cells (PSC) [1, 2]. The PSC possess a heterojunction architecture with three layers [3] i.e ETM (electron transporting material), perovskite absorbing layer and HTM (hole transporting material) with front and back contacts as shown in fig. 1a. The commonly employed ETM and HTM are TiO_2 and organic material spiro-OMeTAD respectively [3]. The synthesis of organic material spiro-OMeTAD is expensive process and enhances the cost of device. Therefore, the researchers are developing different types of HTM in PSCs [4-6].

Graphene can be employed as a novel material in PSCs due to its appropriate electrical and optical characteristics [7]. The graphene can be inserted as ETM, HTM or transfer electrode in PSCs because its work function as well as energy band gap can be tuned under the influence of doping [7-9].

The toxicity of lead in PSC is a major obstacle in its development for commercial application. Therefore, tin based PSC (TPSC) [10] is being explored as a substitute of lead based PSC in future. Keeping this in mind, the

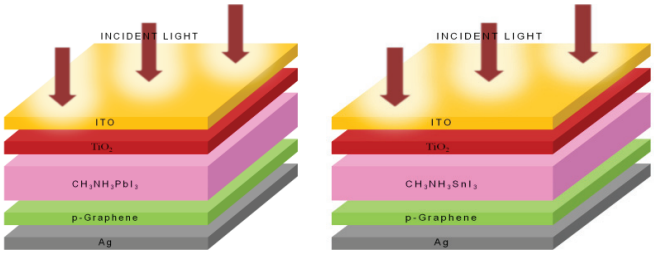
DAE Solid State Physics Symposium 2019
AIP Conf. Proc. 2265, 030609-1–030609-4; https://doi.org/10.1063/5.0017201
Published by AIP Publishing. 978-0-7354-2025-0/$30.00

present work compares the performance of lead-based and lead-free perovskite solar cells in order to develop environmentally friendly technology.

SIMULATION DETAILS

The structures ITO/TiO_2/$CH_3NH_3PbI_3$/p-Gr/Ag and ITO/TiO_2/$CH_3NH_3SnI_3$/p-Gr/Ag are simulated in Automat for Simulation of Heterostructures (AFORS-HET) software [11-13] as shown in fig. 1a and 1b respectively. Both the structures are n-i-p type i.e. light is falling from n-side. The electron-hole pairs are generated in the active layers. The generated electrons are collected by the ETM layer due to the built in voltage developed at the interface between the TiO_2 and perovskite layers. Similarly, generated holes are collected by the HTM layer i.e. p-graphene material.

To check the performance of both the devices, we optimized the active layer thickness. The $CH_3NH_3PbI_3$ and $CH_3NH_3SnI_3$layer thickness are varied in the range of 50-800 nm and 50-700 nm respectively. The variations in open circuit voltage (Voc), short circuit current density (J_{sc}), fill factor (FF) and power conversion efficiency (PCE) of the cell with absorbing layer thickness are observed. The impact of textured front surface is also studied on both the configurations. The detailed device parameters used in the simulation are taken from different theoretical and experimental studies [3, 14-16] and are mentioned in table 1.



(a) (b)

FIGURE.1 n-i-p structure of perovskite solar cell using (a) $CH_3NH_3PbI_3$ and (b) $CH_3NH_3SnI_3$ absorber layer

TABLE 1. Material parameter of each layer

Parameters	TiO_2	$CH_3NH_3PbI_3$	$CH_3NH_3SnI_3$	Graphene
Thickness (nm)	90	50-800	50-700	0.334
Dielectric constant	100	25	6.5	7.1
Electron affinity (eV)	4	3.9	4.17	4.7
Energy Band gap (eV)	3.2	1.5	1.3	0.5
Optical energy band gap (eV)	3.2	1.5	1.3	0.5
Valence band density (cm^{-3})	1×10^{21}	2.5×10^{20}	1×10^{18}	3×10^{19}
Conduction band density (cm^{-3})	2×10^{20}	2.5×10^{20}	1×10^{19}	3×10^{19}
Hole mobility ($cm^2\ V^{-1}\ s^{-1}$)	0.006	50	1.6	1×10^{5}
Electron mobility ($cm^2\ V^{-1}\ s^{-1}$)	0.006	50	1.6	10
Thermal velocity of electron(cm/s)	1×10^{6}	2.4×10^{18}	1×10^{6}	1×10^{7}
Thermal velocity of hole(cm/s)	4×10^{6}	2.4×10^{18}	1×10^{6}	1×10^{7}
Doping concentration(cm^{-3})	5×10^{19}	1×10^{17}	3.2×10^{15}	1×10^{18}
Layer density(g cm^{-3})	2.328	2.328	2.328	2.328

RESULTS AND DISCUSSIONS

With the help of layer parameters listed in table 1, we studied the effect of absorbing layer thickness of both the perovskite on the performance parameters i.e, J_{sc}, V_{oc}, FF and PCE of the solar cell. Fig.2 gives the effect of thickness of $CH_3NH_3PbI_3$ (Pb) and $CH_3NH_3SnI_3$ (Sn) perovskite layers on the different parameters. It is observed that J_{sc} (fig. 2(a)) shows increment in its value with increase in thickness in both the cases ($CH_3NH_3PbI_3$ and

$CH_3NH_3SnI_3$) to a saturation value of 16.99 mA/cm^2 and 19.90 mA/cm^2, respectively. This enhancement results due to comparatively thicker active layer which absorbs more photons. Greater photon absorption creates large electron-hole pair generation in active layer. The better J_{sc} behavior is demonstrated by $CH_3NH_3SnI_3$ based structure.

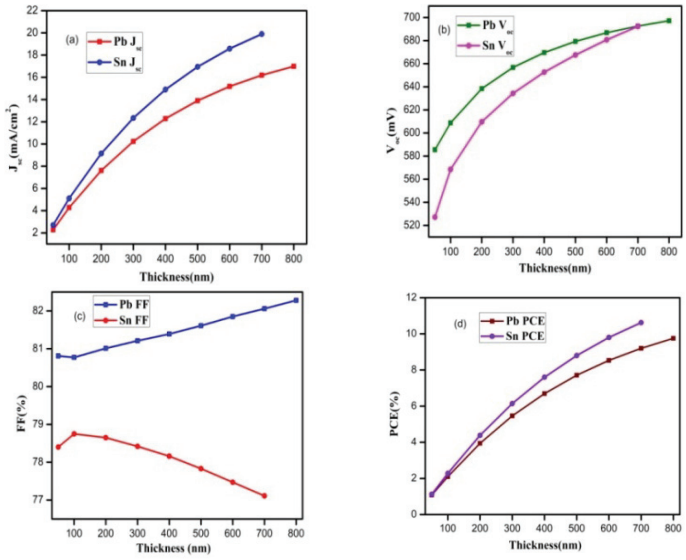


FIGURE 2. Effect of layer thickness of $CH_3NH_3PbI_3$ and $CH_3NH_3SnI_3$ on (a) J_{sc}, (b) V_{oc}, (c) FF and (d) PCE of the perovskite solar cell

The value of V_{oc} varies from 585.5 mV to 697.2 mV and 527.1 mV to 692.2 mV for $CH_3NH_3PbI_3$ and $CH_3NH_3SnI_3$ perovskite respectively and the results are given in fig. 2 (b). However, the FF value slightly reduces from 78.40% to 77.11% in $CH_3NH_3SnI_3$ layer within the thickness range of 50-700 nm. This may probably due to internal recombination in perovskite absorber layer. But in $CH_3NH_3PbI_3$, the value of FF increases from 80.81% to 82.28% with thickness ranging from 50-800 nm and the results are given in fig. 2(c). Fig. 2(d) gives the variation of PCE with thickness for both the perovskite solar cells and a considerable increase in the value of efficiency from 1.07% to 9.74% and 1.12% to 10.62% is noted for $CH_3NH_3PbI_3$ and $CH_3NH_3SnI_3$ respectively. The performance for $CH_3NH_3SnI_3$ based device is better as compared to $CH_3NH_3PbI_3$based device.

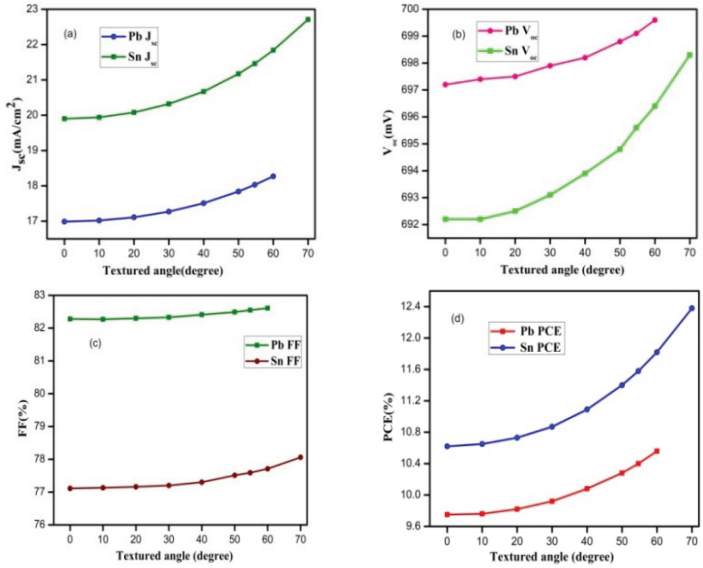


FIGURE 3. Effect of angle of incidence of light on $CH_3NH_3PbI_3$ and $CH_3NH_3SnI_3$ on (a) J_{sc}, (b) V_{oc}, (c) FF and (d) PCE of the perovskite solar cell

The performance of solar cell is also influenced by the nature of the front contact surface because it may be plane or textured. The textured front surface absorbs more photons as compared to the plane front surface. This is due to the fact that when incident light falls on the plane surface of the cell, it may bounce back but if the surface is textured the light undergoes multiple reflections. Hence, in effect, large area is used for the absorption of incident light improving the cell performance. We varied the angle of incidence from 0° to 60° and 0° to 70° for $CH_3NH_3PbI_3$ and $CH_3NH_3SnI_3$ respectively and the results are shown in fig. 3. It is observed that all the cell parameters improve with incidence angle giving J_{sc} values 16.99 mA/cm^2 to 18.27 mA/cm^2 and 19.90 mA/cm^2 to 22.71 mA/cm^2, V_{oc} = 697.2 mV to 699.6 mV and 692.2 mV to 698.3 mV, FF = 82.28% to 82.61% and 77.11% to 78.06% and PCE varying from 9.74% to 10.56% and 10.62% to 12.38% for $CH_3NH_3PbI_3$ and $CH_3NH_3SnI_3$ perovskite solar cell.

CONCLUSIONS

The present manuscript is devoted to the comparison between the $CH_3NH_3PbI_3$ and $CH_3NH_3SnI_3$based solar cells. The behavior of both the cells is discussed under the influence of perovskite layer thickness. We also replaced the complex material spiro-OMeTAD by graphene as hole transporting material. The PCE for both $CH_3NH_3PbI_3$ and $CH_3NH_3SnI_3$ solar cells are 9.74% and 10.62% at optimized active layer thickness of 800 nm and 700 nm respectively. The textured nature of front surface also improved the efficiency to 10.56% and 12.38% for $CH_3NH_3PbI_3$ and $CH_3NH_3SnI_3$solar cells respectively. Therefore, based on the values of performance parameters for both $CH_3NH_3PbI_3$ and $CH_3NH_3SnI_3$ solar cells, the tin based solar cells has found to exhibit better performance for the above suggested structure.

ACKNOWLEDGMENT

One of the authors (**Ramesh Kumar**) is grateful to the **University Grants Commission, INDIA** for providing the financial support through START-UP GRANT under FRPS (No. F.30-364/2017 (BSR), Dated 15/12/2017). The authors gratefully acknowledge the AFORS-HET software developer for providing free access of the software.

REFERENCES

1. A. Kojima, K. Teshima, Y. Shirai and T. Miyasaka J. Am. Chem. Soc. **131**, 6050 (2009).
2. N. J. Jeon et. al., Nat. Energy **3**, 682 (2018).
3. H. J. Du, W. C. Wang and Y. F. Gu, Chin. Phys. B **26**, 028803 (2017).
4. K. Mahmood, B. S. Swain and H. S. Jung, Nanoscale.**6**, 9127 (2014).
5. S. Lu et. al., Chem. Commun.**50**, 6931(2014).
6. P. Qin et. al., Nat. Commun.**5**, 3834 (2014)
7. E. L. Lim et.al., Nano-Micro Lett. **10: 27** (2018).
8. T. Mahmoudi, Y. Wang and Y. B. Hahn, Nano Energy.**47**, 51 (2018).
9. M. F. Bhopal et.al., J. Mater. Chem. C. **5**, 10701 (2017).
10. MH. Kumar et.al., Adv. Mater. **26**, 7122 (2014).
11. R.Varache et. al., Sol. Energ. Mat. Sol. C.**141**, 14 (2015).
12. R. Stangl, J. Haschke and C. Leendertz,Solar Energy, Intech (R.D. Rugescu (Ed.), Croatia, 2009), pp. 432.
13. R.Stangl, M. Kriegel and M. Schmidt, Proceeding of the WCPEC-4, Fourth World Conference on Photovoltaic Energy Conversion (Hawaii, USA), PP. 1350 (2006).
14. S.M. Iftiquar, J. S. Kim and J. Yi, Optik**148**, 54(2017).
15. K. Patel and P.K. Tyagi. Carbon**116**, 744 (2017).
16. Gagandeep, M. Singh and R. Kumar, AIP Conference Proceedings, **2115**, 030548 (2019).

The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1759-0833.htm

Assessing brand love, brand sacredness and brand fidelity towards halal brands

Richa Joshi
Department of Management Studies, National Institute of Technology, Hamirpur, India, and
Prerna Garg
Department of Management Studies, Jaipuria Institute of Management, Ghaziabad, India and Bharati Vidyapeeth Deemed University, Pune, India

Received 12 April 2020
Revised 21 October 2020
Accepted 1 November 2020

Abstract

Purpose – The purpose of this study is to examine the role of contemporary consumer-brand relationships in predicting brand sacredness, brand fidelity and propensity to spread word of mouth (WOM) in the context of "halal" cosmetic brands.

Design/methodology/approach – The proposed research framework has been empirically tested by collecting responses from 403 Muslim respondents of diverse demographics. Structural equation modelling has been used for exploring the underlying relationships between emotions associated with halal brands.

Findings – All the proposed hypotheses were positively significant, thus confirming that brand love is significantly affected by brand trust, brand image, self-congruence and brand experience. Further, brand love acts as a significant determinant in shaping brand sacredness, brand fidelity and WOM.

Research limitations/implications – The study has made a significant contribution by exploring the intensity of brand love and its effect on relationship-maintenance triad in halal cosmetic brands in India.

Practical implications – Marketers must understand the emotional side of brands to create synergy in their marketing efforts. Moreover, in the case of religious or faith-based brands, attainment of the highest order of brand love can play a revolutionary role for long-term growth.

Originality/value – In the context of halal cosmetic brands in India, this study has offered a new perspective by extending the literature on consumer-brand relationships. The use of brand fidelity and brand sacredness has enhanced the effectiveness of emotions with which consumers have been looking at brands for ages.

Keywords Brand image, Word of mouth, Brand trust, Brand fidelity, Brand sacredness, Self-congruence

Paper type Research paper

1. Introduction

Literature in marketing has undergone several changes in the past few decades. Initial phases of the 1970s witnessed a thrust on examining the antecedents and outcomes of customer satisfaction to a great extent. However, it was in the late 1980s and 1990s that branding emerged as a novel area of research with major focus captured by constructs such as brand extension and brand equity (Joshi and Yadav, 2016). During the early 2000s, the emotional aspects of branding such as brand love (Carroll and Ahuvia, 2006), brand attachment (Park *et al.*, 2008; Park *et al.*, 2007), brand trust (Delgado-Ballester and Munuera-Alemán, 2001; Ha, 2004) and brand resonance (Rindfleisch *et al.*, 2006; Moore and Wurster, 2007) started penetrating. Since then, researchers across the globe are exploring consumer-brand

Journal of Islamic Marketing
© Emerald Publishing Limited
1759-0833
DOI 10.1108/JIMA-04-2020-0104

relationships, which involve emotional connections. Theories and concepts from psychology have been borrowed to explain the consumer's behaviour in different situations (Sarkar and Sreejesh, 2014). Consumers also shifted their focus from generic products to branded ones; therefore, it became necessary to understand their emotions. Moreover, researchers and practitioners understood that present-day consumers are not limited to shopping malls, shopping places and showrooms; they are omnipresent; they are everywhere online and offline. Consumers these days interact with brands (Grace *et al.*, 2018) and consider brands to express themselves. Hence, brand managers' role becomes more prominent in communicating and positioning brands in a way that they connect themselves to the customer. Consumers associate brands with their identities and use them to position themselves in the society. Moreover, a brand adds meaning to a product, and consumers worldwide look forward to brands that adhere to their cultures and values (Borzooei and Asgari, 2013). Therefore, "halal brand", as a religious brand holds key importance in examining the consumption patterns of Muslim consumers. Tracing the shift in decision-making by Muslim consumers over the decades, marketers need to focus on "halal branding" and communicate attributes beyond quality and hygiene. Academicians and practitioners all across the world have also aligned their focus in understanding Islamic marketing and branding as a new and separate discipline (Wilson and Liu, 2011).

The word "halal" has an "Islamic" origin, and it refers to everything, which is permissible in Islamic religion (Wilson, 2014; Garg and Joshi, 2018). "Halal" branding has emerged as a differentiation strategy to position products (Garg and Joshi, 2018) to a particular segment of the population. It conveys to the customer that certain guidelines were followed during the production of such products (Wilson and Liu, 2010), hence it has offered a reassurance (Wilson and Grant, 2013) and reduced the propensity of jeopardizing the cultural norms. Interestingly, studying "halal" in a country such as India paves an opportunity for the researchers to understand and explore the opinions of over 180 million of Muslim population (Hassan and Sengupta, 2019). However, existing literature shows that the studies on "faith-based marketing" especially "Islamic marketing/branding" are a recent development in emerging economies. Many of these studies are limited to countries such as Malaysia and Indonesia (Ahmad, 2018; Battour and Ismail, 2016). Therefore, there is a need to explore in the context of other Asian countries. Additionally, halal branding in personal care and cosmetics is also a less-discussed area. However, researchers such as Garg and Joshi (2018) have explored the purchase intention of customers towards "halal" cosmetics in India. Their study was focused on the attitude of the consumer whereas, the surge in the studies based on consumer-brand relationships gathers a need to explore the emotional side of the consumer via novel constructs such as brand love, brand sacredness and brand fidelity. The present study, therefore, advances the literature on "halal" branding by integrating the emotional side of customer while consuming the brands.

The development of relationships is vital for a brand, but the continuance of a relationship is equally important. Hence, the central question is whether the consumer is interested in maintaining their relationship with the brand? The answer to this lies in observing the patterns of consumer emotions i.e.; they vary from jealousy to love and from hate to passion. The collection of all these emotions in the context of branding surfaces a way for the researchers to explore the underlying dimensions of consumer-brand relationships. "Sacredness is considered as the highest form of brand love" attaining such a high position in consumer's thought process requires utmost effort. However, what comes next is equally significant, is consumer exhibiting loyalty only? Or he moves to the next level from loyalty, i.e. fidelity. Brand fidelity explains the intention of preserving the relation through accommodation of the brand, forgiveness, derogation of alternatives, willingness to

sacrifice and taking personal ownership of the brand (Grace *et al.*, 2018). Thus, in the current study, authors have proposed a framework for empirically evaluating complex consumer emotions by taking "halal" cosmetics as a stimulus. As per the authors' knowledge, this is the first study that tests brand fidelity and sacredness in the context of "halal" brands and has attempted to propose a relationship maintenance triad for brands.

2. Theoretical grounding

The theoretical background of the study is based on the constructs that define consumer emotions. The proposed framework of the study has two parts. The first part includes the antecedents of brand love i.e. brand trust, brand experience, brand image and self-congruence whereas, the second part consists of brand sacredness, brand fidelity and positive word of mouth (WOM).

2.1 Antecedents of brand love

2.1.1 Brand trust. Trust is defined as "the extent to which a consumer believes that a certain brand he or she has confidence in satisfies his or her desire" (Carroll and Ahuvia, 2006). The concept has been widely published in branding literature in connection to the trust consumers feel towards the brand. Brand trust is termed as "the willingness of the average consumer to rely on the brand's ability to perform its stated function" (Chaudhuri and Holbrook, 2001). The concept of brand trust is highly prominent in situations of uncertainty hence, it is highly relevant for cosmetic brands. Usage of cosmetics puts a customer in a vulnerable situation hence, a trusted brand decreases uncertainty and increases the confidence of the customers (Chiu *et al.*, 2010; Pavlou *et al.*, 2007). Brand trust is a driver of brand loyalty and loyalty is derived as an outcome of a trusted relationship with the brand (Chaudhuri and Holbrook, 2001). Also, sustained trust for the brand invokes feelings of love towards the brand. Literature in branding also confirms that brand trust influences some important constructs of marketing i.e. brand satisfaction, brand loyalty and brand love (Drennan *et al.*, 2015). Therefore, we hypothesize that:

H1. Brand trust positively influences brand love.

2.1.2 Brand experience. Brand experience was discussed for the first time in the work of Gilmore and Pine (1999). They stated that brand experience are "subjective, internal consumer responses (sensation, feelings and cognitions and behavioural responses), evoked by brand-related stimuli that are part of a brand's design and identity, packaging, communications and environments". However, brand experience does vary and can be positive or negative. Brand experience is conceptually different from brand attitude and brand involvement. It is altogether a different construct but is somewhat related to the other constructs such as brand satisfaction and brand loyalty. Four types of brand experiences were proposed by Brakus *et al.* (2009) i.e. affective, sensory, intelligent and behavioural. However, the present study has considered three dimensions of brand experience as suggested by Huang (2017) i.e. intellectual, behavioural and sensory. A favourable brand experience induces emotional bonding with the brand and develops intimacy, which further arouses brand passion. Hence, we propose the following hypothesis:

H2. Brand experience positively influences brand love.

2.1.3 Brand image. Brand image is defined as "a set of beliefs held about a particular brand" (Kotler, 2004). It is also stated that brand image consists of tangible and intangible perceptions related to the brand formed by affection, cognition and evaluation process of the consumer (Song *et al.*, 2019; Shabbir *et al.*, 2017). Therefore, brand image is crucial for

consumer decision-making and can strengthen the competitive position for a brand. It is formed by the consumer impression of a particular brand. However, it is affected by several factors such as publicity, promotion, reviews of the brand or perceived quality of the brand. A positive brand image inspires a consumer to have a strong emotional bond with the brand and enable them to love the brand. The similar observation has been reported in several studies such as Islam and Rahman (2016), Ismail and Spinelli (2012) and Al-Haddad (2019). Brand image also becomes highly relevant in the case of cosmetic brands, as it is a high involvement category for the consumer. Based on the previous studies, we, therefore, hypothesize that:

H3. Brand image positively influences brand love.

2.1.4 Self-congruence. Consumers make use of brands to express their self-concept (Ekinci *et al.*, 2013). Self-congruity theory postulates that consumers choose brands to protect their self-concept, and when the brand matches their actual self, they show more loyalty towards the brand (Wallace *et al.*, 2017). Self-congruence with a brand is defined as "the coherence/similarity in characteristics of an individual with a brand". In the context of "halal" cosmetics, the customers who consider them to be religious/follow Islamic religion would prefer to purchase "halal" branded cosmetics, as it is expected to enhance their self-concept. The idea is congruent with their actual self. Additionally, this effect propels the individual to experience an emotional bond with the brand. The way an individual is attracted to other similar individuals and creates social/emotional bonds; in a similar manner consumers are inclined to brands that are deemed to be self-congruent (Rauschnabel and Ahuvia, 2014). Also, they experience love/passion and attachment towards these self-congruent brands. Hence, we can infer that self-congruence with the brand affects brand love; the same has been proposed and empirically verified by Wallace *et al.* (2017). Hence, we propose the following hypothesis:

H4. Self-congruence positively influences brand love.

2.2 Relationship maintenance triad

The current study has proposed three distinct relationships, i.e. the connection among brand sacredness, brand fidelity and positive WOM. This triad plays a significant role in the continuance of the relationship with the brand. Relationship maintenance is equally crucial in comparison to the development of the relationship. Brand love ensures that the relationship is developed, and individuals who are deeply committed tend to forgive and accommodate the brand (Rusbult *et al.*, 2012) to continue the relationship with the brand. Therefore, brand sacredness i.e. the highest form of brand love must influence brand fidelity, as a devoted consumer is faithful towards the brand. Such consumer accommodates and forgives the brand even in the situations of sudden failure of the brand.

2.2.1 Brand love. For decades academicians have been studying the like and dislike of customers towards the brands. However, in the past decade, there is a shift in studies concerning branding relationships. Sternberg's (1986) triangular theory of love has encouraged researchers such as Shimp and Madden (1988) and Fournier (1998) to study love in the context of consumers. Carroll and Ahuvia (2006) explained brand love for the first time with the help of an empirical model stating antecedents and outcomes of brand love (Joshi and Garg, 2020). Recent studies on brand love have established brand love as an important construct in consumer-brand relationships. Brand love for a brand provides an advantage to a company over the competitors, as it is a strong feeling, which is far beyond

liking towards the brand. Several studies in the literature show love propels positive WOM (Carroll and Ahuvia, 2006), love ensures loyalty (Bergkvist and Bech-Larsen, 2010; Mody and Hanks, 2019), it gives rise to romantic jealousy (Sarkar and Sreejesh, 2014) and it also induces brand advocacy (Wallace *et al.*, 2017).

2.2.2 Brand sacredness. The origin of "sacred" can be traced back to the theological literature (Belk *et al.*, 1989). Sometimes few special characteristics are possessed by an object, which creates spiritual feelings in an individual such objects are considered "sacred". Religion provides belongingness to an individual (Fromm, 1956). Interestingly, these days' brands have been able to provide such a sense of belongingness to the customer. They empower the customer because brands serve as a self-expression tool (Sarkar *et al.*, 2015; Shachar *et al.*, 2011). Also, brands provide an extraordinary experience to the customer, which religion does provide. Such experience enables them to treat a brand as a sacred entity. Sarkar *et al.* (2015) extended this concept to consumer-brand relationships and called it brand sacralization, i.e. "the process by which brands are perceived as sacred entities". The work on brand sacredness/sacralization was strengthened based on brand love. "Brand love is the extent of deep emotional attachment a satisfied consumer has for a particular trade name" (Albert *et al.*, 2008). Such a strong feeling towards the brands gives rise to brand sacredness, as sacralization is the highest form of brand love (Batra *et al.*, 2012). Such a linkage between brand love and sacredness lead to the development of the following hypothesis:

H5. Brand love positively influences brand sacredness.

2.2.3 Brand fidelity. There is a surge in constructs related to the emotional aspects of branding. Brand fidelity is one of the newest constructs proposed in the literature of consumer-brand relationships. It is defined as "the consumer's faithfulness to a brand partner manifested through various behaviours (i.e. accommodation/forgiveness – performance and price) and cognitions (i.e. derogation of alternatives and cognitive interdependence) that maintain relationship stability and durability" (Grace *et al.*, 2018). Grace *et al.* (2018) proposed that brand fidelity is comprising two different manifestations i.e. cognitive manifestation and behavioural manifestation. The former further includes derogation of alternatives, positive illusions and cognitive interdependence, whereas the latter consists of accommodation/forgiveness and willingness to sacrifice. Hence, brand fidelity was observed as a multi-dimensional construct. However, further dimension reduction by Grace *et al.* (2020) resulted in four distinct factors, which included accommodation/forgiveness based on price, accommodation/forgiveness based on performance, derogation of alternatives and cognitive interdependence. For the present study, the authors have considered one item each for accommodation/forgiveness based on price, derogation of alternatives and cognitive interdependence. Devotion towards the brand leads to faithfulness for the brand; however, achieving the highest form of brand love is difficult. Such attainment ensures that this relationship would be maintained and enhanced through positive WOM. Hence, we hypothesize:

H6. Brand sacredness positively influences brand fidelity.

H7. Brand fidelity positively influences positive WOM.

2.2.4 Positive word of mouth. WOM is defined as "any positive or negative statement made by potential, actual or former customers about a product or a company, which is made available through offline or online channels" (Hennig-Thurau *et al.*, 2004). In the past few years, though, it has transformed into a more effective form, i.e. e WOM or electronic WOM; however, consumers do not easily spread positive WOM in offline and online environments.

The brand has to be superior in performance and quality to have its advocates offline and online. The role of emotional bonds with the brand becomes highly relevant here; as proposed by Carroll and Ahuvia (2006) brand love influences positive WOM. Therefore, the more robust the bond with the brand, the more likely is the probability of positive WOM being spread by the consumer (Kudeshia *et al.*, 2016). The consumers who consider brand superior and tend to behave as devotees of a brand do spread favourable WOM and such consumers think of brands as sacred entities:

H8. Brand sacredness positively influences positive WOM.

3. Methodology

The questionnaire for the current study was prepared through the review of several latest studies on the constructs included in the proposed framework. Scales (as mentioned in Appendix) were borrowed and accordingly modified for the requirements of our research. The stimulus for the study is a cosmetic brand named –"Iba halal care". The rationale for choosing this brand is based on the popularity and brand awareness of this brand as it is India's first halal certified cosmetic and personal care brand launched in India in 2014 (Garg and Joshi, 2018; Hassan and Sengupta, 2019). The halal market in India is still untapped and "Iba halal care" offers 100 different products to its customers (Kalpesh, 2019, September 29), therefore enjoying the first-mover advantage (Hassan and Sengupta, 2019). Currently, the brand is being sold online through shopping websites such as amazon.com, flipkart.com and nykaa.com. "Iba halal care" has also marked its online presence through social networking sites such as Facebook, which has more than 214k followers.

The data collection was done via online and offline mediums from Muslim customers of Delhi national capital region in India who were using this brand for more than one year. Snowball sampling and convenience sampling were used to gather responses, with most of them captured through online medium owing to the sudden emergence of COVID-19 pandemic. For the purpose of maintaining accuracy, an independent sample *T*-test was done to estimate the difference in responses from the online and offline medium. The resulting test values indicated that there was no significant difference in demographic profiles of online vs offline respondents (Deutskens *et al.*, 2006). A usable sample size of 403 was obtained after screening the responses of 700 customers giving a response rate of 57.57%. The use of this brand is not merely confined to Muslim consumers, as consumers practicing non-vegan buying habits also prefer to buy it. India embraces around 14.2% of the Muslim population (Garg and Joshi, 2018) and the study aims at capturing the emotions of Muslim consumers in the context of "halal" brand, hence a sample size of 403 seems adequate. Further, the sample size for the current study was comparable to similar studies done on halal cosmetics by Khan *et al.* (2020) and Sama and Trivedi (2019) Figure 1.

4. Data analysis

4.1 Reliability and common method bias

Internal consistencies of the items for the particular construct were evaluated through Cronbach's alpha values. Table 1 shows the estimated values for all the constructs were more than 0.7 (ranging from 0.798 to 0.884) therefore, reliability was achieved. Factor loadings for all the items were more 0.5, therefore all the items were strongly loaded on their respective factor (Joshi and Yadav, 2020b). Common method bias was estimated through Harman single factor test (Harman, 1960), the obtained value was 18.251%, which shows that no single factor accounted for more than 50% of variance (Joshi and Yadav, 2019).

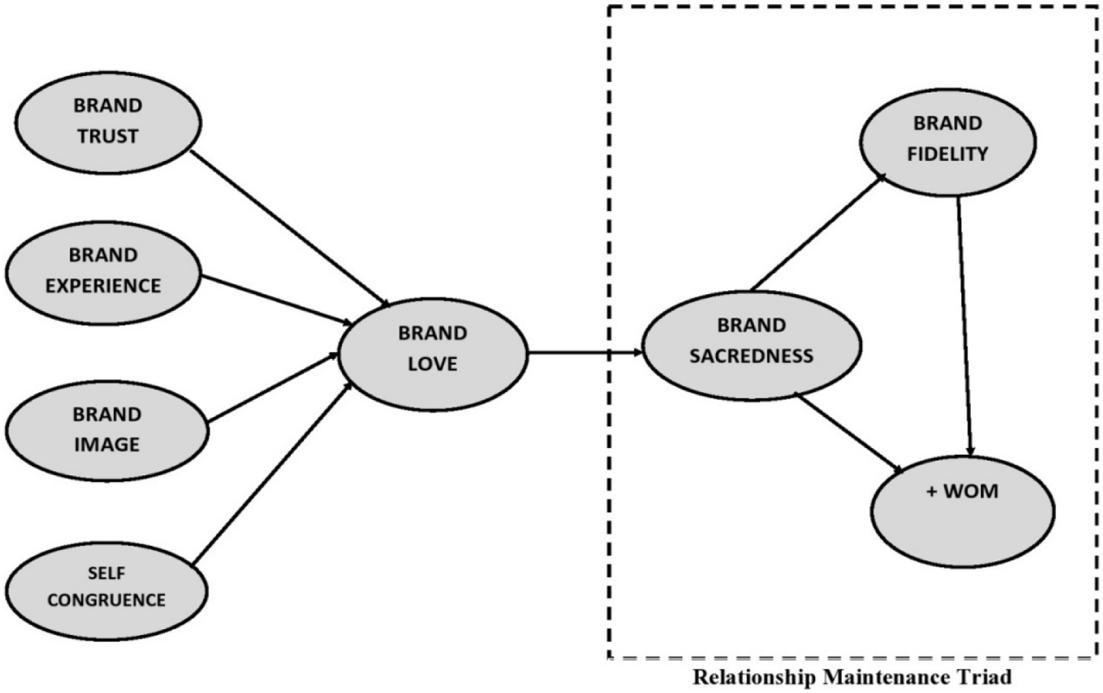


Figure 1.
Proposed framework of the study

Construct	Factor loadings	Cronbach's α
Brand trust	0.848 0.804 0.850	0.835
Brand experience	0.883 0.892 0.868	0.884
Brand image	0.875 0.840 0.846	0.849
Self-congruence	0.860 0.866 0.855	0.850
Brand love	0.776 0.824 0.834 0.604	0.844
Brand sacredness	0.798 0.801 0.815	0.799
Brand fidelity	0.840 0.852 0.774	0.798
Positive WOM	0.831 0.854 0.789	0.800

Table 1.
Reliability and factor loadings

4.2 Usage of structural equation modelling for data analysis

Structural equation modelling has been used for testing the proposed framework, as the technique helps in testing the relationships of several independent and dependent variables simultaneously. It is the best suited technique for confirmatory modelling (Kline, 2005) used to evaluate a hypothesized model with empirical data (Nunkoo and Ramkissoon, 2012). For the current study, the authors have adopted a two-stage approach (Joshi and Yadav, 2020b) proposed by Anderson and Gerbing (1988). In the first stage, the authors have analysed the measurement model while in the second stage, the structural model was tested to analyse the strength and direction of the proposed relationships among research constructs (Suparno, 2020).

4.3 Confirmatory factor analysis

The quality of measurements was assessed through the measurement model. Confirmatory factor analysis was done using analysis of a moment structures 22. Table 2 shows the values obtained, the average variance extracted (AVE) values for all the proposed constructs in the model exceeded the threshold of 0.5 as suggested by Hair *et al.* (2013) therefore, convergent validity was obtained (Garg and Pandey, 2020a). Discriminant validity was assessed through the square root of AVE and the values of inter-construct correlations. The square root of AVE exceeded the values of inter-construct correlations for all the constructs hence, discriminant validity was observed (Fornell and Larcker, 1981; Garg and Pandey, 2020b). The values of composite reliability and maximal reliability were also acceptable according to the prescribed threshold of 0.7 (Joshi and Yadav, 2020a).

4.4 Model fitness

Fitness of model was estimated with the help of several indices such as normed chi square, comparative fit index (CFI), goodness of fit index (GFI), standardized root mean square residual (SRMR) and root mean square error of approximation (RMSEA). The results of empirical testing revealed that χ^2/df, CFI, GFI, SRMR and RMSEA values were 2.753,0.911, 0.883, 0.048 and 0.066, respectively (as shown in Table 3). The obtained values of model fitness indices were within the prescribed thresholds (Ullman, 2001; Hair *et al.*, 2013; Junaid *et al.*, 2019; MacCallum *et al.*, 1996) therefore, the model was a good fit (Figure 2).

4.5 Path estimates

The path estimates of the structural model have been shown in Table 4. The *p*-values indicate that the effect of brand trust, brand experience, brand image and self-congruence on brand love was found to be positively significant; therefore, *H1–H4* were accepted ($P < 0.05$, 0.01). Brand love also have a significant effect on brand sacredness therefore, *H5* was accepted ($P < 0.05$, 0.01). Brand sacredness was observed to have a significant effect on brand fidelity, as $P < 0.05$, 0.01; hence, *H6* was also accepted. Similarly, the effect of brand fidelity and brand sacredness on positive WOM was also found significant, as the $P < 0.05$, 0.01; hence, *H7* and *H8* were also accepted, respectively.

5. Findings and discussion

The present study has enriched the existing literature by proposing a conceptual framework examining brand love's role as nexus between contemporary branding metrics and relationship maintenance triad. Brand love has been previously examined by different researchers (Batra *et al.*, 2012; Thomson *et al.*, 2005), but they somehow failed to unveil the romantic side of it. The findings of the current study confirm that brand trust, brand experience, brand image and self-congruence positively affect brand love. These findings

Constructs	CR	AVE	MSV	MaxR(H)	BT	BE	BI	SC	BLO	BS	BF	WOM
BT	0.836	0.631	0.237	0.846	*0.794*							
BE	0.885	0.719	0.128	0.887	0.223***	*0.848*						
BI	0.850	0.653	0.222	0.851	0.218***	0.102[a]	*0.808*					
SC	0.850	0.654	0.086	0.854	0.293***	0.188**	−0.058	*0.809*				
BLO	0.851	0.597	0.237	0.886	0.486***	0.357***	0.471***	0.266***	*0.773*			
BS	0.803	0.579	0.138	0.828	0.308***	0.127*	0.085	0.213***	0.363***	*0.761*		
BF	0.799	0.571	0.125	0.804	0.036	0.036	−0.093	0.137*	0.082	0.354***	*0.756*	
WOM	0.803	0.578	0.138	0.820	0.105a	0.068	−0.025	−0.005	0.108a	0.371***	0.344***	*0.760*

Notes: Significance of correlations: [a]$p < 0.100$; *$p < 0.050$; **$p < 0.010$; ***$p < 0.001$; CR = Composite reliability, AVE = Average variance extracted, MSV = Maximum shared variance, MaxR(H) = Maximal reliability

Table 2.
Reliability and validity estimates

are in accordance with the existing literature (Drennan *et al.*, 2015; Sallam, 2014; Roy *et al.*, 2013; Ismail and Spinelli, 2012; Zhou *et al.*, 2020) and confirm that consumer experiences strong feelings of affection or love for brands they can trust upon and they can associate themselves with. The image of a brand holds a significant role in shaping the feelings

Table 3.
Model fit indices

Measure	Estimate	Threshold	Interpretation
CMIN/DF	2.753	Between 1 and 3	Excellent
CFI	0.911	>0.90	Excellent
GFI	0.883	Near to 1	Acceptable
SRMR	0.048	<0.08	Excellent
RMSEA	0.066	<0.08, <0.10	Acceptable

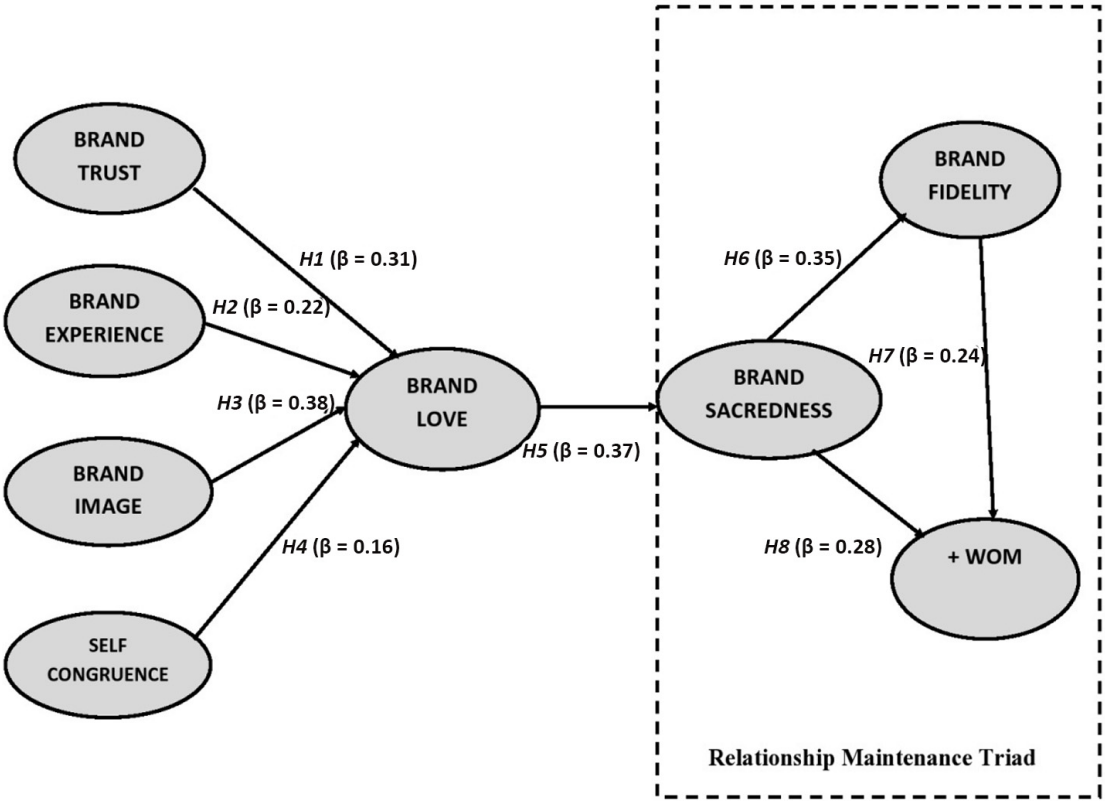


Figure 2.
Empirically tested framework

Table 4.
Path estimates

S. No.	Hypotheses	*P*-value	Results
H1	*Brand trust → brand love*	0.001	Supported
H2	*Brand experience → brand love*	0.001	Supported
H3	*Brand image → brand love*	0.001	Supported
H4	*Self-congruence → brand love*	0.002	Supported
H5	*Brand love → brand sacredness*	0.001	Supported
H6	*Brand sacredness → brand fidelity*	0.001	Supported
H7	*Brand fidelity → positive word of mouth*	0.001	Supported
H8	*Brand sacredness → positive word of mouth*	0.001	Supported

towards a brand; therefore, marketers must avoid a microscopic view of considering conventional branding concepts. To overcome this myopic view, they must also focus on creating great brand experiences that will invoke positive feelings in the minds of the consumers and build brand sacredness. A study by Zhou *et al.* (2020) revealed that self-congruence significantly influences brand love in the case of sportswear brands. The results of the present study also goes in consistency with this study and confirms that consumers with high self-congruity with a halal brand tends to show emotional connect with the brand, thus contributing to strong emotional bond in the form of brand love. The study further opens new avenues in the areas of brand fidelity and the intensity to propagate WOM by confirming significant positive effect of brand sacredness on brand fidelity and spread of WOM. These findings are noteworthy, as most of the literature has failed to capture the role of constructs such as brand sacredness and fidelity in literature pertaining to halal branding. The current study also posits that such constructs are very viable in areas where religiosity is at the core, and emotions act as a foundation to an individual's decision-making process. Brand sacredness being one of the contemporary construct goes far beyond measures such as brand love, brand loyalty and brand relationship quality (Thomson *et al.*, 2005; Batra *et al.*, 2012). It has a unique power of displaying cult-like emotions (Wang *et al.*, 2019), therefore it becomes imperative to examine the underlying effects of brand sacredness by extending it to emotions.

6. Conclusion

A study by Das and Mandal (2016) confirmed a significant effect of brand sacredness on online purchase intention and brand loyalty. They concluded that consumers get attached to brands and consider them as sacred entities, therefore marketers of poor-performing brands must use several strategies for promoting brands online via brand communities. They also argued that marketers need to figure out the role of social media in making the young population devoted to a brand. Therefore, the current study examined an extension to such strong emotion towards brand fidelity and WOM. Moreover, higher-level brand relationships such as love, fidelity and commitment may not be applicable to all brands and are found to be more persistent in brands with high customer involvement (Chaudhary, 2018). In light of this, halal brands being a symbol of religious association trigger the cognitive decision-making of the consumers. This, therefore, gave the researchers enough opportunity to examine the relation maintenance triad and examine higher-level brand relationships. In addition to this, emotions such as brand love and sacredness are difficult to measure, as they communicate the strength of the emotional connection between the consumer and the brand. Therefore, the consumers must be motivated enough to maintain the relationship with the brand, like any other relationship. For this reason, the study proposed a relationship maintenance triad by using brand fidelity, which represents the consumer's cognitive "effort" for maintaining a relationship and acts as an accurate indicator of the consumer's emotional attachment. The use of brand fidelity and brand sacredness in halal branding, therefore, becomes a significant contribution of this study to convey the religious side of halal branding and associated feelings among the Muslim customers. The study has revealed the role of emotions in showcasing the faith associated with halal brands and the amplifying effect of these emotions in creating a relationship maintenance triad.

7. Limitations and future directions

The findings of this study are centric towards examining and exploring consumer-brand relationships with reference to halal brands in India. Future research can however, extend consumer-brand relationships in the context of other geographical regions and cultural needs.

As consumer relationships have so far been extensively researched upon in tangible goods, therefore service industries such as tourism and hospitality, entertainment and health-care must also add a different perspective to the existing research. Although the model's ability to predict brand love, sacredness, fidelity and WOM has been reasonable, more manifestations of online brand communities' emotions and effects can add to its explanatory power. Further, the results cannot be generalized outside the current sampling frame, considering the small sample size of Islamic consumers. Therefore, it would be interesting to examine other demographics of the customers other than religion. All in all, the suggested avenues can immensely contribute to the existing inventory of literature in consumer-brand relationships.

References

Ahmad, N. (2018), "Toward advancing debates on Islamic marketing: a renewed perspective", *Journal of Islamic Marketing*, Vol. 9 No. 1, pp. 152-166.

Albert, N., Merunka, D. and Valette-Florence, P. (2008), "When consumers love their brands: exploring the concept and its dimensions", *Journal of Business Research*, Vol. 61 No. 10, pp. 1062-1075.

Al-Haddad, A. (2019), "Social identification, brand image and customer satisfaction as determinants of brand love", *Creative Business and Social Innovations for a Sustainable Future*, Springer, Cham, pp. 255-262.

Anderson, J.C. and Gerbing, D.W. (1988), "Structural modeling in practice: a review and recommended two-step approach", *Psychological Bulletin*, Vol. 103 No. 3, pp. 411-423.

Batra, R., Ahuvia, A. and Bagozzi, R.P. (2012), "Brand love", *Journal of Marketing*, Vol. 76 No. 2, pp. 1-16.

Battour, M. and Ismail, M.N. (2016), "Halal tourism: concepts, practices, challenges and future", *Tourism Management Perspectives*, Vol. 19, pp. 150-154.

Belk, R.W., Wallendorf, M. and Sherry, J.F. Jr (1989), "The sacred and the profane in consumer behavior: theodicy on the odyssey", *Journal of Consumer Research*, Vol. 16 No. 1, pp. 1-38.

Bergkvist, L. and Bech-Larsen, T. (2010), "Two studies of consequences and actionable antecedents of brand love", *Journal of Brand Management*, Vol. 17 No. 7, pp. 504-518.

Borzooei, M. and Asgari, M. (2013), "Halal branding and purchase intention: a brand personality appeal perspective", *International Journal of Business and Management Invention*, Vol. 2 No. 8, pp. 23-27.

Brakus, J., Schmitt, B. and Zarantonello, L. (2009), "Brand experience: what is it? How is it measured? Does it affect loyalty?", *Journal of Marketing*, Vol. 73 No. 3, pp. 52-68.

Carroll, B.A. and Ahuvia, A.C. (2006), "Some antecedents and outcomes of brand love", *Marketing Letters*, Vol. 17 No. 2, pp. 79-89.

Chaudhary, A.H. (2018), "Retracted article: brand love: fiction or reality?", *Journal of Strategic Marketing*, Vol. 26 No. 8.

Chaudhuri, A. and Holbrook, M.B. (2001), "The chain of effects from brand trust and brand affect to brand performance: the role of brand loyalty", *Journal of Marketing*, Vol. 65 No. 2, pp. 81-93.

Chiu, C.M., Huang, H.Y. and Yen, C.H. (2010), "Antecedents of online trust in online auctions", *Electronic Commerce Research and Applications*, Vol. 9 No. 2, pp. 148-159.

Das, P. and Mandal, S. (2016), "Evaluating the influence of social media on brand sacralisation", *South Asian Journal of Global Business Research*, Vol. 5 No. 3, pp. 424-446.

Delgado-Ballester, E. and Munuera-Alemán, J.L. (2001), "Brand trust in the context of consumer loyalty", *European Journal of Marketing*, Vol. 35 Nos 11/12, pp. 1238-1258.

Deutskens, E., De Ruyter, K. and Wetzels, M. (2006), "An assessment of equivalence between online and mail surveys in service research", *Journal of Service Research*, Vol. 8 No. 4, pp. 346-355.

Drennan, J., Bianchi, C., Cacho-Elizondo, S., Louriero, S., Guibert, N. and Proud, W. (2015), "Examining the role of wine brand love on brand loyalty: a multi-country comparison", *International Journal of Hospitality Management*, Vol. 49, pp. 47-55.

Ekinci, Y., Sirakaya-Turk, E. and Preciado, S. (2013), "Symbolic consumption of tourism destination brands", *Journal of Business Research*, Vol. 66 No. 6, pp. 711-718.

Fornell, C. and Larcker, D.F. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18 No. 1, pp. 39-50.

Fournier, S. (1998), "Consumers and their brands: developing relationship theory in consumer research", *Journal of Consumer Research*, Vol. 24 No. 4, pp. 343-373.

Fromm, E. (1956), *The Art of Loving*, Harper and Row, New York, NY.

Garg, P. and Joshi, R. (2018), "Purchase intention of 'halal' brands in India: the mediating effect of attitude", *Journal of Islamic Marketing*, Vol. 9 No. 3, pp. 683-694.

Garg, P. and Pandey, A. (2020a), "Examining moderating role of personal identifying information in travel related decisions", *International Journal of Tourism Cities*, Vol. 6 No. 3, pp. 621-638.

Garg, P. and Pandey, A. (2020b), "Towards sustainable tourism: an empirical investigation", *Foresight*, doi: 10.1108/FS-04-2020-0042.

Gilmore, J. and Pine, J. (1999), *The Experience Economy: Work is Theatre and Every Business a Stage*, Harvard Business Publishing, Cambridge, MA.

Grace, D., Ross, M. and King, C. (2018), "Brand fidelity: a relationship maintenance perspective", *Journal of Brand Management*, Vol. 25 No. 6, pp. 577-590.

Grace, D., Ross, M. and King, C. (2020), "Brand fidelity: scale development and validation", *Journal of Retailing and Consumer Services*, Vol. 52, p. 101908.

Ha, H.Y. (2004), "Factors influencing consumer perceptions of brand trust online", *Journal of Product and Brand Management*, Vol. 13 No. 5, pp. 329-342.

Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2013), *Multivariate Data Analysis: Pearson New International Edition*, Pearson Higher Ed.

Harman, H.H. (1960), *Modern Factor Analysis*, University of Chicago Press, Chicago, IL.

Hassan, Y. and Sengupta, A. (2019), "India – an untapped market for halal products", *Journal of Islamic Marketing*, Vol. 10 No. 3, pp. 981-1002.

Hennig-Thurau, T., Gwinner, K.P., Walsh, G. and Gremler, D.D. (2004), "Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet?", *Journal of Interactive Marketing*, Vol. 18 No. 1, pp. 38-52.

Huang, C.-C. (2017), "The impacts of brand experiences on brand loyalty: mediators of brand love and trust", *Management Decision*, Vol. 55 No. 5, pp. 915-934.

Islam, J.U. and Rahman, Z. (2016), "Examining the effects of brand love and brand image on customer engagement: an empirical study of fashion apparel brands", *Journal of Global Fashion Marketing*, Vol. 7 No. 1, pp. 45-59.

Ismail, A. and Spinelli, G. (2012), "Effects of brand love, personality and image on word of mouth", *Journal of Fashion Marketing and Management: An International Journal*, Vol. 16, pp. 386-398.

Japutra, A., Ekinci, Y. and Simkin, L. (2019), "Self-congruence, brand attachment and compulsive buying", *Journal of Business Research*, Vol. 99, pp. 456-463.

Joshi, R. and Garg, P. (2020), "Role of brand experience in shaping brand love", *International Journal of Consumer Studies*, doi: 10.1111/ijcs.12618.

Joshi, R. and Yadav, R. (2016), "Reciprocal effects of brand extension on brand equity: a literature review", *International Journal of Marketing and Business Communication*, Vol. 5 No. 2, pp. 20-28.

Joshi, R. and Yadav, R. (2019), "The study of brand extension among generation Y in the Indian market", *International Journal of Indian Culture and Business Management*, Vol. 19 No. 3, pp. 339-353.

Joshi, R. and Yadav, R. (2020a), "Captivating brand hate using contemporary metrics: a structural equation modelling approach", *Vision*.

Joshi, R. and Yadav, R. (2020b), "Brand desire: scale development and empirical examination", *Journal of Asia-Pacific Business*, Vol. 21 No. 3, pp. 169-184.

Junaid, M., Hou, F., Hussain, K. and Kirmani, A.A. (2019), "Brand love: the emotional bridge between experience and engagement, generation-M perspective", *Journal of Product and Brand Management*, Vol. 28 No. 2, pp. 200-215.

Kalpesh, D. (2019), "Ahmedabad's 'halal' firm aims Rs 100 crore business", available at: https://timesofindia.indiatimes.com/city/ahmedabad/abads-halal-firm-aims-rs-100cr-biz/articleshow/71355034.cms (accessed 17 October 2020).

Khan, N., Sarwar, A. and Tan, B.C. (2020), "Determinants of purchase intention of halal cosmetic products among generation Y consumers", *Journal of Islamic Marketing*, doi: 10.1108/JIMA-11-2019-0248.

Kline, R.B. (2005), *Principles and Practice of Structural Equation Modelling*, Guilford, New York, NY, Vol. 2.

Kotler, P. (2004), *Armstrong: Principles of Marketing*, 7th ed., Prentice Hall, NJ.

Kudeshia, C., Sikdar, P. and Mittal, A. (2016), "Spreading love through fan page liking: a perspective on small scale entrepreneurs", *Computers in Human Behavior*, Vol. 54, pp. 257-270.

MacCallum, R.C., Browne, M.W. and Sugawara, H., M. (1996), "Power analysis and determination of sample size for covariance structure modelling", *Psychological Methods*, Vol. 1 No. 2, pp. 130-149.

Mody, M. and Hanks, L. (2019), "Consumption authenticity in the accommodations industry: the keys to brand love and brand loyalty for hotels and airbnb", *Journal of Travel Research*, Vol. 59 No. 1.

Moore, D. and Wurster, D. (2007), "Self-brand connections and brand resonance: the role of gender and consumer emotions", *Advances in Consumer Research*, Vol. 34, p. 64.

Nunkoo, R. and Ramkissoon, H. (2012), "Structural equation modelling and regression analysis in tourism research", *Current Issues in Tourism*, Vol. 15 No. 8, pp. 777-802.

Park, C.W., MacInnis, D.J. and Priester, J. (2008), "Brand attachment: constructs, consequences, and causes", *Foundations and Trends® in Marketing*, Vol. 1 No. 3, pp. 191-230.

Park, C.W., MacInnis, D.J. and Priester, J.R. (2007), "Brand attachment and management of a strategic brand exemplar", *Handbook on Brand and Experience Management*, Edward Elgar Pub, Cheltenham.

Pavlou, P.A., Liang, H. and Xue, Y. (2007), "Understanding and mitigating uncertainty in online exchange relationships: a principal-agent perspective", *MIS Quarterly*, Vol. 31, pp. 105-136.

Rauschnabel, P.A. and Ahuvia, A.C. (2014), "You're so lovable: anthropomorphism and brand love", *Journal of Brand Management*, Vol. 21 No. 5, pp. 372-395.

Rindfleisch, A., Wong, N. and Burroughs, J.E. (2006), *Seeking Certainty via Brands: An Examination of Materialism and Brand Resonance*, Association for Consumer Research.

Roy, S.K., Eshghi, A. and Sarkar, A. (2013), "Antecedents and consequences of brand love", *Journal of Brand Management*, Vol. 20 No. 4, pp. 325-332.

Rusbult, C.E., Agnew, C.R. and Arriaga, X.B. (2012), "The investment model of commitment processes", in Van Lange, P.A.M., Kruglanski, A.W. and Higgins, E.T. (Eds), *Handbook of Theories of Social Psychology*, Sage, Thousand Oaks, CA, pp. 218-231.

Sallam, M.A. (2014), "The effects of brand image and brand identification on brand love and purchase decision making: the role of WOM", *International Business Research*, Vol. 7 No. 10, p. 187.

Sama, R. and Trivedi, J.P. (2019), "Factors affecting consumers' loyalty towards halal cosmetics: an emerging market perspective", *International Journal of Business and Emerging Markets*, Vol. 11 No. 3, pp. 254-273.

Sarkar, A. and Sreejesh, S. (2014), "Examination of the roles played by brand love and jealousy in shaping customer engagement", *Journal of Product and Brand Management*, Vol. 23 No. 1, pp. 24-32.

Sarkar, J.G., Sarkar, A. and Ponnam, A. (2015), "Exploration of brand sacralization among the young adult consumers in the context of emerging Asian market", *Young Consumers*, Vol. 16 No. 3, pp. 264-280.

Shabbir, M.Q., Khan, A.A. and Khan, S.R. (2017), "Brand loyalty brand image and brand equity: the mediating role of brand awareness", *International Journal of Innovation and Applied Studies*, Vol. 19 No. 2, pp. 416-423.

Shachar, R., Erdem, T., Cutright, Keisha, M. and Fitzsimons, G.J. (2011), "Brands: the opiate of the nonreligious masses?", *Marketing Science*, Vol. 30 No. 1, pp. 92-110.

Shimp, T.A. and Madden, T.J. (1988), "Consumer – object relations: a conceptual framework based analogously on Sternberg's triangular theory of love", in Houston, M.J. (Ed.), *Advances in Consumer Research*, Association for Consumer Research, Provo, UT, pp. 163-168.

Song, H., Wang, J. and Han, H. (2019), "Effect of image, satisfaction, trust, love, and respect on loyalty formation for name-brand coffee shops", *International Journal of Hospitality Management*, Vol. 79, pp. 50-59.

Sternberg, R.J. (1986), "A triangular theory of love", *Psychological Review*, Vol. 93 No. 2, pp. 119-135.

Suki, J.S.N.M. and Sasmita, J. (2015), "Young consumers' insights on brand equity: effects of brand association, brand loyalty, brand awareness, and brand image", *International Journal of Retail and Distribution Management*, Vol. 43 No. 3, pp. 276-292.

Suparno, C. (2020), "Online purchase intention of halal cosmetics: S-O-R framework application", *Journal of Islamic Marketing*, doi: 10.1108/JIMA-09-2019-0192.

Thomson, M., MacInnis, D.J. and Park, C.W. (2005), "The ties that bind: measuring the strength of consumers' emotional attachment to brands", *Journal of Consumer Psychology*, Vol. 15 No. 1, pp. 77-91.

Ullman, J.B. (2001), "Structural equation modeling", *Using Multivariate Statistics*, 4th ed., Allyn and Bacon, Needham Heights, MA.

Wallace, E., Buil, I. and de Chernatony, L. (2017), "Consumers' self-congruence with a 'liked' brand", *European Journal of Marketing*, Vol. 51 No. 2, pp. 367-390.

Wang, C.L., Sarkar, J.G. and Sarkar, A. (2019), "Hallowed be thy brand: measuring perceived brand sacredness", *European Journal of Marketing*, Vol. 53 No. 4, pp. 733-757.

Wilson, J.A. (2014), "The halal phenomenon: an extension or a new paradigm?", *Social Business*, Vol. 4 No. 3, pp. 255-271.

Wilson, J.A.J. and Grant, J. (2013), "Islamic marketing – a challenger to the classical marketing canon?", *Journal of Islamic Marketing*, Vol. 4 No. 1, pp. 7-21.

Wilson, J.A.J. and Liu, J. (2010), "Shaping the halal into a brand?", *Journal of Islamic Marketing*, Vol. 1 No. 2, pp. 107-123.

Wilson, J.A.J. and Liu, J. (2011), "The challenges of Islamic branding: navigating emotions and halal", *Journal of Islamic Marketing*, Vol. 2 No. 1, pp. 28-42.

Zhou, F., Mou, J., Su, Q. and Wu, Y.C.J. (2020), "How does consumers' perception of sports stars' personal brand promote consumers' brand love? A mediation model of global brand equity", *Journal of Retailing and Consumer Services*, Vol. 54, p. 102012.

Further reading

Leventhal, R.C., Wallace, E., Buil, I. and de Chernatony, L. (2014), "Consumer engagement with self-expressive brands: brand love and WOM outcomes", *Journal of Product and Brand Management*, Vol. 23 No. 1, pp. 33-42.

Sarkar, A., Ponnam, A. and Murthy, B.K. (2012), "Understanding and measuring romantic brand love", *Journal of Customer Behaviour*, Vol. 11 No. 4, pp. 324-347.

Yoo, B. and Donthu, N. (2001), "Developing and validating a multidimensional consumer-based brand equity scale", *Journal of Business Research*, Vol. 52 No. 1, pp. 1-14.

Appendix

(1) Brand trust (adapted from Drennan *et al.* (2015))
- This is a reliable cosmetic brand.
- This cosmetic brand is likeable.
- This cosmetic brand is a very good brand.

(2) Brand experience (adapted from Huang, 2017)
- I find this cosmetic brand interesting in a sensory way.
- I engage in a lot of thinking when I encounter this cosmetic brand.
- I engage in physical actions when I use this cosmetic brand.

(3) Brand image (adapted from Suki and Sasmita (2015))
- This cosmetic brand has a differentiated image in comparison with the other brand.
- This cosmetic brand has a clean image.
- This cosmetic brand is well-established.

(4) Self-congruence (adapted from Japutra *et al.* (2019))
- This cosmetic brand is consistent with how I see myself.
- This cosmetic brand is a mirror image of me.
- This cosmetic brand is similar to me.

(5) Brand love (adapted from Carroll and Ahuvia (2006))
- This is a wonderful cosmetic brand.
- I love this cosmetic brand.
- This cosmetic brand makes me feel good.
- This cosmetic brand is totally awesome.

(6) Brand sacredness (adapted from Das and Mandal (2016))
- I am deeply devoted to this cosmetic brand I prefer to buy.
- I consider my preferred cosmetic brand as a sacred entity.
- I am emotionally involved with the cosmetic brand I prefer.

(7) Brand fidelity (adapted from Grace *et al.* (2020))
- When this cosmetic brand has had a price increase, it has been well-justified.
- I would feel offended if someone said something bad about this cosmetic brand.
- This cosmetic brand is one of a kind and, in my opinion, there is no competition.

(8) Positive WOM (adapted from Carroll and Ahuvia (2006))
- I have recommended this cosmetic brand to lots of people.
- I "talk up" this cosmetic brand to my friends.
- I try to spread the good-word about this cosmetic brand.

About the authors

Dr Richa Joshi is an Assistant Professor in National Institute of Technology (NIT) Hamirpur (HP). She has more than 10 years of teaching experience. The areas of her interest are brand equity, brand extension and consumer brand relationships. She has completed her doctoral degree from Delhi Technological University. She has obtained her graduation from Delhi University and post-graduation from GGSIP University with the qualification of UGC-NET in Management. She has presented papers in various national and international conferences at IIM Indore, IIM Kashipur, MDI Gurgaon and IIT Delhi. She has also contributed "ABDC" and "Scopus" indexed research papers in

several journals of Sage, Inderscience and Emerald publications. She has been an active reviewer of journals of Sage and Emerald. Richa Joshi is the corresponding author and can be contacted at: richardias86@gmail.com; richajoshi@nith.ac.in

Ms Prerna Garg is a research scholar at Bharati Vidyapeeth University, Pune and Assistant Professor in Jaipuria Institute of Management, Ghaziabad. She possesses a rich experience of over 10 years with Consumer Behaviour, electronic marketing in Travel industry and Branding as her areas of interest. She has published many scholarly articles in journals of repute (including Scopus listed and ABDC indexed journals) and has presented research papers in several International/National conferences at IIMs, IMT, FMS and many more. She has actively participated in AICTE sponsored SDP, FDP, Seminars and QIPs.

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

arXiv:2012.02291v1 [cs.IR] 29 Nov 2020

Cluster Based Deep Contextual Reinforcement Learning for top-k Recommendations

Anubha Kabra, Anu Agarwal, Anil Singh Parihar

Department of Computer Science, Delhi Technological University
anubhakabradtu@gmail.com,anuag7699@gmail.com, anil@dtu.ac.in

Abstract. Rapid advancements in the E-commerce sector over the last few decades have led to an imminent need for personalized, efficient and dynamic recommendation systems. To sufficiently cater to this need, we propose a novel method for generating top-k recommendations by creating an ensemble of clustering with reinforcement learning. We have incorporated DB Scan clustering to tackle vast item space, hence increasing the efficiency multi-fold. Moreover, by using deep contextual reinforcement learning, our proposed work leverages the user features to its full potential. With partial updates and batch updates, the model learns user patterns continuously. The Dueling Bandit based exploration provides robust exploration as compared to the state-of-art strategies due to its adaptive nature. Detailed experiments conducted on a public dataset verify our claims about the efficiency of our technique as compared to existing techniques.

Keywords: Recommendation System, Deep Neural Networks, Deep Reinforcement Learning, Clustering

1 Introduction

With the ever-growing availability of commodities and services online, the consumer community has conveniently started shifting to online mediums for fulfilment of their needs. To capture the interest of consumers online, prompt and real time recommendation systems are needs. The current recommendation systems are non dynamic i.e: they do not update themselves with time. This discards the trend that might uplift the recommendation which is more relevant to the consumer. To keep the consumer retention intact, it is necessary that the recommendation systems update themselves from time to time to learn the changing user behaviour. These changing trends can be captured using the upcoming reinforcement learning techniques. Reinforcement Learning focuses on updating the policy in batches rather than splitting the data into training and testing. When such split is made, the policy does not generally learn anything from the testing data. Following are the challenges that the system faces with static recommendation systems :

1. Ignoring Trend : Since the system is trained on the entire training set all at once, the trends present in the consumer behaviour data get mixed up and averaged-out. This leads to non-relevant recommendation for consumers. For industry, tackling the non stationary environment is very important or it may lead to consumers loosing interest. Reinforcement learning techniques update the system frequently, hence learning the changes in the environment with time.
2. Disregarding Features : Most of the existing recommendation systems rely only on the consumer and item frequency or rating. They usually don't have a provision of utilising additional features about the consumers that may be available, that may help in recommendation. Currently, the recommendation systems rely on the narrow scope of how frequently a consumer or item is seen, and expilcit user feedback in terms of ratings [5]. With reinforcement based learning, these problems can be curbed by using contextual RL, which considers all features available in the data set before making valid recommendations.
3. No accommodation for new consumers : Due to the training and testing steps being discrete, if a new user joins the system, the entire system needs to be retrained. This is a very cumbersome process. Reinforcement learning based strategies can easily handle this situation by updating the system with each new trial. So, if a new user joins the system, the agent will learn about the user with time without re-training.
This paper focuses on exploration using dueling bandit gradient descent for reinforcement learning based recommendation strategy. The proposed work also narrows the item-set to be considered for recommendation by using DB Scan clustering. The novelty lies in combining both strategies to overcome the drawbacks of current recommendation strategies and make it fit for continuously changing online environment. The combined model meets the industry standards due to its swift working, while exhibiting recommendations highly applicable to users.

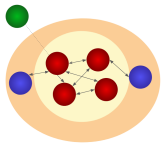


Fig. 1: DB SCAN

2 Related Work

Existing Recommendation Systems can be broadly divided into major categories. The first category consists of the traditional recommendation algorithms which are either popularity based or collaborative. Popularity based systems always recommend the same items at given time, irrespective of the user. On the other

hand, collaborative-Filtering (CF) based systems [2, 12] look for other users which are similar to the given user and recommend him/her the items already purchased by them. It is still one of the most popular and successful techniques. Content-based filtering [6, 6, 13] forms another category of recommenders which generates recommendations based on items already purchased by the user. However, this technique generally doesn't work well for platforms serving items of multiple categories as this method does not generalize well across different categories. Yet another traditional algorithm is Hybrid Filtering [3, 9] which uses a combination of above two techniques. Matrix Factorization [7] extends CF as it attempts to identify latent features from user-ratings data by expressing the user-item interaction matrix as a product of multiple lower dimensional matrices. Further attempts [19] have been made to incorporate the use of neural networks with the aforementioned techniques to improve the results.

The second major and relatively modern category includes recommendation systems which use reinforcement learning [11, 15, 16] and are inherently dynamic in nature.The recommendation scenario maps well to a usual Reinforcement Learning problem where an agent learns to behave in an environment via a series of sequential observations. These sequential observations can be mapped to the user-item interaction logs generated over time. Further, this task has been modelled by some as Markov Decision Problem [10, 14] and others as a Multi-Armed Bandit problem [8]. Attempts have also been made to incorporate session context into the system to provide more relevant recommendations [1, 17, 18]. Our proposed work combines contextual Multi armed bandits strategy with DB-SCAN clustering Dueling Bandits exploration strategy to provide highly personalised recommendations in a more efficient manner.

3 Methodology

3.1 DB Scan Clustering

DB scan is a clustering algorithm that is density-based. The distance between points is used to find the neighbors (points withing close vicinity). This measurement is generally done on the basis of Euclidean distance. Furthermore, it specifies the out-of-bound points and low-density regions. DB scan has two parameters : **minPts:** For the formation of a cluster, least number of points to be considered. **eps:** The distance between two points to be called neighbours. This distance measures the closeness. DB scan corresponds well with the real-life data. Due to the unpredictable structure in such data with a lot of noise, DB scan algorithm can fit these irregularities well, which is generally denied in other clustering algorithms. Eg: K Means clustering generally fails when the data is not spherical in nature, whereas DB Scan can capture this. Figure 1 shows the Red points to be the core points which satisfy the minPts criteria and are in the eps vicinity. The Blue points satisfy the eps distance but exceed the minPts criteria and hence are not included in the cluster (Yellow Area). The Green Point

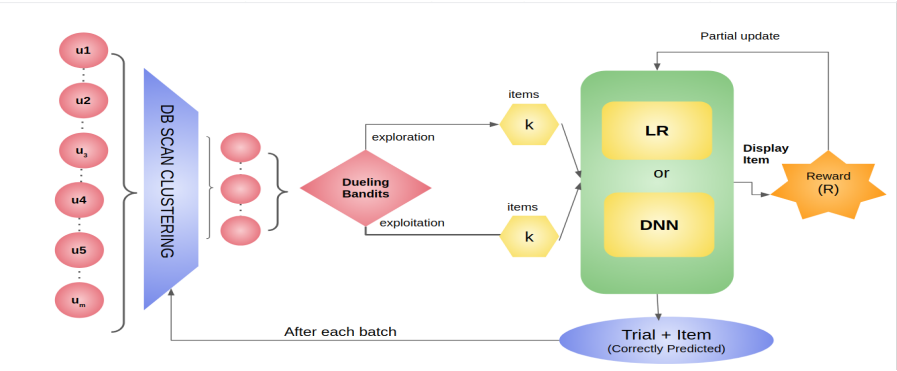


Fig. 2: Proposed Model Pipeline

does not satisfy any of these conditions, and hence is referred to as noise.

Initialization
D : dataset
p_u : Un-visited Points in the dataset
p_v : Visited points in dataset
C : cluster index
for *non-visited* p_u *in* D **do**
  p_{u_i} changes to p_{v_i}
  N : neighbours(p_i,eps)
  if $N > minPts$ **then**
    N = N U N'
  if *found p' does not belong to any cluster* **then**
    p' ∈ C
end

Algorithm 1: DB Scan Algorithm (eps, minPts)

3.2 Dueling Bandit Gradient Descent

Dueling Bandit Gradient Descent is an adaptive exploration strategy unlike epsilon greedy, UCB, TS etc. It attempts to present the best recommendations to the user by combining exploration and exploitation in appropriate proportions. As the name suggests, in this technique, two models work simultaneously. First model M takes decision based on patterns learnt from the data whereas the second model M' is exploratory in nature. For every row in the dataset, both models generate a list of k items each. These lists are then combined by the reinforcement learning agent using probabilistic interleaving [4] to generate a final list of k items. This combined list is presented to the user and partial feedback is obtained for each item on the list. If more items from the list generated by M' get positive, then the parameters (P) of model M are updated to incline it

towards M' as shown in Equation 1, else the model is kept unchanged.

$$P_{new} = P_{old} + \beta P' \tag{1}$$

Where other exploration strategies tend to lower the performance of the recommender in the long run because of being predominantly random in nature, this technique strikes a good balance between exploration and exploitation and adapts as and when required.

3.3 Model Pipeline

STEP 1 : Clustering Due to the vast number of applications which can be recommended to user, DB Scan clustering can help reduce the number of items to be taken into consideration before recommendation. The initial batch size does not implement DB Scan clustering. Collection of correctly predicted user context and item is done.Clusters of items are made using context features of the first batch. For the next batch, when a trial sample is passed through this clustering mechanism, it first clusters applications relatable to the user and its features, from which best k items are selected to be shown to user with that particular features. The re-clustering is done after every batch is processes with the building collection of trail and corresponding correctly predicted arm. This technique works very well to improve efficiency since the pool of applications considered is not the entire item-set but a smaller chunk of it. Moreover it helps to understand the user behaviour with applications better.

STEP 2: Exploration or Exploitation After clustering, decision is made on whether to pursue exploration(selecting an item for user with some form of randomised way) or exploitation (selecting best item for the user) however this is not as straightforward as epsilon-greedy. The strategy used for this decision is Dueling Bandits Gradient Descent which generates a list combined from both exploration and exploitation models. The exploit network adapts itself further in the process if required based on the user feedback.

STEP 3: Reward Formulation The k selected items are passed through the base model to predict if user has chosen any one of them or not.There are two base models experimented with. First one is a basic Logistic Regression(LR) based model while the other is a Deep Neural Network(DNN). DNN gives higher performance than LR. If any of the k items matches with the actually selected item, the reward is 1, else the reward is 0.This reward, along with the trial features of that particular user is fed back into the model for continuous learning.

STEP 4: Partial Update

The model is updated in random batches on default size 1000 selected from all the trials experienced till now. The trial and reward information is stored over time in memory. Random mini batch of 1000 from this set after every 5000 trials and updated back to the base model. This helps to keep the trend of the market intact and learn a pattern similarity between context features of different users.

4 Experimentation

4.1 Dataset

The dataset used consists of user and mobile application interaction logs. The Data is collected from TalkingData SDK [1] integrated within mobile apps TalkingData serves under the service term between TalkingData and mobile app developers.Here the apps are considered as items to be recommended and user and their information as the context used for the same. The categorical features are one hot encoded and continuous features are scaled before being given as input.

4.2 Baselines

Static Techniques
• *LR:* Logistic Regression is widely used in the industry as a baseline method due to its easy implementation and high efficiency. It is implemented using Keras.
• *FM:* Factorization Machines is a feature handy recommendation system which takes in all contextual features and predicts clicks/non-click.
• *W&D:* Wide & Deep combines dependent learning using textbfwide : a normal representation of the features and **deep**: embedding representation of feature to predict the click label.

Table 1: Abbreviation and Code Number for each policy .Proposed policies are given in **bold**.

Code	Abbreviation	Policy
1	LR	Logistic Regression
2	W&D	Wide and Deep
3	FM	Factorization Machines
4	B-UCB	Bootstrapped Upper Confidence Bound
5	B-TS	Bootstrapped Thompson Sampling
6	E-Greedy	Epsilon Greedy
7	FEE	First Explore then Exploit
8	AE	Active Explorer
9	**DB+LR**	**Dueling Bandits+LR**
10	**DB +DNN**	**Dueling Bandits+DNN**
11	**DB_Scan + DB +DNN**	**DB Scan+ Dueling Bandits+DNN**

Dynamic Techniques
• *Contextual Bandits* These are Reinforcement Learning techniques with base classifier as Logistic Regression. The following exploration strategies are explored:
• *Epsilon Greedy :*The idea here is to explore with certain fixed possibility. The exploration threshold is generally kept very small. Here default is set to 0.2. This has many variations for improved exploration.

[1] https://www.kaggle.com/c/talkingdata-mobile-user-demographics/data

• *First explore then exploit :* The exploration in this case is done initially, and later only exploitation is taken into consideration. The system assumes that initial exploration is good enough for system to learn.
• *Bootstrapped UCB [1] :* UCB works on upper confidence bound strategy for exploration.The bootstrapping ensures small pool of trails to be taken at a single time.
• *Bootstrapped TS [1] :* Thompson sampling creates distribution for each item, which changes according to the reward updation for the selected trial. This strategy involves forming a prior distribution for each arm.
• *Active Explorer [1] :* This strategy focuses on weighted label based probabilities.

4.3 Evaluation Metrics

CTR : Click behaviour can be used to analyze how well a recommendation system is performing. CTR helps to capture the user click behaviour.

$$CTR_i = \frac{Total\,number\,of\,clicks}{Total\,number\,of\,occurrences} \tag{2}$$

The average of CTR values of all the items is as follows:

$$CTR = \frac{\sum_{i=0}^{N} CTR_i}{N} \tag{3}$$

Precision@k : Precision@k for top-k recommendations is calculated as follows:

$$Precision@k = \frac{number\,of\,clicks\,in\,top-k\,recommended\,items}{k} \tag{4}$$

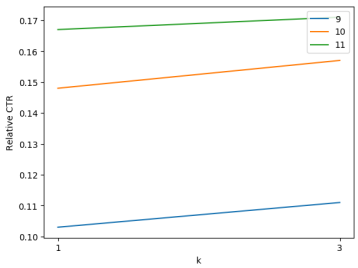
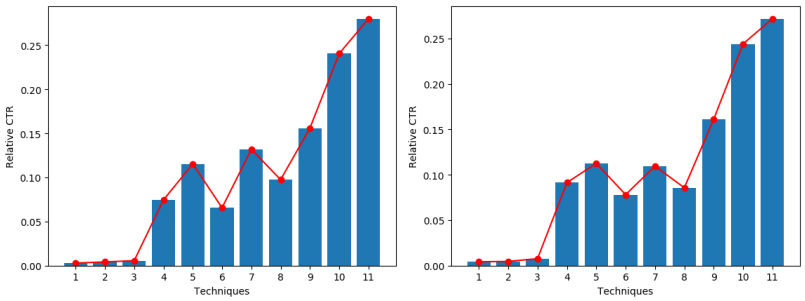


Fig. 3: Comparative analysis of proposed methods for 1 and 3 item recommendations

4.4 Experimentation Settings

Three simulations were used for experimentation as shown in Table 1. The techniques **9** and **10** use different base model.Technique **9** has used Logistic Regression to simulate [1] strategy. **10** uses Deep Neural Network(DNN) as the base model. Since DNN gives significantly better results, technique **11** uses DNN as base model with DB Scan Clustering. Experiments have been conducted for top-1 and top-3 recommendations. While static strategies have a train-test split of 70:30. The dynamic techniques rely on learning throughout the process and no split has been done. To evaluate the results however, 30% of the data is considered, which is kept identical for all experiments.



(a) $k = 1$ (b) $k = 3$

Fig. 4: Relative CTR trends for different values of k

5 Results and Analysis

It can be clearly deduced from the CTR and Precision@k values mentioned in Table 2 that the methods which rely on static (one-time) model training such as LR, W&D and FM, perform poorly in comparison with the rest of the reinforcement learning based techniques. Thus, it can be concluded that dynamic model training is pivotal when it comes to applications such as recommender systems, because the user's underlying behaviour pattern keeps changing with time.Further, on comparing the CTR values for different exploration strategies, it can be seen that the adaptive techniques such as Active Explorer perform better than purely random exploration strategies like ϵ-greedy. However, the best performance is given by Dueling Bandit Gradient Descent strategy as it diversifies the network and inclines it towards exploration if favourable user feedback is obtained. The relative CTR graphs in Figure 4 for different values of k show that in-spite of a slight decrease in the overall values from k=1 to k=3, the general trend still remains the same. Non-RL based techniques show

Table 2: Result for each policy mentioned with its respective CTR and Accuracy values .

Abbreviation	CTR		Precision@k	
	k=1	k=3	k=1	k=3
LR	0.0022	0.0042	0.28	0.28
W&D	0.0032	0.0010	0.33	0.36
FM	0.0036	0.0051	0.35	0.32
B-UCB	0.0243	0.0276	0.54	0.59
B-TS	0.0243	0.0276	0.67	0.70
E-Greedy	0.0288	0.0323	0.44	0.48
FEE	0.0294	0.0388	0.69	0.68
AE	0.0285	0.3117	0.48	0.49
DB+LR	**0.0432**	**0.0566**	**0.75**	**0.77**
DB +DNN	**0.0965**	**0.0972**	**0.82**	**0.85**
DB_Scan + DB + DNN	**0.1342**	**0.0157**	**0.89**	**0.92**

just about no change in results with changing k values. Since RL based techniques incorporate feedback for every recommended item, the change in values is comparatively more significant for them.Figure 3 shows a direct performance comparison between different variations of deep contextual bandits proposed in this paper. Even though DB + DNN seems to be the best strategy as it give the highest values throughout, DB_SCAN + DB + DNN is the most efficient and practical alternative. Inclusion of clustering straightaway reduces the number of comparisons by a factor of x/N where x is the average size of a cluster.

6 Conclusion

The proposed work focuses on improving recommendations of items to users in real time scenario. Using DB Scan clustering on the data reduces the processing cost making the recommender efficient. Moreover, Reinforcement Learning techniques with Dueling Bandit Gradient Descent exploration and partial feedback tends to learn the user pattern and adapts to it quickly. Dueling Bandit exploration gives this work an edge as it performs adaptive exploration with lesser number of parameters requiring manual tuning. This leads to relevant item suggestions to users without cumbersome retraining. For future work, various other clustering methods could be experimented with for item space reduction. Ensemble of exploration strategies can also be tried.

References

1. D. Cortes. Adapting multi-armed bandits policies to contextual bandits scenarios. *arXiv preprint arXiv:1811.04383*, 2018.
2. A. S. Das, M. Datar, A. Garg, and S. Rajaram. Google news personalization: scalable online collaborative filtering. In *Proceedings of the 16th international conference on World Wide Web*, pages 271–280, 2007.

3. G. De Francisci Morales, A. Gionis, and C. Lucchese. From chatter to headlines: harnessing the real-time web for personalized news recommendation. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 153–162, 2012.
4. A. Grotov and M. de Rijke. Online learning to rank for information retrieval: Sigir 2016 tutorial. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 1215–1218, 2016.
5. F. M. Harper and J. A. Konstan. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4):1–19, 2015.
6. W. IJntema, F. Goossen, F. Frasincar, and F. Hogenboom. Ontology-based news recommendation. In *Proceedings of the 2010 EDBT/ICDT Workshops*, pages 1–6, 2010.
7. Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.
8. L. Li, W. Chu, J. Langford, and R. E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670, 2010.
9. J. Liu, P. Dolan, and E. R. Pedersen. Personalized news recommendation based on click behavior. In *Proceedings of the 15th international conference on Intelligent user interfaces*, pages 31–40, 2010.
10. Z. Lu and Q. Yang. Partially observable markov decision process for recommender systems. *arXiv preprint arXiv:1608.07793*, 2016.
11. T. Mahmood and F. Ricci. Improving recommender systems with adaptive conversational strategies. In *Proceedings of the 20th ACM conference on Hypertext and hypermedia*, pages 73–82, 2009.
12. B. Marlin and R. S. Zemel. The multiple multiplicative factor model for collaborative filtering. In *Proceedings of the twenty-first international conference on Machine learning*, page 73, 2004.
13. R. J. Mooney and L. Roy. Content-based book recommending using learning for text categorization. In *Proceedings of the fifth ACM conference on Digital libraries*, pages 195–204, 2000.
14. P. Rojanavasu, P. Srinil, and O. Pinngern. New recommendation system using reinforcement learning. *Special Issue of the Intl. J. Computer, the Internet and Management*, 13(SP 3), 2005.
15. N. Taghipour and A. Kardan. A hybrid web recommender system based on q-learning. In *Proceedings of the 2008 ACM symposium on Applied computing*, pages 1164–1168, 2008.
16. N. Taghipour, A. Kardan, and S. S. Ghidary. Usage-based web recommendations: a reinforcement learning approach. In *Proceedings of the 2007 ACM conference on Recommender systems*, pages 113–120, 2007.
17. H. Wang, Q. Wu, and H. Wang. Learning hidden features for contextual bandits. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 1633–1642, 2016.
18. C. Zeng, Q. Wang, S. Mokhtari, and T. Li. Online context-aware recommendation with time varying multi-armed bandit. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 2025–2034, 2016.
19. S. Zhang, L. Yao, A. Sun, and Y. Tay. Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys (CSUR)*, 52(1):1–38, 2019.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:11/November -2020 Impact Factor- 5.354 www.irjmets.com

CONCENTRATED SOLAR POWER FOR MID-DAY MEAL PROGRAM IN DELHI

Rajat Malik[*1], Raviratna Subir[*2], Rishabh Goswami[*3], J.P. Kesari[*4]

[*1,2,3]B.Tech Student(4th Year), Department of Mechanical Engineering, Delhi Technological University, Delhi-110042, India.

[*4]Associate Professor, Department of Mechanical Engineering, Delhi Technological University, Delhi –110042, India.

ABSTRACT

Solar energy is the future for clean energy which must be implemented for sustainable development. The world is bombarded by solar energy everyday in large quantities. But, only some part of it can be utilised by us and converted to useful energy. In this paper we considered Concentrated Solar Power generated cooking system for the midday meal system in Delhi. This paper also include case studies of 5 systems which use CSP system for various activities.

Keywords: Concentrated Solar Power(CSP), Collector, Scheffler Dish, DNI, Analysis.

I. INTRODUCTION

Solar energy is one of the major ways to generate clean power. The Sun provides us enough energy to run the whole world for 1 year, in just 1 hour. According to the US department of Energy Each hour 430 quintillion Joules of energy from the sun hits the Earth. That's 430 with 18 zeroes after it!

In comparison, the total amount of energy that all humans use in a year is 410 quintillion Joules. But the problem is the conversion of that energy into useful energy. There are many different ways. We decided to study CSP for this work. It generates solar power by using mirrors or lenses to focus a huge area of sunlight onto a receiver. Electricity is produced when the concentrated light is transformed to heat (solar thermal energy), which drives a heat engine (steam turbine) connected to an electrical power generator or powers a thermochemical reaction. After using reflectors to generate heat one can use that for areas other than electricity like dry cleaning, cooking etc.

II. LITERATURE REVIEW

In Alalewi, Ahmad 2014[1] Concentrated Solar power they discussed about different efficient ways of using solar power like parabolic trough, and tower technology.

This work described the four main types of concentrating solar collectors and discussed about the innovation which occurred in each over the last decade. Improvements in the related fields of concentrated photovoltaics and thermoelectrics was also presented, along with advances made in thermal energy storage methods, energy conversion cycles and CSP applications.[2]

The paper presents an analysis of the technical potential of concentrating solar energy (CSP) on a worldwide scale. The paper describes the assessment methodology and the technical and economic CSP model, and shows the results of this analysis for the different world regions.[3]

[5] several emissions study and economic considerations of csp was discussed.

The importance and potential of solar concentration in India was discussed. Also, various types of concentrators were studied.[6]

[8] In this review paper the authors discussed the theoretical framework of CSP technology with PTC(Parabolic Trough Collector) system, for purpose of understanding it. And recent developments on heat transfer enhancement methods for improvement of CSP technology with PTC system were discussed. As the rupture of glass covers was frequently observed during application, methods of thermal deformation restrain for tube receivers were reviewed also.

[9] This paper described the CSP technologies and STC (Solar Tower Collector), and it describes the importance and use of the predictions of solar irradiations in simulating the required plant configuration of a STC, to achieve optimum results.

[10] In this paper, the authors gives a review on the current thermal energy storage systems installed in CSP plants. New concepts for TES integration also are proposed, especially coupled technology for higher operating temperature and cascade TES of modularized storage units for intelligent temperature control.

According to the literature, concentrated solar power systems (CSP) have been successfully employed for bioactive compounds (e.g. essential oils) extraction, drying process, water desalination and decontamination, baking, and cooking.

Further considerations for improving the performance and applicability of CSP in the food industry are also discussed in the present study[11]

This paper reviews the current understanding of the concerns about climate change and the role being played by fossil fuel use. [12]

The energy produced by thermal solar plants doesn't need to be limited solely to hours of sunlight. it's possible to conceive a storage system and it's possible to increase the assembly of warmth beyond the hours of full sunshine. The use of molten salt offers the potential to scale back electricity cost and to extend the energy performance in an eco-compatible way. Salt is a smaller amount expensive and more environmentally benign than currently used HTFs but unfortunately the high melting point results in significant O&M challenges and requires an innovative freeze protection system.[15]

This review provides an summary of the status of state-of-the-art hybrid CSP-desalination systems. Published studies on hybrid plants under two categories: only freshwater generation and cogenerating of power and freshwater were reviewed in this paper.[16]

This paper reviews central receiver designs for concentrating solar energy applications with high-temperature power cycles.[17]

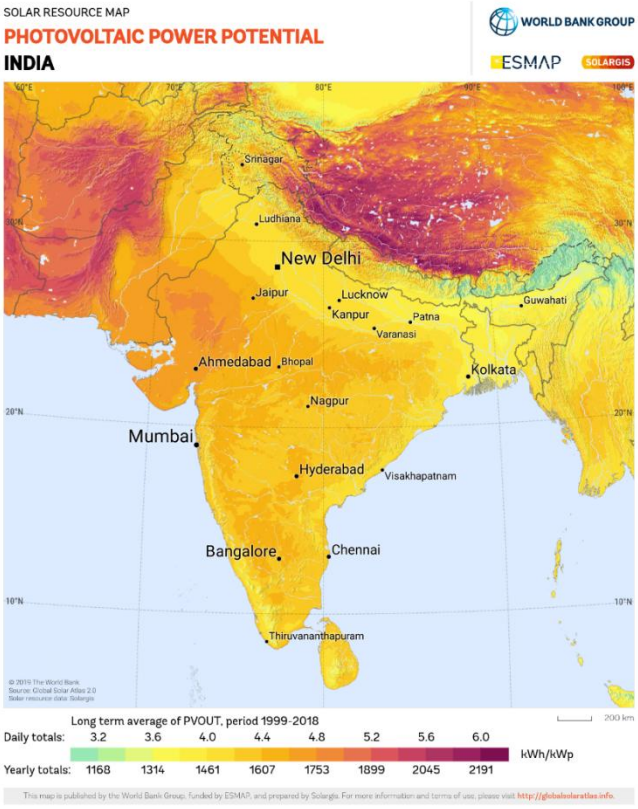


Figure-1: Solar Irradiance Map of INDIA

(https://solargis.com/maps-and-gis-data/download/india)

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:11/November -2020 Impact Factor- 5.354 www.irjmets.com

III. CASE STUDIES

1. Tirumala Tirupati Devasthanam kitchen at Tirupati:

Tirumala Tirupati Devasthanam is one among the foremost reverted pilgrimage place of India situated at Venkatachalam in Andhra Pradesh . The shrine is visited by 70,000 pilgrims on a mean day. Given the high amount of individuals visiting and staying at the shrine, the wants for food and power is high. The temple has thus installed solar powered cooking system. The installation of solar concentrators was accomplished in September 2002 and was launched in the subsequent month. The system requires only 10 kg per sq solar concentrating solar steam cooking system. It makes use of 73 parabolic concentrators and lift the water temperature to 550°C to 650°C. This technique is integrated with the prevailing boiler to make sure that cooking continues even at night and during rain and cloudy weather. The main aim for this solar instalment was to lessen the usage of LPG upto 50%. It therefore helps in saving $7000 a month. The system produces 3600 kg steam each day and thus, helps in saving 100,000 kg of cooking gas at an annual basis. It's cost is estimated to be around 1.3 crore. Wherein, Ministry of Non-renewable energy provided Rs. 58 lakhs as subsidy. It produces 4000kgs of steam per day at 453 Kelvin. The water is transformed into steam with the assistance of 106 solar dishes. The fuel saved (specifically diesel) accounts to 1.2 liters annually . 50,000 kilo grams of rice with Sambhar and Rasam are cooked each passing day in a very unconventional way that's with the use of of steam which reaches the kitchen through pipes. Chefs also find it more convenient because it is a way that is quiet efficient and time saving as compared to the conventional or traditional style of cooking.

The travellers visiting Tirumala are from various states, places, dialects, societies, status. As indicated by figures from Tirumala, just about 80,000 devotees visit the temple on typical days while the tally goes past a lakh during the Brahmotsavam. The Tirumala is arranged at most noteworthy 2,980 feet over the ocean level and appreciates salubrious and empowering climate. Tirumala Tirupati Devasthanam has the world's biggest sun based steam cooking framework utilizing the Scheffler Community Kitchen plan. The framework is constituted of 106 housetop mounted parabolic concentrators that produce steam for preparing up to 30,000 meals per day.



Figure-2: Solar Heaters at Tirumala Temple

(http://dilipkumar.in/articles/tirupati/tirumala-temple-going-greener-way-uses-solar-and-wind-power.html?archive=%3D%2F)

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:11/November -2020 Impact Factor- 5.354 www.irjmets.com



Figure-3: Community Kitchen at Tirupati

(https://www.vina.cc/2012/02/03/tirupati-temple-goes-green/)

2. ACME SOLAR PLANT:

Located in Bikaner, Rajasthan ACME CSP Power Plant of 2.5 MW is present. It is based on Concentration Tower technology. The construction of the project started in 2009 and became operational in year 2011. Technical Characteristics of the plant are mentioned below. The 46m high solar tower gets solar radiation from 14,280 heliostats with each heliostat aperture area of 1.136 m^2 which makes solar field aperture area to be approximately 16,222 m^2 (14,280 X1.136). This solar thermal plant is considered as the first demonstration project that India had after the launch of Jawaharlal Nehru National Solar Mission (phase-I bidding) on January 11, 2010.

The plant works on the steam Rankine cycle with cooling done via cooling tower. Its maximum cycle pressure reaches to 60 bar and temperature to 440°C. Following is the overview and technical characteristics of ACME Solar Tower Project, Bikaner:

Project: ACME Solar Tower

Location and Country: Bikaner, Rajasthan (India)

Lat/Long Location: 28°11′ 2.0″ North, 73°14′ 26.0″ East

Land Area: 12 hectares

Technology: Power tower

Turbine Capacity: 2.5 MW

Heat Transfer Fluid Type: Water/Steam

Cooling Method: Wet cooling (Cooling Tower)

Receiver Inlet Temperature: 218 °C

Receiver Outlet Temperature: 440°C

Power Cycle Pressure: 60 bar

Storage Type: None

Heliostat Solar-Field Aperture Area: 16,222 m^2

No. of Heliostats: 14,280

Heliostat Aperture Area: 1.136 m^2

Tower Height: 46 m

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:11/November -2020 Impact Factor- 5.354 www.irjmets.com



Figure-4: Solar Tower Bikaner

(http://solarthermalpitsugoko.blogspot.com/2017/01/acme-solar-thermal-tower.html)

3. JULICH SOLAR TOWER

CSP plant in Julich, Germany is capable of producing 1.5 MW using Power Tower technology. It is the country's only plant that is based on solar power tower technology. Construction of the Julich Power Tower began in July 2007 and it was finally installed in December 2008. Heliostat field comprises a total of 2153 Heliostats. Each heliostat has an aperture area of 8.2 m^2 which finally makes a total of 17650 m^2 aperture area. On top of the 60 m tower, the power block steam turbine is combined with the generator that is capable of producing power in the form of electrical energy. Air is used as a HTF(Heat Transfer Fluid) which is heated to 680◦C and then is used to heat water up to 480◦C to be used in the power generation cycle. The efficiency of the plant isn't too great since the heat transfer coefficient of air is low. The plant also has a heat storage unit with a 1.5-hour capacity. Heat is stored by making hot air flow between a packing of ceramic elements which are hence, heated. Following is the overview and technical characteristics of Julich Solar Tower: -

Project: Saguaro Power Plant

Location and Country: Julich, Rhineland (Germany)

Lat/Long Location: 50°54′ 54.0″ North, 6°23′ 16.0″ East

Land Area: 17 hectares

Technology: Power Tower

Turbine Capacity: 1.5 MW

Heat Transfer Fluid Type: Air

Cooling Method: Dry cooling

Receiver Outlet Temperature: 680°C

Storage Type: Ceramic heat sink (Capacity-1.5hour)

Heliostat Solar-Field Aperture Area: 17,650 m^2

No. of Heliostats: 2,153

Heliostat Aperture Area: 8.2 m^2

Tower Height: 60 m



Figure-5: Bird's eye view

(https://www.dlr.de/sf/en/desktopdefault.aspx/tabid-8560/15527_read-44867/)



Figure-6: Close-up Image

(https://www.dlr.de/content/en/research-facilities/solar-tower-juelich.html)

4. COMMUNITY SOLAR COOKER:

Solar cooker is a device that uses direct sunlight, converting it into usable thermal energy as a means to heat, cook or pasteurize the food and drinks. Most solar cookers, are low-tech, inexpensive and less powerful. However, some of them are as powerful and expensive as traditional stoves. Advanced and large-scale solar cookers can cook for multitudes of people. Because of their low-cost and no fuel usage, many non-profit organizations have come to the forefront in order to promote their worldwide use so as to help in decreasing the air pollution and to reduce the deforestation and desertification given rise by gathering firewood for cooking. These cookers also lend a helping hand in the reduced fuel costs for low income people. There are varying kinds of solar cookers that exist which include parabolic solar cookers, solar ovens, and panel cookers, among others. The salient features of the cooker are as below:

1. Community cooking

2. Indoor Cooking
3. Fast Cooking
4. Cooking of traditional Food
5. Automatic tracking
6. Seasonal Adjustment
7. Multiple Use

Dish Solar Cooker: It is a parabolic dish solar cooker and its dimensions are 1.4 meter aperture diameter and 0.28 meter focal length. The contemplative material used for fabrication of this cooker is anodized aluminium sheet with reflectivity of more than 75%. Tracking of the cooker is manual , therefore it has to be adjusted in 15-20 minutes during cooking time. Nearly, 2 to 3 litres of water could be boiled in half an hour as it has delivering power of about 0.6 KW. The temperature raised at around at 350 to 400°C(sufficient for roasting, frying and boiling) at the bottom of the vessel. The needs of 10 to 15 people could be met as the cooker has a thermal efficiency of around 40% and can be used from one hour after sunrise to one hour before sunset on clear days. Dismantling and Assembling the cooker can be easily done by anybody. Therefore it can be packed nicely and transported anywhere in the country. It is user friendly, as the place of vessel that is to be kept for cooking is at a level which is convenient for the people to use. The cooker could be asset for people living in rural as well as urban areas and also for small businesses like dhabas, tea shops, etc.. It can save up to 10 LPG cylinders per year on full use at small establishments as the cost of the cooker is around Rs. 6000- 7000/-.

5. COMPOUND PARABOLIC CONCENTRATORS USED FOR INDUSTRIAL PROCESS HEATING IN KANDLA:

It is proven that the industrial sectors are dependent largely on conventional fuels to meet the requirement of their heating applications.

Compound Parabolic Concentrator (CPC) is one such proven, highly efficient, and economical solar thermal technology. For generating hot water up to 90°C–100°C, CPC is very favorable.

Need for CPC System:

Before being changed to solar technology the heating requirement at Hindustan Unilever's Kandla plant was met by using diesel. But they decided to switch over to solar technology for heating 20 kl of process water during one shift operation. By heating water from 30°C to 85°C, 11 lakh Kcal, is accounted for. To meet the heating requirements, VCare Global proposed a comprehensive solution, centered on Compound Parabolic Concentrator technology. The solution comprised of 160 collectors providing a total aperture area of 480 m^2. The system runs in primary loop consisting of cold water storage, cold water pumping to the CPC in series and parallel laid on three arrays and collecting the generated hot water in the hot water tank.



e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:11/November -2020 Impact Factor- 5.354 www.irjmets.com



Figure-7: CPC Collectors Installed at Hindustan Unilever, Kandla, Gujarat(undp.org)

After exploring various technologies for solar process heating including Flat Plate Collectors, Evacuated Tube Collectors, Parabolic dishes, Parabolic Trough Concentrators and Fresnel Linear Concentrator. CPC was finalised as it was the most suited alternative for such application. For any Solar Thermal Heating system Space may be a challenge. The Kandla project was no exception. However, this issue was mitigated by designing the layout of CPC panels in the best possible space minimizing manner. The challenge was installation of system at site as soil conditions had to be analysed to determine the type of civil foundation so that work could be decided. Another challenge was limited working time that was allowed at plant (only during daytime) and in fact being monsoon many days of site work was delayed as civil work and insulation work couldn't be undertaken during that period but still the timeline had to be maintained.

The CPC was designed to offer 11 lakh kcal per shift operation on an average basis. To achieve the specified temperature of 85°C on the secondary side, it had been designed to boost the temperature of hot water up to 92°C within the primary circuit, using solar energy generated from CPC. The facility has started making benefits by using the Solar Thermal Hot Water System of saving in daily diesel consumption at about 127 ltr/day and also reduction in emission of carbon dioxide about 126 Ton carbon dioxide per annum.

IV. CALCULATIONS

Let us assume that around 40 scheffler dishes is installed by the foundation for the cooking purpose in their mega kitchens.

We will first calculate the amount of fuel saved by one reflector dish and then with the help of that we can calculate the net fuel saved by 40 dishes in one year time and finally At the end using the cost of scheffler dishes and cost of fuel saved we can calculate the payback period.

- No. of Scheffler dishes = 40

  New Delhi DNI= 5.5 Kwh/m^2/day
- Let the seasonal angle deviation of sun to be 5^0

  Taking one Scheffler dish,

Assuming number of sunny days in a year = 250

Approximate cooking hours = 8/day

Aperture area = 16 cos(43.23 + δ) where the δ is the seasonal angle deviation

Taking the value of δ as 5 degrees,

Hence, aperture area = 15cos(43.23 + 2.5)= 10.47m^2

Energy given by 1 Scheffler reflector = = 57.59 kWh/day

= 57.59 Kwh/day =207.324 MJ/day

The calorific value of LPG = 46.1 MJ/ Kg

Amount of LPG saved per day using 1 Scheffler reflector = 207.324/46.1 kg = 4.497 kg

Cost of 1 kg LPG = Rs. 41.83

Amount saved per day using 1 Scheffler reflector= 4.58 * 41.83

= Rs.191.58

Amount saved in 250 days= 250*191.58=47896.47

Amount saved in 250 days by 40 Scheffler dishes =

19,15,859 RS.

Cost of 1 Scheffler dish = Rs. 2,70,000

Cost of 40 Scheffler dish= Rs.1,08,00,000

Taking 12,00,000 as other costs included

Payback period = 6.26years

V. CONCLUSION

In this paper we did some case study in order to understand different types of CSP implemented in the world. We did a case study for each of the projects. In the end we gave theoretical model for cooking mid-day meal for children of Delhi via using a CSP model based on Scheffler dish concentrator. With a cost-efficient payback period as 6.26 year. Lastly, we would like to tell that due to India's geographical position and solar irradiation, implementing CSP in various fields in India will definitely help us move forward, and help future generations have clean environment.

VI. REFERENCE

[1] Alalewi, Ahmad. Concentrated Solar Power (CSP). Higher Institute for Applied Science and Technology, 2014.

[2] Barlev, David, Ruxandra Vidu, and Pieter Stroeve. "Innovation in concentrated solar power." Solar energy materials and solar cells 95.10 (2011): 2703-2725.

[3] Trieb, Franz, Christoph Schillings, Marlene O'sullivan, Thomas Pregger, and Carsten Hoyer-Klick. "Global potential of concentrating solar power." German Aerospace Centre (DLR) (2009).

[4] Adinberg, R. (2011). Simulation analysis of thermal storage for concentrating solar power. Applied Thermal Engineering, 31(16), 3588-3594.

[5] Vallentin, Daniel, and Peter Viebahn. "Economic opportunities resulting from a global deployment of concentrated solar power (CSP) technologies—the example of German technology providers." Energy Policy 38.8 (2010): 4467-4478.

[6] Ummadisingu, Amita, and M. S. Soni. "Concentrating solar power–technology, potential and policy in India." Renewable and sustainable energy reviews 15.9 (2011): 5169-5175.

[7] https://solargis.com/maps-and-gis-data/download/india

[8] Fuqiang, Wang, et al. "Progress in concentrated solar power technology with parabolic trough collector system: A comprehensive review." Renewable and Sustainable Energy Reviews 79 (2017): 1314-1328.

[9] Zhang, H. L., Baeyens, J., Degrève, J., & Cacères, G. (2013). Concentrated solar power plants: Review and design methodology. Renewable and sustainable energy reviews, 22, 466-481.

[10] Pelay, Ugo, Lingai Luo, Yilin Fan, Driss Stitou, and Mark Rood. "Thermal energy storage systems for concentrated solar power plants." Renewable and Sustainable Energy Reviews 79 (2017): 82-100.

[11] Barba, Francisco J., et al. "Solar radiation as a prospective energy source for green and economic processes in the food industry: From waste biomass valorization to dehydration, cooking, and baking." Journal of cleaner production 220 (2019): 1121-1130.

[12] Wuebbles DJ, Jain AK. Concerns about climate change and the role of fossil fuel use. Fuel Processing Technology 2001; 71:99-119.

[13] Forrester, J. (2014). The value of CSP with thermal energy storage in providing grid stability. Energy Procedia, 49, 1632-1641.

[14] Islam, Md Tasbirul, et al. "A comprehensive review of state-of-the-art concentrating solar power (CSP) technologies: Current status and research trends." Renewable and Sustainable Energy Reviews 91 (2018): 987-1018.

[15] Cavallaro, Fausto. "Fuzzy TOPSIS approach for assessing thermal-energy storage in concentrated solar power (CSP) systems." Applied Energy 87, no. 2 (2010): 496-503.

[16] Mohammadi, Kasra, Mohammad Saghafifar, Kevin Ellingwood, and Kody Powell. "Hybrid concentrated solar power (CSP)-desalination systems: A review." Desalination 468 (2019): 114083.

[17] Ho, Clifford K., and Brian D. Iverson. "Review of high-temperature central receiver designs for concentrating solar power." Renewable and Sustainable Energy Reviews 29 (2014): 835-846.

[18] Avila-Marin, Antonio L., Jesus Fernandez-Reche, and Felix M. Tellez. "Evaluation of the potential of central receiver solar power plants: configuration, optimization and trends." Applied energy 112 (2013): 274-288.

[19] López-Herraiz, María, Azucena Bello Fernández, Noelia Martinez, and Manuel Gallas. "Effect of the optical properties of the coating of a concentrated solar power central receiver on its thermal efficiency." Solar Energy Materials and Solar Cells 159 (2017): 66-72.

[20] Ho, Clifford K. "A review of high-temperature particle receivers for concentrating solar power." Applied Thermal Engineering 109 (2016): 958-969.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/346714622

Crack Detection of Structures using Deep Learning Framework

Conference Paper · December 2020

CITATIONS
0

READS
96

4 authors, including:



Avinash Kumar Jha
Delhi Technological University
5 PUBLICATIONS **1** CITATION
SEE PROFILE



Ashutosh Trivedi
Delhi Technological University
48 PUBLICATIONS **284** CITATIONS
SEE PROFILE

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

dynamically loaded retaining wall View project

2D and 3D Consolidation View project

All content following this page was uploaded by Avinash Kumar Jha on 08 December 2020.

The user has requested enhancement of the downloaded file.

Proceedings of the Third International Conference on Intelligent Sustainable Systems [ICISS 2020]
DVD Part Number: CFP20M19-DVD; ISBN: 978-1-7281-7088-6

Crack Detection of Structures using Deep Learning Framework

Arunish Kumar[1], Amit Kumar[1], Avinash Kumar Jha[1], Ashutosh Trivedi[1]

[1]Department of Civil Engineering, Delhi Technological University, India
{kumararunish7, amitvijaranyia, avinash.jha.6696}gmail.com, atrivedi@dce.ac.in

ABSTRACT

The transportation system has drawn attention due to the growing threats of cracks to the roads and bridges. To accurately detect cracks was made possible by the introduction of deep learning. In this paper, we propose the modified LeNet 5 model for detecting cracks in roads and bridges. We summarize our work in the following points: - 1. We worked on the three datasets, i.e., Automated Bridge Crack Detection Dataset, Concrete Crack Images for Classification Dataset, and Asphalt Crack Dataset. 2. We applied a modified version of LeNet-5 model on Automated Bridge Crack Detection Dataset, Concrete Crack Images for Classification Dataset, and Asphalt Crack Dataset. 3. Then, we analyzed our results and compared the same with Principal Component Analysis and without using Principal Component Analysis. We are also highlighting the region of crack and non-crack using green and red color, respectively. Our proposed model takes both the factors of time and accuracy into consideration while producing the output.

Keywords— modified LeNet, principal component analysis (PCA), deep learning, crack detection

1. INTRODUCTION

Machine learning has been used widely these days to solve various real world problems [19]. From solving vast engineering problems to finance and healthcare, machine learning is used everywhere [20-22]. In this paper, the practical application of deep learning for crack detection has been discussed.

1.1 Inspiration

Cracks are common distresses in concrete,rock masses[15], asphalt as well as bridges [16]. Road surface, aging, climate are the various reasons for the occurrence of different types of cracks [7]. Asphalt cracks are mainly due to stress and friction. Friction value increases with the increasing value of coefficient of friction [17]. The cracks can reduce the efficiency of materials. The water, by penetrating the deeper layers of the materials through cracks, decreases the compaction of the material, which results in the decrease in load-bearing capacity of the whole structure. Hence it is obvious to maintain the structures from time to time. If any cracks occur, it should be first detected and acted upon suitably to remove it.

1.2 Crack Detection Algorithm

The detection of cracks can be performed by both visual inspection and automatic survey. Although both methods present good distress analysis results, the automatic crack detection system is quicker, lower costing than traditional human vision detection [15]. Therefore, automatic crack detection has attracted much attention from scientific and technical corporations in recent years [7]. For the crack detection purpose, we have modified the LeNet-5 model, and used it. We have also applied Principal Component Analysis, while predicting the cracks on the various materials that we have used, to reduce model training time without compromising much with the accuracy. Trained model can also be used to locate the cracks in an image by highlighting it with red color. In the end, we are comparing the results that we have got for the materials with using Principal Component Analysis and without using Principal Component Analysis.

2. LITERATURE REVIEW

Eisenbach et al. [1] have examined different assessment strategies applied on the pavement surface and have sorted them into three gatherings: - crack image thresholding, patch-based classification, and depth-based algorithms. Their paper's shortcomings were that the pictures captured by them were distinct in 64*64 size, which does not give much information. Yokoyama [2] built up the identifier that consequently identifies cracks from the photos of solid surfaces utilizing CNN. However, for the stained concrete, the detection rate was meager. Cha et al. [3] have utilized CNN classifiers that are less affected by clamor brought about by lighting, shadow, projecting to identify cracks from images and different sorts of auxiliary harms like voids, spalling, and corrosion of cement in steel structures. These techniques probably will not have the option to treat the non-homogenous solid surface concerning color and texture. Pauly et al. [4] illustrated the potential of deep learning by using deeper networks for the pavement crack detection.

However, when the training and testing images are taken from different areas, the network does not perform well. Hoang [5] has preprocessed the Otsu method's image threshold, which is an intensity adjustment method called

Min-Max Gray Level Discrimination (M2GLD). This strategy's downside was that the clients need to fine-tune boundaries changing proportion (RA) and edge boundary (τ). By utilizing this examination, it is discovered that RA = 2 and τ = 0.5 can convey acceptable crack identification on the building surface. Be that as it may, for various kinds of pictures, such as black-top or Asphalt, these two qualities should be balanced. Zheng et al. [6], to boost and extract the data features, they have suggested FCN, RCNN, and RFCN frameworks. Although the accuracy is improved with the help of RCNN and FCN but the framework takes more time to train. Fan et al. [7] proposed U Hierarchical Dilated Network (U-HDN) to detect cracks. U Hierarchical Network uses the architecture of encoder-decoder along with hierarchical feature learning. However, U Hierarchical Network although produces acceptable results, but due to the neural network's complicated structure, the computational cost is high, and the efficiency is low. Li et al. [8] concocted a technique for distinguishing the surface breaks for stacked coal with the help of a vibration failure process, which is based on a vibration failure system and SVM. However, they did not use Principal Component Analysis to analyze further which feature of the crack plays a major role so that they can reduce the training time. Chen et al. [9] detected cracks for the submerged metallic plane from nuclear inspection videos with the help of a deep learning model called NB-CNN. However, they did not use Principal Component Analysis to further reduce their training time. Zhou et al. [10] analyzed three-dimensional (3D) laser image range data with an automated crack detection technique, which uses frequency domain filtering and contouring analysis. However, the above-proposed method fails in case of detecting shallow cracks. Lin et al. [11], with the help of a sandstone specimen under three-point bending, determined the cohesive zone's characteristics and cohesionless crack's length. However, the primary outcomes need additional examinations because the results of some numerical modeling get challenged. Zhang et al. [12] compared the Handcrafts method's crack detection performance with the deep learning model and concluded that the latter performed better. The dataset that they worked upon was small, which consisted of just 500 images.

3. METHODOLOGY AND EXPERIMENTATION

3.1 Datasets

3.1.1 Automated Bridge Crack Detection Dataset

The original dataset comprises of 2068 Bridge Crack Images, collected by the Phantom 4 Pro's CMOS surface array camera, with a resolution of 1024*1024 [14].

3.1.2 Concrete Crack Images for Crack Classification Dataset

The dataset possesses images of concrete. The dataset has two categories, i.e., positive and negative. The positive category contains cracked images, and the negative has non-cracked images of concrete. The dataset contains 40,000 images of concrete of pixel size 227*227 [18].

3.1.3 Asphalt Crack Dataset

This dataset consists of 400 images of Asphalt cracks and Non-Cracks images, that are captured using Nikon camera, and can be used for training crack detection models [13].

3.2 Deep Learning Approach

3.2.1 Modified LeNet-5

Following Table 1 depicts the architecture of the modified version of LeNet-5 model when applied on Concrete Crack Images for Classification Dataset. Input shape to the first layer of the model is of size (*, 227, 227, 3), here * represents the total number of images used while training the model.

TABLE I. ARCHITECTURE OF MODIFIED LE-NET5

Layer (Type)	$\frac{patch\ size}{stride}$	**Output Shape**	**Param #**
Conv2D	(5 × 5)/1	(*, 223, 223, 6)	456
Average Pooling2D	(2 × 2)/2	(*, 111, 111, 6)	0
Conv2D	(5 × 5)/1	(*, 107, 107, 16)	2416
Average Pooling2D	(2 × 2)/2	(*, 53, 53, 16)	0
Flatten	_	(*, 44944)	0
Dense	_	(*, 120)	5393400
Dense	_	(*, 84)	10164
Dense	_	(*, 1)	85
Total Parameters #		5,406,521 (5.4 M)	
Trainable Parameters #		5,406,521 (5.4 M)	
Non-Trainable Parameters #		0	

3.2.1.1 Convolutional Layer

The convolutional layers are the primary building blocks of a convolutional neural network structure. It is the convolutional layers that learn the features that are suitable for differentiating between a crack and non-crack image. Initial convolutional layers learn the local features required for this, whereas the deeper layers learn from the global features required for differentiating cracks and non-cracks [4]. The stride is the additional hyperparameter of the layer. The number of receptive field columns and rows covering the input width and height is defined by stride. Fewer receptive field applications and smaller output sizes are due to larger stride, which reduces computational cost at the price of loss of some input data features.

3.2.1.2 Pooling Layers

The pooling layer reduces the number of parameters and computation time by down-sampling the representation. A pooling layer usually takes a patch of adjacent pixels and tries to reduce it to a single pixel. A pooling layer is generally included between two consecutive convolutional layers. There are two types of pooling layers, Average Pooling layer and Max Pooling Layer. We have used average pooling layer between two successive convolutional layers.

3.2.1.3 Activation Layer

The typical Artificial Neural Network is given non-linearity with the help of sigmoidal functions. However, Nair and Hinton stated that the computations are slowed down by saturation of nonlinearities. A non-linear function called RELU was introduced recently [3]. The example of a nonlinear function is depicted in Figure 1. Apart from its negative input values, the RELU gives no bounded output than other nonlinear functions, generating bounded output values. RELU's gradients are always zeros and ones. Due to these factors, RELU has faster calculations and better accuracy than sigmoidal functions. We have used the Sigmoidal function in the output layer and in all other layers, RELU function is used. The formula for RELU function is:

$$R(z) = \begin{cases} z & z > 0 \\ 0 & z <= 0 \end{cases} \quad (1)$$

The formula for Sigmoid function is:

$$\boldsymbol{R(z)} = \frac{\mathbf{1}}{\mathbf{1} + e^{-z}} \quad (2)$$

3.2.1.4 Loss function & Optimisation function

Loss functions provide a quantity that is sought to be minimised during the training phase of any model. This minimisation is done with the help of some optimization function. We have used BinaryCrossEntropy as the loss function and AdamOptimiser as the optimiser. The formula of BinaryCrossEntropy loss function:

$$CE = -\sum_{i=1}^{C'=2} t_i log(s_i) = -t_1 log(s_1) - (1 - t_1) log(1 - s_1) \quad (3)$$

where, CE is Binary Cross Entropy, C1 and C2 are are the two classes[0, 1](in our case, it is crack or non-crack), t1 and s1 are groundtruth and score for C1 class respectively and t2(1-t1) and s2(1-s2) are groundtruth and score for C2 class respectively.

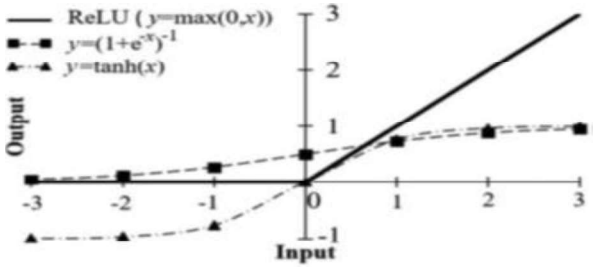


Figure 1. Non-Linear Activation function

3.3 Machine Learning Approach

3.3.1 Principal Component Analysis

Principal Component Analysis helps us to reduce dimensionality. If there are vast numbers of features, there will be a high chance that they are co-related. Here we reduce the data, and also we want to keep similar kinds of information with us. In Principal Component Analysis, if we have a considerable number of features, we try to reduce it into fewer features that would also explain our data very well. Although we might be losing some amount of information while reducing the dimensions, the benefits of using it eclipses the disadvantages to a great extent. The advantages of using Principal Component analysis are lesser training time, lesser memory requirement, and easy to visualize. Before using Principal Component Analysis, usually feature scaling is applied.

3.3.2 Math behind Principal Component Analysis

Principal Component Analysis requires the directions in which the data changes the maximum, i.e., the variance. The Principle Component is found out using the covariance matrix, which will tell how a particular feature varies with each and every feature. Covariance matrix will be symmetrical as covariance (f1, f2) = covariance (f2, f1). Once the covariance matrix is calculated, the next step is to find the eigenvalues and eigenvectors. Main eigenvectors are defined as the direction in which the variance is maximum. Eigenvectors and Eigenvalues will be founded using Singular Value Decomposition. Eigenvectors will be a unit vector. The

first eigenvector will give first max variance, and the second eigenvector will give second max and so on. Eigenvalues will represent how much variance will be explained by a certain dimension. In case the eigenvalue of a certain direction is very less, which means it has very less variance or information hence it can be removed.

3.4 Training & Testing of the Model

We have applied modified LeNet-5 on the Automated Bridge Crack Detection dataset, Concrete Crack Images Classification dataset, and Asphalt Crack dataset. Figure 2 shows the Detection of Concrete Cracks Flowchart. The idea of using modified LeNet-5 was to check the performance of LeNet-5 by detecting the cracks on different types of materials. We wanted to make sure that modified LeNet-5 produces consistent results for all types of materials that we have chosen to work upon. The modified LeNet -5 was implemented using the Keras framework.

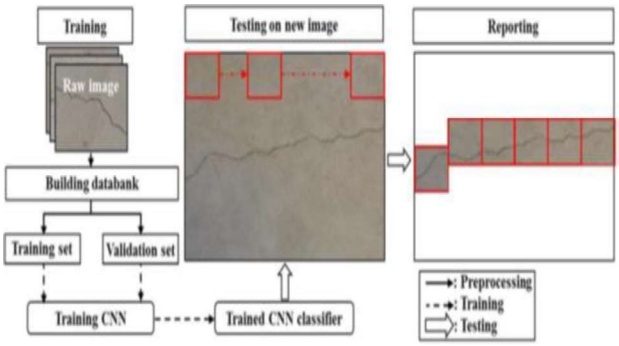


Figure 2. Detection Concrete Cracks Flowchart

For the Concrete Crack Images for Classification dataset, the modified LeNet -5 was trained on different sets of training data with or without Principle Component Analysis. We trained the modified LeNet-5 models with 8000, 16000, 24000, 32000 and 40000 training sets of images. We are comparing the accuracy and time, first by applying Principal Component Analysis before modified LeNet-5, second by using modified LeNet-5 alone, for all the datasets we have used. We have used Stochastic Batch Gradient Descent for training all the datasets with the batch size of 32. For the Concrete Crack Images Dataset, the min delta's value was chosen as 0.01, and the patience value was taken as 4. The Concrete Crack Images dataset was tested on ten epochs. Simultaneously, the Automated Bridge Crack Detection Dataset was also given the same value of min delta, patience, and epochs as the Concrete Crack Images Dataset. Whereas for the Asphalt Dataset, the training was done without using the early stopping function, with the epochs value 30, and it was observed that the model was giving maximum training and testing accuracy at 29th epoch when Principal Component Analysis was applied and at 10th epoch when Principal Component Analysis was not applied. To overcome overfitting, the orientation of every image is changed before passing it to the model, so that the model never sees the same image again, for all the datasets that we have used. Figure 3,Figure 4, Figure 5 shows predictiction of model on Concrete Crack Images for Classification Dataset, Asphalt Crack Dataset and Automated Bridge Crack Detection Dataset respectively, without using Principal Component Analysis. Whereas Figure 6, Figure 7, Figure 8 shows the prediction of the model on Concrete Crack Images for Classification Dataset, Asphalt Crack Dataset and Automated Bridge Crack Detection Dataset respectively, by using Principal Component Analysis.

3.5 Predictions on Images

3.5.1 Predictions on images without using Principal Component Analysis

Figure 3 shows the predictions of the model on Concrete Crack Images for Classification dataset without using Principal Component Analysis. Figure 4 shows the predictions of the model on Asphalt Crack dataset without using Principal Component Analysis. Figure 5 shows the predictions of the model on Automated Bridge Crack Detection dataset without using Principal Component Analysis.

3.5.2 Predictions on images using Principal Component Analysis

Figure 6 shows the predictions of the model on Concrete Crack Images for Classification dataset using Principal Component Analysis. Figure 7 shows the predictions of the model on Asphalt Crack dataset using Principal Component Analysis. Figure 8 shows the predictions of the model on Automated Bridge Crack Detection dataset using Principal Component Analysis.

3.6 Locating the cracks in the images

We first loaded the image of some structure, which had cracks. Then we resized the image to 1200*1200 pixels. After reshaping the image, we divided the image into 144 small images, each of size 100*100 pixels. These small images then were passed through our trained model for prediction, which we have implemented from scratch using Keras to detect the crack and non-crack region. If the model is predicting cracks for the image passed to it, then we are coloring the region as red; otherwise, it has been colored with green color. Then these all 144 small images were recombined to make 1 full image. Figure 9 below represents the raw and crack highlighted image.

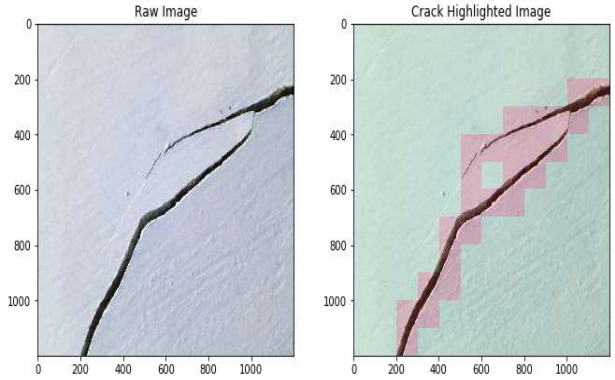


Figure 9. Raw images and crack highlighted image

4. RESULTS AND DISCUSSION

4.1 Metrics used for judging the performance of the model

4.1.1 Precision

It calculates the precision of the predictions with respect to the labels. The formula for Precision is:

$$Precision(P) = \frac{TP(true\ positives)}{TP(true\ positives) + FP(false\ positives)} \quad (4)$$

4.1.2 Recall

It calculates the recall of the predictions with respect to the labels. The formula for Recall is:

$$Recall(R) = \frac{TP(true\ positives)}{TP(true\ positives) + FN(false\ negatives)} \quad (5)$$

4.1.3 Accuracy

It calculates how often the predictions are correct. The formula for Accuracy is:

$$Accuracy(A) = \frac{TP + TN}{TP + FP + TN + FN} \quad (6)$$

TP, TN, FP, and FN are True Positive, True Negative, False Positive, and False Negative respectively. In our case, Positive is when the image doesn't have cracks and Negative is when the image does have cracks. True Positive is when the image doesn't have cracks and the model predicted it as non-crack. True Negative is when the image does have cracks and the model also predicted it has cracks. False Positive is when the image doesn't have cracks but the model predicted that it has cracks. False Negative is when the image does have cracks but the model predicted that it has no-cracks. True Positive Prediction, True Negative Prediction, False Positive Prediction, False Negative Prediction is shown by Figure 10,11,12 13 respectively.

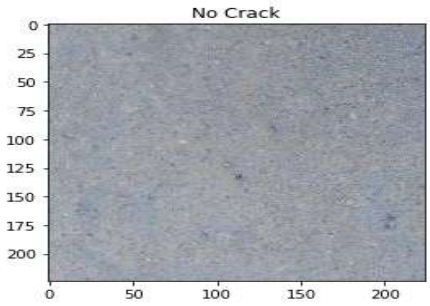


Figure 10. True Positive Prediction

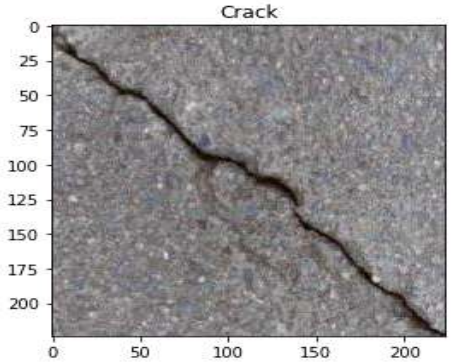


Figure 11. True Negative Prediction

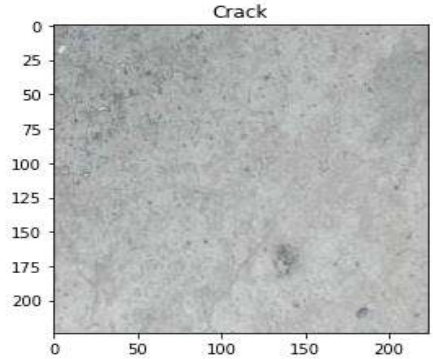


Figure 12. False Positive Prediction

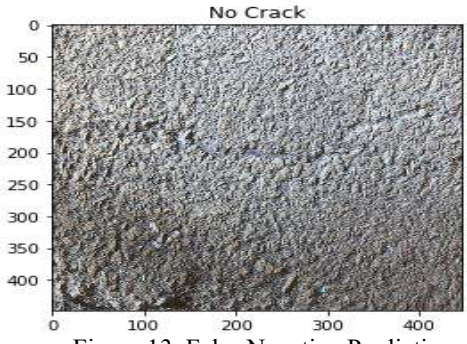


Figure 13. False Negative Prediction

4.2 Training Time for judging the performance of the model

It is the time taken by the model to learn all the parameters. We have recorded the time taken by the model to train for all the datasets which are discussed in section 3.1. Figure 14, 15, 16 and 17 show the accuracy, precision, recall and training timecurve for Concrete Crack Images for Classification dataset, with and without using PCA. Figure 18 and 19 show the performance of the model, based on various parameters (as discussed in section 4.1 & 4.2), for Asphalt Crack Dataset and Auttomated Bridge Crack Dataset.

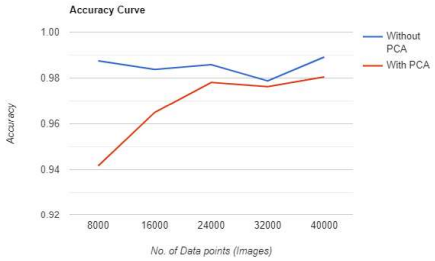


Figure 14. Accuracy Curve

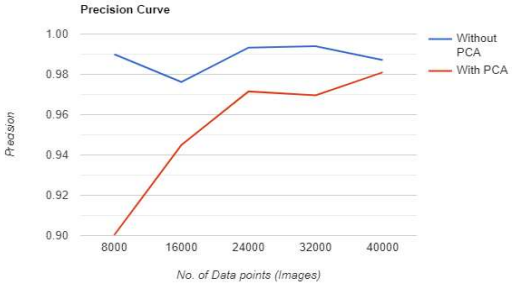


Figure 15. Precision Curve

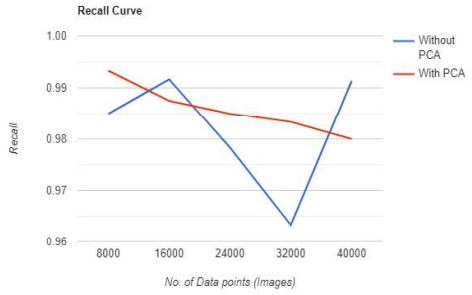


Figure 16. Recall Curve

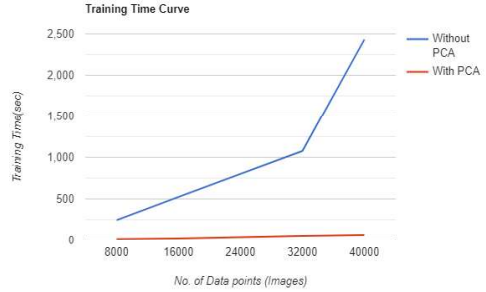


Figure 17. Training Time Curve

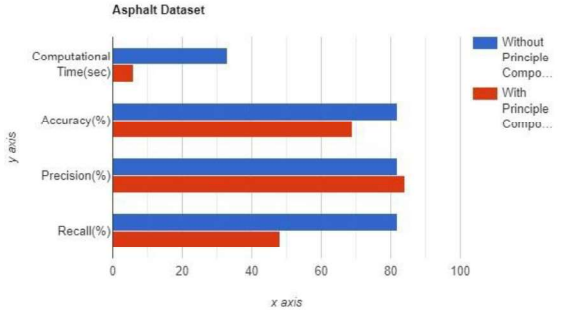


Figure 18. Performance on Asphalt Crack dataset

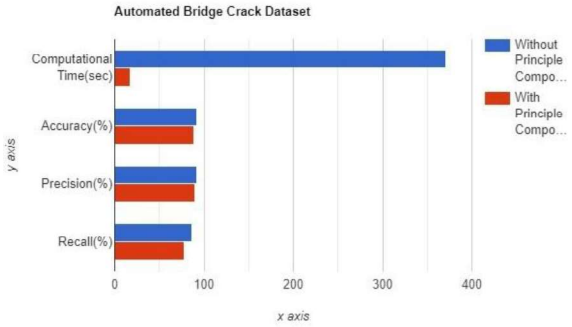


Figure 19. Performance on Automated Crack Detection Data

5. CONCLUSION AND FUTURE SCOPE

Our modified LeNet model was performing well on Automated Bridge Crack Detection dataset and Concrete Crack Images for Classification Dataset. While for the Asphalt Crack dataset, it was performing average. The PCA applied on these datasets has enabled us to get our datasets trained in a shorter time without any considerable loss in accuracy for the Concrete Crack Images for Classification dataset and Automated Bridge Crack Detection dataset. Although the Principal Component Analysis model reduces training time for the Asphalt Crack dataset, we are also losing a substantial amount of accuracy. Our modified LeNeT model was only predicting whether there is a crack or not in materials but it is not giving any further information about the type of cracks or causes of the occurrence of cracks, which contributes to our paper's limitation. Another limitation is that the LeNet model's performance was average in predicting cracks in Asphalt. In our subsequent work, we will try to eradicate these shortcomings of our paper.

6. ACKNOWLEDGEMENTS

The authors would like to thank the Department of Civil Engineering, Delhi Technological University (DTU) and Central Library, DTU for providing access to all the resources used for our study. The financial support received under the research project grant (F. No. DTU/IRD/619/2105) of DTU, New Delhi is thankfully acknowledged

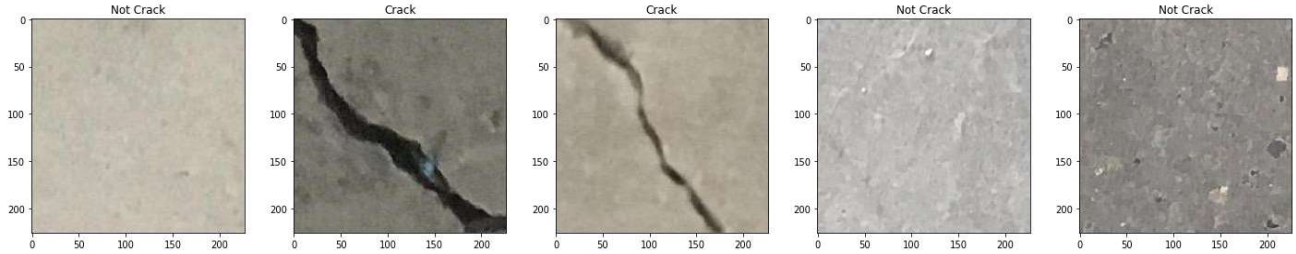


Figure 3. Predictions on Crack images for Classification dataset without using PCA

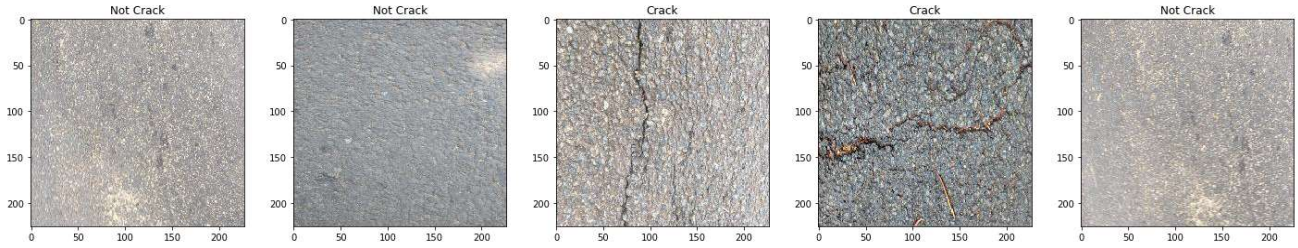


Figure 4. Predictions on Asphalt Crack dataset without using PCA

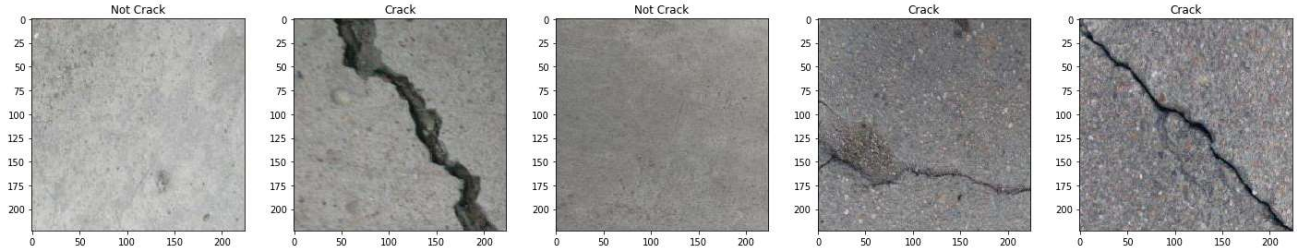


Figure 5. Predictions on Automated Crack Detection dataset without using PCA

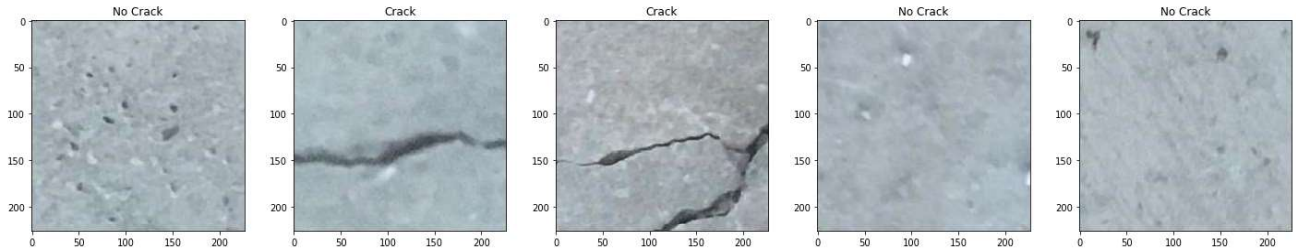


Figure 6. Predictions on Crack images for Classification dataset using PCA

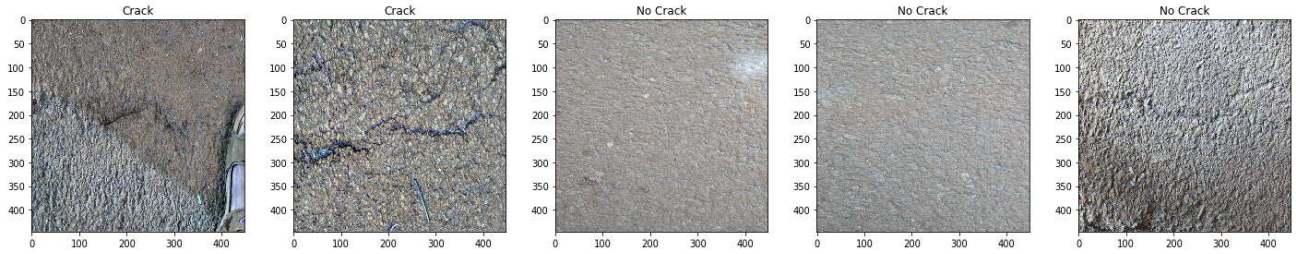


Figure 7. Predictions on Asphalt Crack dataset using PCA

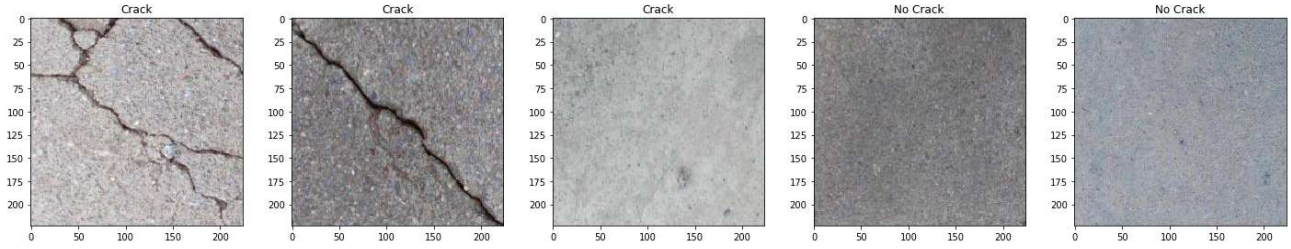


Figure 8. Predictions on Automated Crack Detection dataset using PCA

REFERENCES

[1] Eisenbach, M., Stricker, R., Seichter, D., Amende, K., Debes, K., Sesselmann, M., Ebersbach, D., Stoeckert, U., and Gross, H. (2017) "How to get pavement distress detection ready for deep learning? A systematic approach" 2017 International Joint Conference on Neural Networks (IJCNN)

[2] Yokoyama, S., and Matsumoto, T. (2017) "Development of an Automatic Detector of Cracks in Concrete Using Machine Learning". Procedia Engineering, 171, 1250-1255

[3] Cha, Y., Choi, W., and Büyüköztürk, O. (2017). "Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks" Computer-Aided Civil and Infrastructure Engineering Engineering, 32(5), 361-378

[4] Pauly, L., Peel, H., Luo, S., Hogg, D., and Fuentes, R. (2017) "Deeper Networks for Pavement Crack Detection" in 34th International Symposium on Automation and Robotics in Construction (ISARC)

[5] Hoang, N. (2018). "Detection of Surface Crack in Building Structures Using Image Processing Technique with an Improved Otsu Method for Image Thresholding". Advances in Civil Engineering, 2018.

[6] Zheng, M., Lei, Z., and Zhang, K. (2020) "Intelligent detection of building cracks based on deep learning" Image and Vision Computing, 103, 103987

[7] Fan, Z., Li, C., Chen, Y., Wei, J., Loprencipe, G., Chen, X., and Di Mascio, P. (2020) "Automatic Crack Detection on Road Pavements Using Encoder-Decoder Architecture" Materials, 13(13), 2960

[8] Li, C., and Ai, D. (2017). "Automatic crack detection method for loaded coal in vibration failure process". PLOS ONE, 12(10), e0185750

[9] Chen, F., and Jahanshahi, M. (2018). "NB-CNN: Deep Learning-Based Crack Detection Using Convolutional Neural Network and Naïve Bayes Data Fusion". IEEE Transactions on Industrial Electronics, 65(5), 4392-4400

[10] Zhou, S., and Song, W. (2020). "Robust Image-Based Surface Crack Detection Using Range Data". Journal of Computing in Civil Engineering, 34(2), 04019054

[11] Lin, Q., Wan, B., Wang, S., Li, S., and Fakhimi, A. (2019). "Visual detection of a cohesionless crack in rock under three-point bending". Engineering Fracture Mechanics, 211, 17-31

[12] Zhang, L., Yang, F., Daniel Zhang, Y., and Zhu, Y. (2016). "Road crack detection using deep convolutional neural network". 2016 IEEE International Conference on Image Processing (ICIP)

[13] A, J., G, T., M S, D., Nair, B., and D. S, H. (2020). "Asphalt Crack Dataset". Data.mendeley.com<https://data.mendeley.com/datasets/xnzhj3x8v4/2> (Sep. 27, 2020)

[14] Xu, H., Su, X., Wang, Y., Cai, H., Cui, K., and Chen, X. (2019). "Automatic Bridge Crack Detection Using a Convolutional Neural Network". Applied Sciences, 9(14), 2867

[15] Trivedi, A. (2015). "Computing in-situ strength of rock masses based upon RQD and modified joint factor: Using pressure and damage sensitive constitutive relationship". Journal of Rock Mechanics and Geotechnical Engineering, 7(5), 540-565

[16] A. K. Jha, S. Adhikari, S. Thapa, A. Kumar, A. Kumar, and S. Mishra, "Evaluation of Factors Affecting Compressive Strength of Concrete using Machine Learning" in 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA), 2020: IEEE, pp. 70-74.

[17] Tyagi, L., Butola, R., and Jha, A. (2020). "Mechanical and tribological properties of AA7075-T6 metal matrix composite reinforced with ceramic particles and aloevera ash via Friction stir processing". Materials Research Express, 7(6), 066526

[18] Özgenel, Ç. (2020). "Concrete Crack Images for Classification". *Data.mendeley.com*, <https://data.mendeley.com/datasets/5y9wdsg2zt/1> (Sep. 27, 2020)

[19] A. Ghimire, S. Thapa, A. K. Jha, A. Kumar, A. Kumar, and S. Adhikari, "AI and IoT Solutions for Tackling COVID-19 Pandemic", in 2020 International Conference on Electronics, Communication and Aerospace Technology, 2020: IEEE.

[20] A. Ghimire, S. Thapa, A. K. Jha, S. Adhikari, and A. Kumar, "Accelerating Business Growth with Big Data and Artificial Intelligence", in 2020 Fourth International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2020: IEEE, pp. 441-448.

[21] S. Thapa, P. Singh, D. K. Jain, N. Bharill, A. Gupta, and M. Prasad, "Data-Driven Approach based on Feature Selection Technique for Early Diagnosis of Alzheimer's Disease", in 2020 International Joint Conference on Neural Networks (IJCNN), 2020: IEEE, pp. 1-8.

[22] S. Thapa, S. Adhikari, A. Ghimire, and A. Aditya, "Feature Selection Based Twin-Support Vector Machine for the diagnosis of Parkinson's Disease", in 2020 8th R10 Humanitarian Technology Conference (R10-HTC), 2020: IEEE

Emerging Biosensors for Point of Care Cancer Detection

Varinder Garg[1], Harish Kumar[2], Surinder Rana[3], Bikramjeet Singh Kalsi[4], Sidharth Bhasin[5], Mehak Sood[6], Ramandeep Kaur[7], Siddhant Rajkumar[8]

[1] MBBS, MD Principal Investigator, ICMR-Centre for Innovation and Bio-Design, Room No. 2022 and 2027, Advanced Cardiac Centre, PGIMER, Sector 12, Chandigarh, 160012. Email: me_dvg@yahoo.co.in
[2] B.E., M.E., PhD CSE Dept, UIET Panjab University, Chandigarh 160014.
[3] M.D., D.M Gastroenterology, Professor, Department of Gastroenterology, PGIMER, Sector 12, Chandigarh, 160012. Email: rana.surindersingh@pgimer.edu.in
[4] B. Tech., Computer Science, UIET Panjab University, Chandigarh 160014.
[5,8] B. Tech in Biotechnology, Department of Biotechnology, Delhi Technological University
[6] M.Sc. Chemistry, Department of Chemistry, Post Graduate Government College for Girls, Sector-11, Chandigarh.
[7] B.Sc Honors in Microbiology and Food Technology, MCM DAV College Sec-36A, Chandigarh

ABSTRACT

Cancer is a group of diseases characterized by the six hallmarks namely, self-reliance for growth signals, unresponsiveness to anti-growth signals, evasion of programmed cell death or apoptosis, sustained angiogenesis, infinite replication potential and tissue invasion and metastasis (Hanahan & Weinberg, 2000). The original set has been extended to include two more hallmarks viz. altered energy metabolism pathways and escaping immune-targeted destruction (Hanahan & Weinberg, 2011). Cancer is one of the leading causes for mortality worldwide, with approx. 9.6 million deaths reported globally in 2018 (Bray et al., 2018; Ferlay et al., 2018). The number of deaths has doubled from 1990 to 2016 in India (Dhillon et al., 2018). Its prevalence has been estimated at 18.1 million new cases globally per year in 2018 with 1.15 million new cases per year in India (Ferlay et al., 2019). This number is expected to double by 2030 (Smith & Mallath, 2019). Moreover, cancer treatment puts an enormous financial burden on the patient. The Out of Pocket (OOP) Expenditure can range anywhere from $7500 to $25000 per month for patients with multiple prescribed cancer therapies (Carrera, Kantarjian, & Blinder, 2018). Furthermore, five-year cancer survival rate ranges from as high as 98% in prostate cancer to as low as 9% for pancreatic cancer (Siegel, Miller, & Jemal, 2020).

Keywords: Pancreatic Cancer, Metastasis, Apoptosis.

EMERGING BIOSENSORS FOR POINT OF CARE CANCER DETECTION

Biosensors for Point of Care Testing

Researchers nowadays, are keenly interested in detection of cancer at the very early stage using biosensors because of their high efficiency, sensitivity, accuracy and specificity. In the new era of genomic oncology, genetic biomarkers are becoming the core of cancer biomarkers; that can be identified by the specially designed biosensors.

1.1 Breast Cancer

Breast cancer is very common cancer among women, implicating 2.1 million women each year and also causes greatest number of cancers related deaths. Breast cancer is second most common cancer among all. According to WHO, in 2018 it is estimated 15% of the women died of breast cancer than of all cancer related deaths, that's approximately 627000 in number. Some important biomarkers responsible for breast cancer are BRAC1, BRAC2, CA15-3, CA125, CA27.29, CEA, NY-BR-1, ING-1, HER2 (Human epidermal growth factor receptor), ER (estrogen receptor) / PR (progesterone receptor).

In breast cancer patients, CA15-3 is an important biomarker. Clinically CA15-3 is most often used to monitor patient in cases of advanced breast cancer. It has been seen that the concentrations of CA15-3 increase by 10% in stage1 cancer, 20% in stage 2, 40% in stage 3, 75% in stage 4 of breast cancer. The relationship between CA15-3 levels and breast cancer has shown that the individuals with the values of <30U/mL, before the treatment onset had significantly higher survival times than individuals who had higher levels. For the determination of biomarkers in the breast cancer with high sensitivity and diagnostic accuracy, the biosensors with nanomaterial and their specific characteristics such as metal oxide, metal nanoparticles, nanospheres and integrated nanostructures, such as graphene or reduced graphene oxide, composed of metal oxide and multiwalled carbon nanotubes were provided.

1.1.1 The Detection of Breast Cancer Biomarkers with Microfluidic Immuno–Biochip

The early detection of Epidermal growth factor receptor 2 (EGFr2 or ErbB2) protein with label- free microfluidic immunosensor which is with femtomolar sensitivity. This sensor applies exclusively structured immune-electrode which is made up of porous hierarchical graphene foam modified with electrospun carbon-doped titanium dioxide nanofibers(nTiO2) which is working as an electrochemical electrode.

Because of the brilliant biocompatibility, the applications which are ideal for the electrochemical sensor are high reaction kinetics, anatase nTiO2 and good stability for proteins. The 3D and the porous features of graphene foam allow nTiO2 to penetrate and attach to the surface of graphene foam by physical adsorption. The graphene foams combine with the functional nTiO2 which yields in large surface area, high charge transfer resistance, and porous access to sensing surface by the analyte, which results in new possibilities for the development of electrochemical immunosensors. The antibody of ErbB2 on the graphene foam – nTiO2 composite gets covalently immobilized by the enabling of EDC-NHS chemistry. To get an immunosensor's design, the working electrode was designed to hang above the gold counter electrode in a microfluidic channel. The immunosensor went through electrochemical impedance spectroscopy and differential pulse voltammetry to quantify breast cancer biomarkers. The high sensitivities of the two methods are 0.585μA μ/M cm/2 and 43.7kΩ μ/M cm/2 in which wide concentration range of target ErbB2 antigen from 1×10 ¯15 M to 0.1×10¯6 M and from 1×10¯13 M to 0.1×10¯6 M respectively. Usage of the anti ErbB2, results in high specificity, even in the presence of ErbB3 and ErbB4. Many applications will be derived from the integration of graphene foam – nTiO2 composite into microfluidic devices, in the field of electrochemical detection of chemical and biological species.

1.1.2 Early Diagnosis by DNA Biosensors Based on AuNPs

For early detection breast cancer biomarkers by electrochemical detection of HER2, two different DNA modified gold nanoparticles and graphene oxide loaded on glassy carbon electrodes were prepared. In this electrochemical DNA biosensor, a "sandwich type" detection strategy was employed and its response was measured by Amperometric detection. For the sensitive detection of the breast cancer biomarker ErbB2 and the control marker CD24, the electrochemical signal enhancement was achieved via gold nanoparticles and graphene oxide system. The modified graphene oxide was described by using Fourier transform infrared spectroscopy, transmission electron microscopy, Raman spectroscopy, scanning electron microscope, UV-visible spectroscopy and energy-dispersive X-ray spectroscopy. Some important steps involved in modification of glassy carbon electrode with gold nanoparticles, graphene oxide and DNA probes, target and reporter probe were electrochemically characterised using electrochemical impedance spectroscopy and cyclic voltammetry.

The detection limits of 0.16nM and 0.23nM were obtained with sensitivity 378nA/nM and 219nA/nM for ErbB2 and CD24 respectively by using Amperometric detection of a horse radish peroxide label.

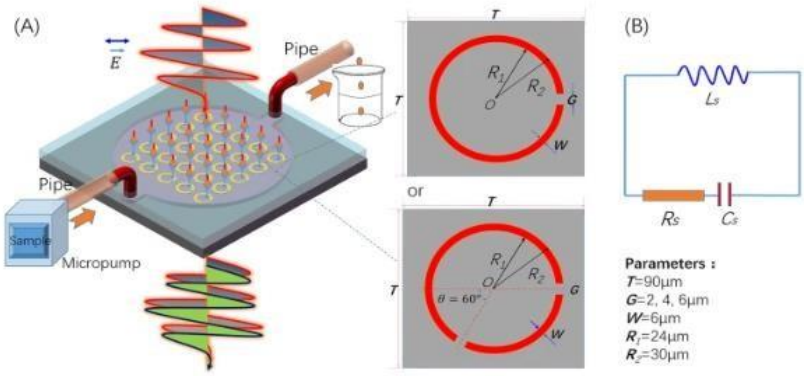
1.2 Liver Cancer

Liver cancer is considered to be the fourth most lethal cancer globally, and hepatocellular carcinoma (HCC) accounts for 75–85% of liver cancer cases. The most approachable HCC biomarkers are (1) the AFP, its isoform lens culinaris agglutinin-reactive fraction of alpha- fetoprotein (AFP-L3); and (2) des-γ-carboxy prothrombin (DCP). However, there are many other molecules that might be taken into account, including glypican 3 (GPC3), glutamine synthase (GS), heat shock protein 70 (HSP70), cytokeratin 19 (CK19), Golgi protein 73 (GP73), midkine, osteopontin (OPN), squamous cell carcinoma antigen (SCCA), Annexin A2, fibroblast growth factor 3/4 (FGF3/4), micro-RNAs (miRNAs), Long non-coding RNAs (lncRNAs), circulating tumor cells (CTCs), cell-free DNA (cfDNA), and other biomarkers.

Recently, specially designed metamaterials in the Terahertz range are being widely used as biosensors for liver cancer detection. It works on the principle of resonating frequencies, where the frequency range of Terahertz Wave (0.1THz to 10 THz) matches with the vibrational frequency of biomolecules released from biomarkers. The intensity of peak gives the estimation of the concentration of the biomarkers in the body. This technique provides the promising results, moreover, it a label-free, non-contact and non-destructive inspection technique on the specific biomolecules. To obtain the sharp peaks, the metamaterials can be designed as asymmetric structures, which can cause fano resonance; provided the structure can resonate and eliminating the substrate effect. Because of the limitation of the THz sensors to work on dry or semi-hydrated sample due to strong water absorption at THz range, Microfluidic chips are integrated with biosensors providing advantages such as low cost, less sample requirement, higher accuracy and rapid testing. The metamaterials are designed with antibody specific to antigen of biomarker in the aqueous phase.

Designing of Metamaterial Biosensor Microfluidic Chip: The metal-split-ring resonators(MSSRs) and poly-(dimethylsiloxane) microchannel form the basic structure of biosensor. The dimethylsiloxane (PDMS) form the microfluidics layer over the metamaterial to prevent the loss of THz wave transmission.

The THz wave which is electrically polarised is launched on the THz chip in integration of microfluidics chip, where the electric field parallel to MSSRs excites the inductive-capacitive resonance, which induced the current in the nearby MSSRs. This leads to electric energy and magnetic energy transfer between the inductance and capacitance. The chip detects the strong electrical field generated among the dielectric gap and provides us the signal. The metamaterial was fabricated on highly resistive silicon substrate. The designing involves 4-6 µm width(W) of split ring, inner radius and outer radius of range 24-26 µm and 28-30µm respectively. The size of gap can be changed from 2µm to 6µm based on the accuracy requirement.



(**A**) The sketch of THz metamaterials biosensor chip integrated with microfluidics; (**B**) Equivalent circuit with {RLC}s for the SRRs.

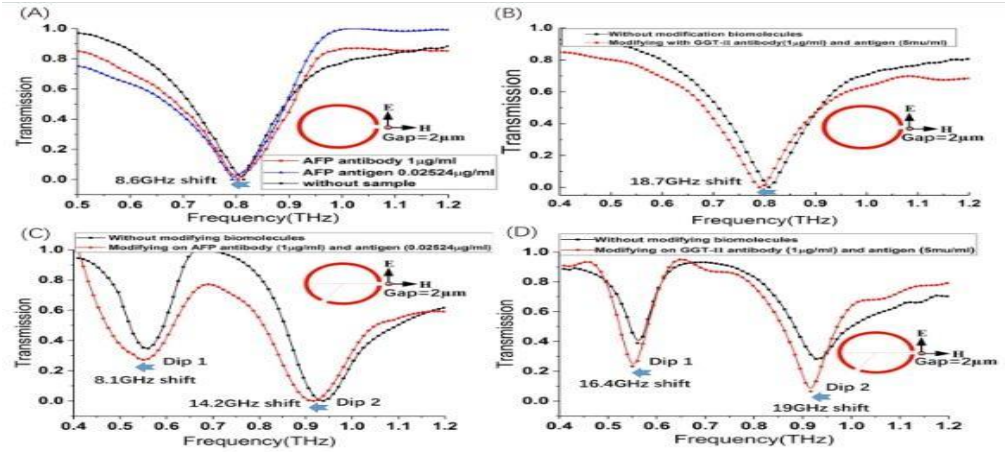
RESULTS

Liver cancer antibody Alpha fetoprotein (AFP) (1 µg/mL solution in PBS buffer) were coupled to the surface of the SRRs through chemical reaction between amidogen beside the alkaline aminophenol (Arg and Lys) of IgG and the active carboxyl; and maintaining 16 hours at 4 °C. AFP antigen or Glutamine transferase isozymes II (GGT-II) antigen with different concentrations was injected into the chamber to incubate for more than 40 min. The reflection spectra or transmission spectra were recorded through THz-TDS setup during different time. The MSSRs biosensors integrated with microfluidics is based on THz spectral dip frequency shifts, which can be changed with refractive index. A figure of merit (FOM) obtained by dividing the sensitivity by the resonance frequency line width is widelyused to characterize SRRs sensing capabilities. FOM is described as:

FOM=S/FWHM

where, FWHM repents full width at half maximum.

Larger the value of FOM, precise will be the results. The results of AFP and GGT II at 2 µm are shown in the graph since we know, lower the gap, higher will be precision. The testing results of the two AFP and GGT II show in agreement with the experimental results. However, the accuracy can be enchanced using advanced Metamaterials operating on fano resonances.



The typical tested results of liver cancer biomarker. (**A**) The results for AFP of SRRs with a 2 µm gap; (**B**) The results for GGT-II liver cancer marker testing of SRRS with a 2 µm gap; (**C**) The results for AFP of SRRs with 2 gaps (2 µm); (**D**) The results for GGT-II liver cancer marker testing of SRRS with 2 gaps (2 µm).

1.1 Lung Cancer

Lung cancer is the leading cause of cancer-related deaths, with mortality being closely associated with smoking habits (Barta, Powell, & Wisnivesky, 2019). The common symptoms include coughing up blood, weight loss, fever, fatigue and anorexia among others. The main problem associated with LC is its misdiagnosis as tuberculosis in countries where TB is prevalent. This is due to the fact that both diseases share common symptoms (Prabhakar, Shende, & Augustine, 2018). However, it can be distinguished on the basis of patient's medical history and whether the person smokes or not as TB can occur in non-smokers. Moreover, TB

Yet another biosensor which prove effective are the microcantilever sensors for liver cancer detection. In microcantilever, a cavity was designed in the free end of the cantilever for local body immobilization. This local immobilization approach reduces the adsorption induced variation of Stiffness constant, k in comparison to the whole surface adsorption.

EMERGING BIOSENSORS FOR POINT OF CARE CANCER DETECTION

A novel cantilever array-based bio-sensor was batch-fabricated with IC compatible MEMS technology for precise liver cancer biomarker detection, that is piezoelectrically driven into vibration and those vibrations can be detected by Wheatstone bridge. The resonance frequency depends on two factors- stiffness constant,k and mass, m.

BIOMARKER DETECTION: For the detection of Biomarkers various concentrations of antigens were taken into account. The figure depicts the measured frequency shift of locally immobilized anti-AFP levers to the various concentrations of antigen AFP, and a linear regression was obtained with a relative uncertainty less than 5%.

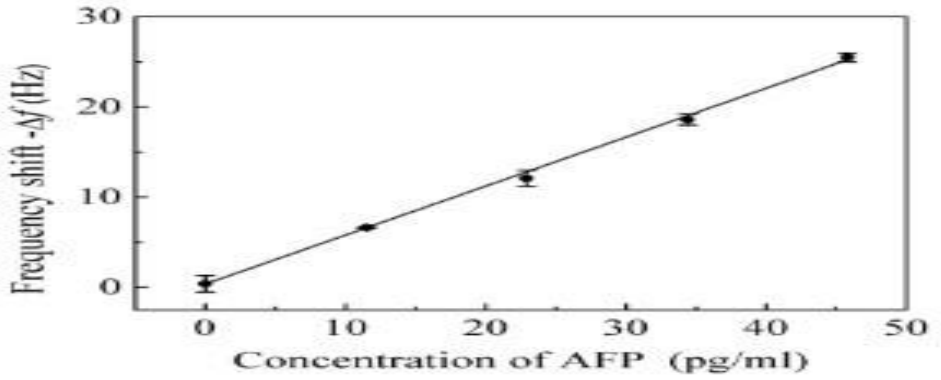


Figure 2: The relative resonance frequency shift versus AFP antigen concentration of 0-50 pg/ml.

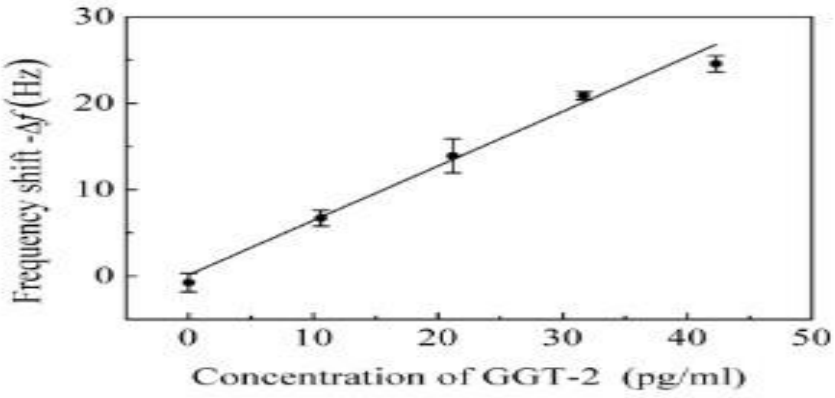


Figure 3: The relative resonance frequency shift versus GGT-2 antigen concentration of 0-50 pg/ml.

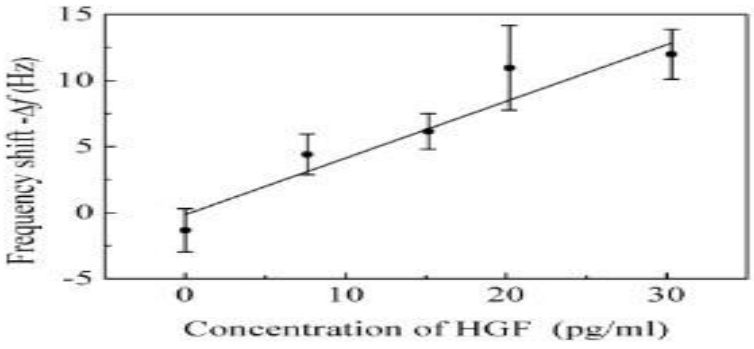


Figure 4: The relative resonance frequency shift versus HGF antigen concentration of 0 30 pg/ml.

Eliminating the effect of k can help detecting the various liver cancer biomarkers more efficiently.

can act as starting point for LC (Prabhakar et al., 2018). Lung cancer can be further classified into LCT, SCLC and NSCLC based on histology (Roointan et al., 2019).

Biomarkers and biosensors can be employed for early diagnosis of lung cancer due to their fast response time and high sensitivity. The biomarkers can be broadly classified into genetics and epigenetics biomarkers and protein-based biomarkers. Some of the common biomarkers utilized are CEA or Carcinoembryonic Antigen, NSE or Neuron-specific enolase, CYFRA21- 1, VEGF, Annexin-II and miRNAs like miR-205, miR-210, miR-708 (Roointan et al., 2019).

1.1.1 Modified Surface Plasmon Resonance based Biosensors

SPR based biosensors have received a lot of interest lately due to their high sensitivity, non- invasiveness and real time results analysis. A label-free, optical SPR works on the principle of detection of changes in refractive index when analyte molecules bind to the immobilized biorecognition element on the surface of biosensor chip (Piliarik, Vaisocherová, & Homola, 2009). These biosensors are cost-effective, portable, and easy-to-use with high specificity and accuracy. Moreover, their specificity and sensitivity can be greatly enhanced by making slight modifications.

Chiu and Yang constructed a novel SPR biosensor using carboxyl-MoS_2 nanocomposite film (instead of using just MoS_2) for detection of LC biomarker CYFRA 21-1. This modification resulted in a fifteen-fold enhancement in the binding affinity K_a and more than two-fold enhancement in SPR angle shift. The LOD of this biosensor (=0.05 pg/mL) was found to be 10^4 times more sensitive than traditional ELISA (=0.60 ng/mL) (Chiu & Yang, 2020).

Miti et al. developed a miRNA biosensor based on localized SPR. This biosensor used HCR or Hybridization Chain reaction for amplification. It works on the principle that two DNA hairpins, in the presence of target sequence, will initiate assembly. This can then be successfully detected using SPR. This biosensor was successful in detecting LC biosensor miRNA-17 and the overall procedure was completed within an hour (Miti et al., 2020). SPR based biosensors hold a lot of potential to be widely adopted as POC diagnostic devices for detection of various types of cancer biomarkers.

1.2 Prostate Cancer

Prostate cancer is the second most common cancer among men. Generally, the tumour formed is latent i.e. it does not mature and affects the patient's life. Thus, it is very important to identify whether the tumour is latent or not as early identification can help save lives. Despite the

limited specificity and high rate of overdiagnosis, Prostate Specific Antigen (PSA) is the most commonly used biomarker for the detection of prostate cancer. To counter this issue new prostate cancer antigen have been proposed. These biomarkers have higher accuracy of PSA in the management of early Prostate Cancer Antigen. Some of the commercially available Prostate cancer antigens are: PCA3 score, four-kallikrein panel and Prostate Health Index. There are also new emerging biomarkers such as PSA glycoforms, microRNAs, TMPRSS2: ERG fusion gene, androgen receptor variants, circulating tumour cells, and PTEN gene. These biomarkers can change the early management of Prostate cancer. Using these biomarkers biosensors can be designed. These biosensors help in diagnosis of cancer.

1.2.1 Bioconjugation Single use biosensor

Detection of Prostate cancer is done most commonly with the help of PSA despite its high rate of false positive tests. Apart from PSA, alpha-methylacyl-CoA racemase (AMACR) is also highly expressed biomarker in the Prostate cancer cells.

The novel bioconjugate single use Biosensor is capable of detecting both PSA and AMACR in undiluted human serum. The preparation of biosensor via bioconjugation mechanism does not take much time and can be prepared just prior to the test. This type of biosensor is accurate, single use and cost effective. It also requires very small quantity of test medium coupled with its short preparation time it becomes a very attractive choice for detection of biomarkers produced by prostate cancer cells.

Detection of AMACR is conventionally done with the help of ELISA, chemiluminescence immunoassay, radioimmunoassay, and fluoro-immunoassay. Even though these methods have fairly high accuracy the process is fairly laborious, expensive and time consuming. Thus the cost effective biosensor technology became more popular for detection of biomarkers in early stages of disease monitoring. The most common way to detect AMACR is using Square Wave Voltammetry to directly detect AMACR using aptamers immobilized on a polymer substrate.

Bioconjugation mechanism conjugate two or more molecules into a single molecule which contains properties of all the components. This method makes for zero-length linkage between the protein and the electrode element of the biosensor. Bioconjugation also has improved clinical applications as it shortens the preparation time and enhances the coverage of biosensor surface. The interaction between antigen and antibody is the biorecognition mechanism. Anti- PSA and Anti-AMACR were modified with N-succinimidyl S-acetylthioacetate(SATA) with the help of bioconjugation mechanism. This leads to a thiol linked anti-PSA or anti-AMACR

which directly links to the thin gold sheet electrode element of the biosensor during incubation. Preparation including the incubation takes about one day.

Thus, the use of bioconjugation in preparation and pulse voltammetry in I measurement results in a single use, cost effective and highly sensitive and selective biosensor for detection and management of early-stage prostate cancer with the help of PSA and AMACR biomarkers. This type of biosensor is proving it self to be a very effective diagnostic tool for the screening application of prostate cancer.

1.2.2 Graphene based biosensor

Graphene is playing an important role in the biosensor field with remarkable physical, optical, electrochemical and magnetic properties.

The use of graphene can significantly enhance the electrochemical signal of various electrodes because of various properties such as high electrochemical surface area, descent conductivity, high electron transfer rate, and a broad range of chemical functional groups on the surface of graphene. There are certain limitations in the use of graphene as it has irreversible self-agglomerations, non-specificity, less colloidal stability repeatability and poor reliability.

Based on graphene two types of biosensors can be made, firstly, graphene material based on GO, GQDs and rGO are assembled in the surface of the biosensor, mainly the electrode or transistor channel, to construct a novel biosensor interface with improved molecular receptors. Based the increased specific surface area and unique π–π orbital interactions on the interface excellent biosensor performance can be achieved. Secondly graphene materials can be applied as carriers for construction of novel nanocomposite. This is approach is very suitable for sensitive protein biomarker analysis because of its unique catalytic and chemical reactions and biosensor signal amplification.

ACKNOWLEDGEMENTS

The authors acknowledge the guidance provided by Centre for Innovation and Bio-Design (ICMR-CIBioD) for the expert review of the manuscript and providing the necessary infrastructure.

DECLARATION OF INTERESTS

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

REFERENCES

Barta, J. A., Powell, C. A., & Wisnivesky, J. P. (2019). Global Epidemiology of Lung Cancer. *Annals of Global Health, 85*(1), 8. doi:10.5334/aogh.2419

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin, 68*(6), 394-424. doi:10.3322/caac.21492

Carrera, P. M., Kantarjian, H. M., & Blinder, V. S. (2018). The financial burden and distress of patients with cancer: Understanding and stepping-up action on the financial toxicity of cancer treatment. *68*(2), 153-165. doi:10.3322/caac.21443

Chiu, N.-F., & Yang, H.-T. (2020). High-Sensitivity Detection of the Lung Cancer Biomarker CYFRA21-1 in Serum Samples Using a Carboxyl-MoS(2) Functional Film for SPR- Based Immunosensors. *Frontiers in bioengineering and biotechnology, 8*, 234-234. doi:10.3389/fbioe.2020.00234

Dhillon, P. K., Mathur, P., Nandakumar, A., Fitzmaurice, C., Kumar, G. A., Mehrotra, R., . . . Dandona, L. (2018). The burden of cancers and their variations across the states of India: the Global Burden of Disease Study 1990–2016. *The Lancet Oncology, 19*(10), 1289-1306. doi:10.1016/S1470-2045(18)30447-9

Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Piñeros, M., . . . Bray, F. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *144*(8), 1941-1953. doi:10.1002/ijc.31937

Ferlay, J., Ervik, M., Lam, F., Colombet, M., Mery, L., Piñeros, M., . . . Bray, F. (2018). Global Cancer Observatory: Cancer Today. Retrieved from https://gco.iarc.fr/today Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell, 100*(1), 57-70. doi:10.1016/s0092-8674(00)81683-9

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell, 144*(5), 646-674. doi:10.1016/j.cell.2011.02.013

Miti, A., Thamm, S., Müller, P., Csáki, A., Fritzsche, W., & Zuccheri, G. (2020). A miRNA biosensor based on localized surface plasmon resonance enhanced by surface-bound hybridization chain reaction. *Biosens Bioelectron, 167*, 112465. doi:10.1016/j.bios.2020.112465

Piliarik, M., Vaisocherová, H., & Homola, J. (2009). Surface plasmon resonance biosensing. *Methods Mol Biol, 503*, 65-88. doi:10.1007/978-1-60327-567-5_5

Prabhakar, B., Shende, P., & Augustine, S. (2018). Current trends and emerging diagnostic techniques for lung cancer. *Biomed Pharmacother, 106*, 1586-1599. doi:10.1016/j.biopha.2018.07.145

Roointan, A., Ahmad Mir, T., Ibrahim Wani, S., Mati Ur, R., Hussain, K. K., Ahmed, B., . . . Akhtar, M. H. (2019). Early detection of lung cancer biomarkers through biosensor technology: A review. *J Pharm Biomed Anal, 164*, 93-103. doi:10.1016/j.jpba.2018.10.017

Siegel, R. L., Miller, K. D., & Jemal, A. (2020). Cancer statistics, 2020. *70*(1), 7-30. doi:10.3322/caac.21590

Smith, R. D., & Mallath, M. K. (2019). History of the Growing Burden of Cancer in India: From Antiquity to the 21st Century. (5), 1-15. doi:10.1200/jgo.19.00048

EMERGING BIOSENSORS FOR POINT OF CARE CANCER DETECTION



Textile & Leather Review

ISSN: 2623-6257 (Print) 2623-6281 (Online) Journal homepage: www.textile-leather.com Journal doi: 10.31881/TLR

Enzyme-Accelerated Acid Hydrolysis of Untanned Proteinaceous Wastes from Tanning Industry

Md. Nur-E-Alam, Nasifa Akter, Kanish Fatema, Md. Abul Kashem Azad, Shimul Chakma, Md. Anwar Arfien Khan

How to cite: Nur-E-Alam M, Akter N, Fatema K, Azad MAK, Chakma S, Khan MAA. Enzyme-Accelerated Acid Hydrolysis of Untanned Proteinaceous Wastes from Tanning Industry. Textile & Leather Review. 2020. https://doi.org/10.31881/TLR.2020.13

How to link: https://doi.org/10.31881/TLR.2020.13

Published: 1 December 2020

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

Enzyme-Accelerated Acid Hydrolysis of Untanned Proteinaceous Wastes from Tanning Industry

Md. NUR-E-ALAM[1]*, Nasifa AKTER[1], Kanish FATEMA[1], Md. Abul Kashem AZAD[1], Shimul CHAKMA[1,2], Md. Anwar Arfien KHAN[3]

[1]Leather Research Institute, Bangladesh Council of Scientific and Industrial Research (BCSIR), Savar, Dhaka-1350, Bangladesh
[2]Delhi Technological University, Delhi, 110042, India
[3]Institute of Mining, Mineralogy and Metallugy, BCSIR, Joypurhat
*nalam1980@yahoo.com

Original scientific article
UDC 675.088:66.094.941:547.96
DOI: 10.31881/TLR.2020.13
Received 15 Jun 2020; Accepted 30 Aug 2020; Published Online 2 Oct 2020; Published 1 Dec 2020

ABSTRACT

Traditionally, tanning industry has been producing considerable amounts of solid wastes, which raises serious concerns on account of their environmental impact. Out of these, untanned raw trimmings account for about 5-7% of the total quantity of raw materials processed. This waste could be a value-added cheap source of collagen, which has numerous industrial applications if properly and scientifically utilized. This research work deals with the utilization of raw trimmings of solid waste from tanneries in the process of enzymatic hydrolysis, performed by using acetic acid and protease, in order to obtain protein hydrolysate. The hydrolysis was carried out with varying acid concentrations, acid solutions, temperatures and times. The maximum obtained protein hydrolysate was about 88% at 1.5 M acid concentration, 4% enzyme ratio, and 60 °C.

KEYWORDS

Raw trimmings, Untanned proteinaceous solid wastes, Acid hydrolysis

INTRODUCTION

Tanning raw hides and skins is one of the oldest processes of our ancestors. The search for methods of conserving hides and skins started in the early Stone Age, around 8,000 BCE. Around five thousand years later, the people of Egypt and Mesopotamia are said to have invented plant-based tanning, using the bark or gum of various trees [1]. The modern tanning process involves several steps in converting raw hides and skins into imputrescible substance and also generates large quantities of solid and liquid wastes. On average, the processing of one metric ton of rawhide produces 200 kg of tanned leather, 250 kg of non-tanned waste, 200 kg of tanned waste leather, and 50,000 kg of liquid waste (Table 1). About 50% of leather mass is lost during the tanning process [2].

Table 1. Environmental input of leather processing, adapted from [3]

Raw hide (1 ton)				
Leather	Solid wastes / Byproducts		50^3 liquid effluent	
200 kg	Untanned wastes:		COD	235-250 kg
	Raw trimmings	120 kg	BOD	100 kg
	Fleshing	70-230 kg	Suspended solids	150 kg
	Tanned wastes:		Chromium	5-6 kg
	Tanned splits	115 kg	Sulfide	10 kg
	Trimmings + Shavings	100 kg		
	Dyed / Finished:			
	Buffing	2 kg		
	Trimmings	32 kg		

The tannery sector of Bangladesh is considered as the most polluting of industries, which is well established, and ranked fourth in terms of earning foreign exchange. The first tannery in Bangladesh was established at Narayanganj by R.P. Shaha in 1940. After that, the tannery industry was shifted to Hazaribagh, Dhaka [4]. At present all tanneries of Hazaribagh have been relocated to Savar Tannery Estate in order to protect the environment. As mentioned earlier, tanning processes generate huge amounts of liquid and solid wastes and if these wastes are not properly treated it will degrade the surrounding environment. Although tanneries have been moved to a new location, solid waste management is not properly maintained.



Raw trimmings of hides and skins



Solid waste in a nearby tannery's yard

Figure 1. Solid wastes from tanneries discarded without any treatment

Raw hides and skins are by-products of meat industry. Before pre-tanning operations, the hides and skins are trimmed (for the convenience of machine operations), which generates huge amounts of un-tanned proteinaceous solid wastes. These trimmings account for about 5-7% of the total quantity of raw materials processed. If these proteinaceous materials are not utilized appropriately, that will create hazardous pollution to the environment [5].

Leather solid wastes can be processed into valuable products such as glue, gelatin, artificial fibrous leathers, and collagen for various industrial uses. A number of authors have reported about various way of turning these wastes into valuable products. Enzymes, such as papain and neutrases [6], acids, such as phosphoric acid [7], sulphuric acid [8], propionic acid [9] and hydrochloric acid [10], and alkali, such as sodium hydroxide [11], magnesium oxide, calcium oxide [8] etc. are used as hydrolyzing agents in the treatment of solid waste generated by the tanning industry. Eco-friendly ultrasound technology can also be used to accelerate the

hydrolyzing process [12]. In this research work, acetic acid (CH_3COOH) and protease enzyme are used as hydrolyzing agents in the process of extracting the protein hydrolysate from untanned raw trimmings from tanneries in Bangladesh.

MATERIALS AND METHODS

Material preparation

Raw trimmings from tanneries were collected from the Savar tannery estate, Savar, Dhaka. These were first washed with an excess amount of tap water in order to remove salt, dirt, dung, blood and other impurities. Hairs were removed by the liming process, using calcium oxide (CaO), and finally unhaired trimmings were delimed by washing them several times with water. After that, they were air dried in the open.
Then, the dried trimmings were cut into small pieces for hydrolysis. Acetic acid (CH_3COOH), lime and protease enzyme were purchased from the local market. Glassware (pipettes, beakers, conical flasks, measuring cylinders, test tubes etc.) that was used was the product of Borosil/Ranken. A magnetic hotplate, a stirrer, Kjeldahl apparatus etc. were used. Acetic acid is commonly used as a souring agent in the process of making vinegar, pickled vegetables and sauce, and as a raw material for spices, in diluted concentrations (4 to 8% by mass). For this reason, acetic acid was chosen as a hydrolysis agent in this study.

Experimental procedure

The study was performed in a batch process in a series of beakers equipped with stirrers by stirring dried raw trimmings and using acetic acid in varying concentrations and at varying temperatures. The detailed experimental methodology is given in Figure 2.

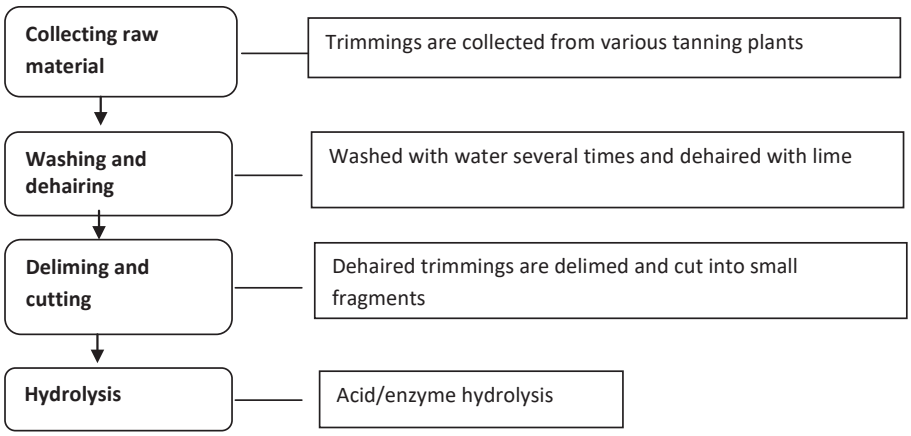


Figure 2. Experimental procedures of hydrolysis

Effect of acid concentration:
The effect of acid concentration on the protein hydrolysis can be examined by varying the concentration of acid in the fixed acid solution at a fixed temperature. 50 gm of cleaned and dried raw trimmings was dissolved in 5 ml of 0.25M, 0.5M, 1.0M and 1.5M acetic acid solution. An extra 400 ml of distilled water was added to each and the temperature was set to 40 °C.

Effect of acid solution:
50 gm of cleaned and dried raw trimmings was put in the beaker with the optimal acid concentration and at a fixed temperature of 40 °C.

Effect of temperature:
The examination of the effect of temperature on the protein hydrolysis was carried out with the optimal acid concentration and its solution.

Effect of hydrolysis time:
The examination of the effect of hydrolysis time (hour) was carried out with the optimal acid concentration, its solution and temperature.

Enzymatic hydrolysis:
The examination of the effect of enzyme concentration was carried out with best the optimal concentration, its solution, time and temperature.

Analysis

The protein of the extracted hydrolysate was identified by the Biuret test and the protein concentration was determined by the Kjeldahl method in a Gerhardt digester (Germany).

a) Biuret test
The Biuret test is based on the ability of Cu(II) ions to form a violet-colored chelate complex with peptide bonds (-CONH- groups) in alkaline conditions. This test confirms the presence of proteins in the sample. In this test, 2 ml of extracted hydrolysate solution was taken in a dry test tube. 3 drops of 10% NaOH and 3-6 drops of 0.5% $CuSO_4$ were added to the sample test tube [13].

b) *Kjeldahl method*
The Kjeldahl method is used to determine the nitrogen content in organic and inorganic substances. For over a hundred years the Kjeldahl method has been used for the determination of nitrogen in a wide range of samples such as foods and drinks, meat, feeds, cereals and forages. It is also used for nitrogen determination in wastewaters, soils and other samples. The Kjeldahl method has three main steps (Figure 3): digestion, distillation, and titration.

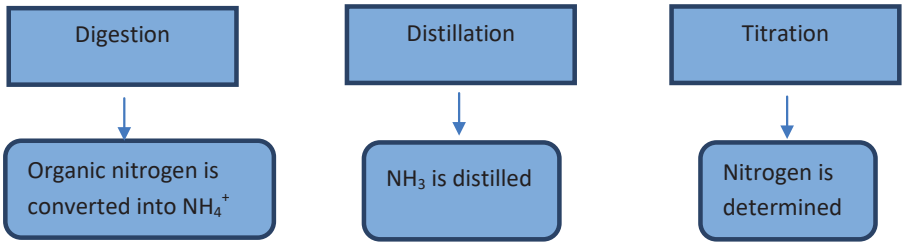


Figure 3. Main steps of Kjeldhal method

Digestion
The aim of the digestion procedure is to break all nitrogen bonds in the sample and convert all of the organically bonded nitrogen into ammonium ions (NH_4^+), carbon dioxide and water (equation 1). During digestion, the organic material carbonizes, which can be visualized by the transformation of the sample into black foam. After that, the foam decomposes and finally a clear liquid indicates the completion of the chemical reaction. For this purpose, the sample is mixed with sulfuric acid at temperatures between 350 and 380 °C. The higher the temperature, the faster the digestion can be obtained.

The speed of the digestion can be greatly enhanced by the addition of salt and catalysts. Potassium sulfate (K_2SO_4) is added in order to increase the boiling point of sulfuric acid, and catalysts (e.g. $CuSO_4$) are added in order to increase the speed and the efficiency of the digestion procedure [14].

$$\text{Protein (-N)} + H_2SO_4 = (NH_4)_2SO_4 + CO_2 + H_20 \quad (1)$$

Distillation

In the distillation step, the ammonium ions (NH_4^+) are converted into ammonia (NH_3) by introducing alkali (NaOH), as showed in equation (ii).

$$(NH_4)_2SO_4 + 2NaOH = 2\ NH_3 + Na_2SO_4 + 2\ H_20 \quad (2)$$

The ammonia (NH_3) is captured in absorbing solution like boric acid, sulfuric acid or hydrochloric acid into the receiver vessel by means of steam distillation. Boric acid (H_3BO_3) of 2-4 % concentration is commonly used for capturing the ammonia, forming solvated ammonium ion (equation iii).

$$H_3BO_3 + NH_3 + H_20 = NH_4^+ + B(OH)_4^- \quad (3)$$

Titration

The concentration of captured ammonium ions is determined by either direct titration or back titration. Both methods indicate the ammonia is present in the distillate with a color change and allow for calculation of unknown protein concentrations.

In direct titration, boric acid is used for capturing ammonia by forming ammonia-borate (NH_4^+:$H_2BO_3^-$) complex, which is neutralized by the addition of sulfuric acid, producing a change in color.

In back titration, sulfuric acid is used for capturing ammonia. The residual sulfuric acid (the excess that did not react with NH_3) is titrated with sodium hydroxide standard solution and the amount of ammonia is calculated by difference.

Working procedure

In this work, 0.5 g of hydrolysate sample was taken in a Kjeldahl flask and digested with 15 ml of concentrated sulfuric acid in the presence of a mixture of K_2SO_4 and $CuSO_4$ in the ratio of 5:1 and distilled into a 4% boric acid solution. The formed borate anions were titrated with 0.05M H_2SO_4 by a direct titration procedure which was converted to nitrogen in the sample [13]. The nitrogen content and the crude protein of the raw trimmings were determined by the following formulae [15].

Calculation of nitrogen content (w_n):

$$w_n = \frac{2(v_1 - v_0)c\ M}{m} \quad (4)$$

where W_n is the nitrogen content, in grams per kilogram, of the test sample; V_1 is the volume, in milliliters, of the sulfuric acid required for the determination; V_0 is the volume, in milliliters, of the sulfuric acid required for the blank test; c is the concentration, in moles per liter, of the sulfuric acid used for the titrations; M is the molar mass, in grams per mole, of nitrogen (M 14 g/mol); m is the mass, in grams, of the test portion.

Calculation of crude protein (w_p):

$$w_p = 6.25\ w_n\ g/kg \quad (5)$$

or

$$w_p = 0.625\, w_n\% \tag{6}$$

where W_p is the crude protein content, expressed in grams per kilogram or in percentage.

Yield

The percentage of yield can be derived from the following equation [16]:

$$Percentage\ yield = \frac{Actual\ yield}{Theoretical\ yield} \text{ X } 100 \tag{7}$$

where actual yield is the amount of product obtained from hydrolysis and the theoretical yield is 68% [2].

RESULTS AND DISCUSSION

FTIR analysis

The FTIR spectrum of the protein hydrolysate obtained from untanned raw trimmings of solid waste from tanneries in the region 500-4000 cm^{-1} is shown in Figure 4. Wavelength is in the 3600–3200 cm^{-1} region, resulting from superimposed OH and NH^{3+} stretching bands. Signals at 1645 cm^{-1} and 1541 cm^{-1} correspond to the carbonyl group (C=O) and N–H, which is the evidence of protein structure [17].

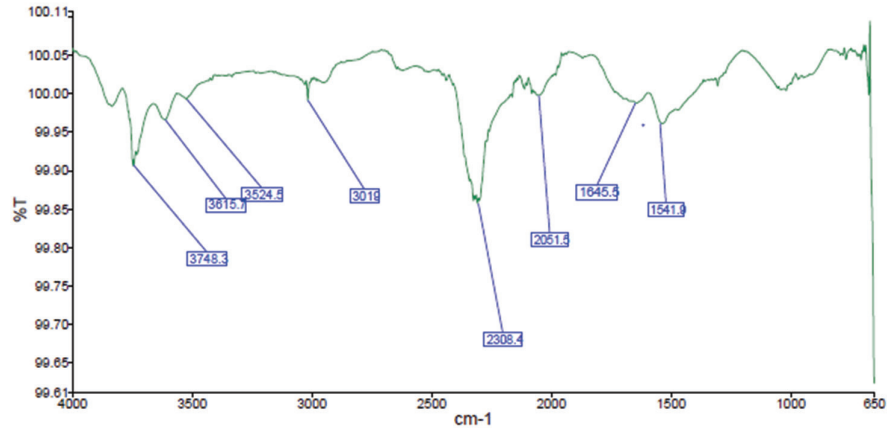
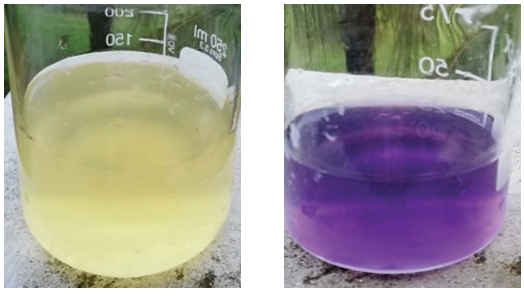


Figure 4. Effect of acetic concentration on protein hydrolysis

Biuret and Kjeldahl methods

Figure 5 shows the change in color of the extracted hydrolysate from grey (a) to purple (b) after the addition of biuret reagent. This change in color represents the presence of proteins in the hydrolysate extracted from raw trimmings.



(a) (b)

Figure 5. Biuret test (a) before addition of biuret reagent, (b) after addition of biuret reagent

Protein hydrolysis of raw trimmings from tanneries by acetic acid was performed with the parameters of acid concentration, acid concentration solution, temperature, and hydrolysis time. Protein concentrations (crude protein) of the extracted hydrolysates were determined by the Kjeldahl method.

Effect of acetic acid concentration on protein hydrolysate

The effect of acetic acid concentration (M) on protein hydrolysis of raw trimmings from tanneries was carried out by hydrolyzing 50 gm of raw trimmings with various acid concentrations (M) with fixed acid solution (5ml) and at a fixed temperature (40 °C). It is shown in Figure 6 that the percentage of protein was increasing with the increase in the concentration of acetic acid from 0.25 to 1.5 molar (M). Since acetic acid is a weak acid and used as a souring agent in vinegar, pickled vegetables etc. in diluted concentrations, it is taken as a hydrolyzing agent in this study. As analytical grade acetic was used in this research, the concentration of acid was not more than 1.5M. The maximum of about 32% protein hydrolysate was extracted from 1.5M acid concentration and the yield was about 23%.

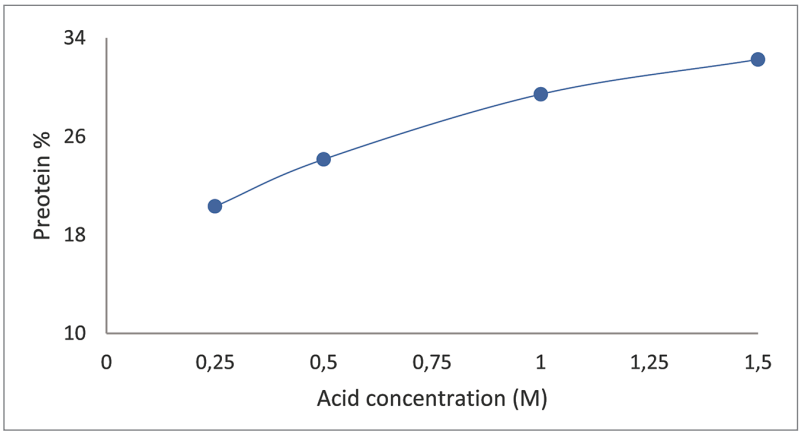


Figure 6. Effect of acetic concentration on protein hydrolysis

Crude protein calculation

Table 2. Volume of H_2SO_4 (ml) required for crude protein determination by titration and yield for effect of acetic acid concentration on hydrolysis

Sample	Volume of 0.05M H_2SO_4 (ml)			Average (ml)	N content (g/kg), Equation (iv)	Crude protein (%), Equation (vi)	Crude protein (g/50gm)	Yield (%), Equation (vii)
	Initial	Final	Diff.					
Blank	0	0	0	0	-	-	-	-
0.25 a	0	11.8	11.8	11.6	32.5	20.3	10.2	14.9
0.25 b	12	23.4	11.4					
0.5 a	30	44	14	13.8	38.6	24.2	12.1	17.8
0.5 b	0	13.6	13.6					
1.0 a	20	36.6	16.6	16.8	47	29.4	14.7	21.6
1.0 b	0	17	17					
1.5 a	25	43.6	18.6	18.4	51.5	32.2	16.1	23.7
1.5 b	0	18.2	18.2					

Effect of concentration solution on protein hydrolysate

The effect of acetic acid concentration solution (ml) on protein hydrolysis of raw trimmings from tanneries is shown in Figure 7. Solution of 5 ml, 10 ml, 15 ml and 20 ml of 1.5M acetic acid were taken for each 50 gm of raw trimmings. From Figure 7, it can be observed that the percentage of yield was increasing with the increase in acid solution and the maximum crude protein percentage was found with 20 ml solution of 1.5M acetic acid and its yield was about 34%.

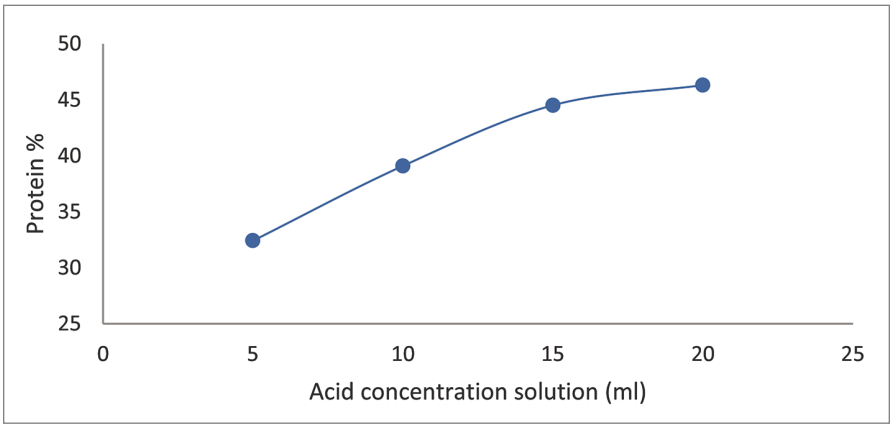


Figure 7. Effect of acetic acid solution on protein hydrolysis

Effect of temperature on protein hydrolysate

Figure 8 shows the effect of temperature on hydrolysis of tannery raw trimmings using best acid concentration (1.5M) and concentration solution (20ml) with temperature variation from 40 to 110^0 C. On heating, trimmings, mainly composed of collagen, are disintegrated and dissolved very quickly. The maximum percentage of protein hydrolysate was obtained about 58 at temperature 80^0 C and yield was about 43%. Temperature above this, the denaturation of protein was occurred and hence yield declined. As raw trimmings are biological materials, the temperature should not increase to high due to avoiding denature of protein.

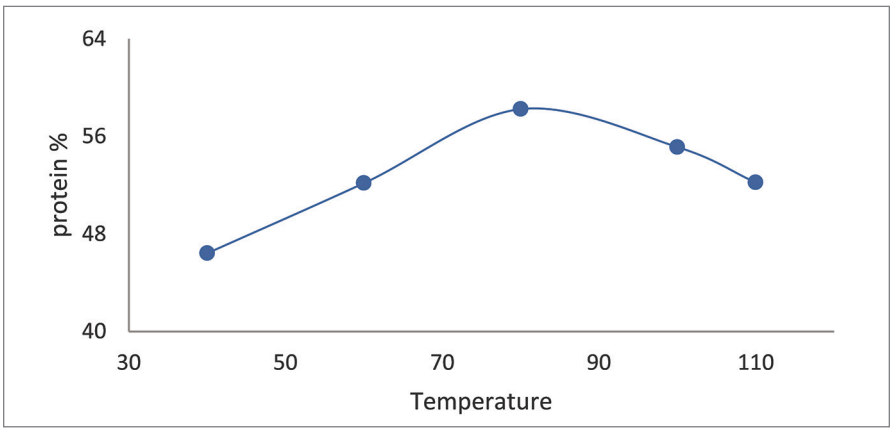


Figure 8. Effect of temperature on protein hydrolysis

Effect of hydrolysis time on protein hydrolysate

Figure 9 shows the effect of hydrolysis time (hour) on hydrolysis with 1.5M acetic concentration, 20 ml acid solution and at 80 °C temperature. About 71% of protein hydrolysate was obtained after 14 hours of hydrolysis and the yield was about 52%. After that the trend of increase was negligible.

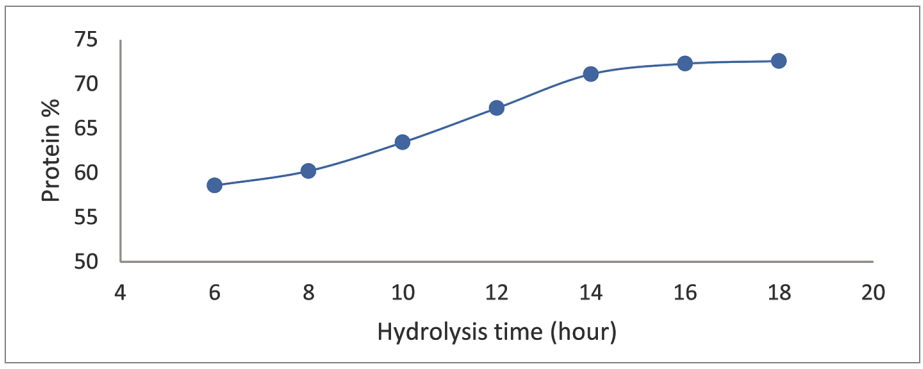


Figure 9. Effect of hydrolysis time on protein hydrolysate

Effect of enzyme concentration (%) on hydrolysis

10 ml solution of 1 to 6% enzyme on the basis of sample weight was prepared for hydrolysis at 1.5 M acetic acid concentration as shown in Figure 10. Other operating parameters, temperature and hydrolysis time, were 80 °C and 4 hours respectively. Among the various enzyme ratios, about 55% protein was extracted from 4% ratio of enzyme and the yield was about 41%. After that, the percentages were negligible.

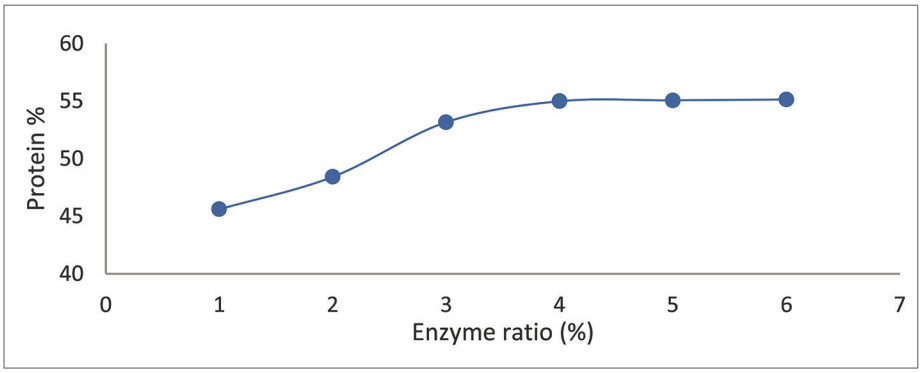


Figure 10. Effect of enzyme concentration (%) on hydrolysis

Effect of enzyme solution (ml) on hydrolysis

Figure 11 shows the effect of enzyme solution on hydrolysis. After the addition of 20 ml enzyme solution, the percentage of protein, which was about 63%, was not increased significantly and the yield was about 46%.

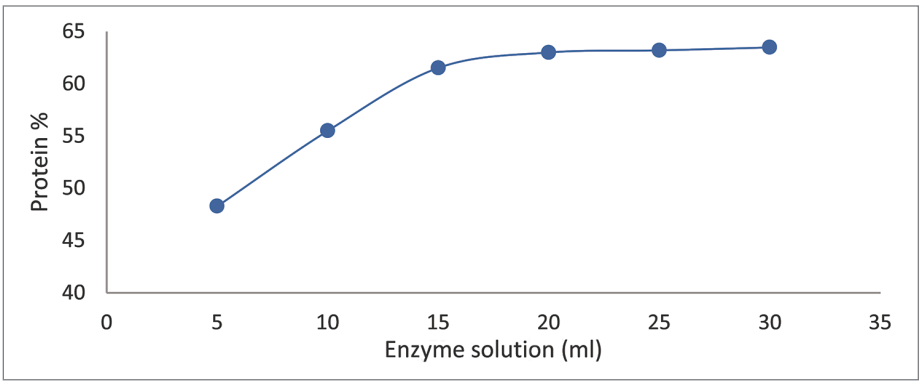


Figure 11. Effect of enzyme solution (ml) on hydrolysis

Effect of heating time on enzymatic hydrolysis

Heating time in enzymatic hydrolysis significantly increases the protein extraction, as shown in Figure 12. The hydrolysis was conducted with the temperature not above 50-60 °C due to enzyme activity in the 1st stage and the final hour used for solidification of the digested sample at 80 °C in the 2nd stage. About 22% of protein hydrolysate was obtained by increasing hydrolysis duration from 4 hours to 6 hours with 1.5 M acetic acid concentration and 20 ml of its solution. The maximum of 88% of protein hydrolysate was obtained after 12 hours of hydrolysis and the yield was about 65%.

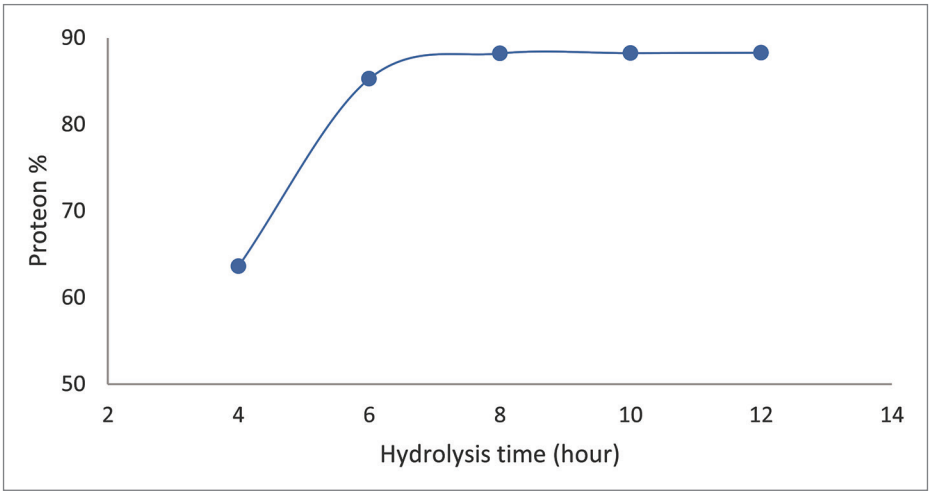


Figure 12. Effect of heating time on enzymatic hydrolysis

Table 3. Crude protein analysis by Kjeldhal method (acetic acid hydrolysis), sample weight 50 gm

Conc. (M)	Solution (ml)	Temp (°C)	Time (h)	Crude protein (%)	Crude protein (g/50gm)	Yield (%)
0.25	5	40	6	20.3	10.2	14.9
0.5	5	40	6	24.2	12.1	17.8
1.0	5	40	6	29.4	14.7	21.6
1.5	5	40	6	32.2	16.1	23.7
1.5	5	40	6	32.4	16.2	23.8
1.5	10	40	6	39.2	19.6	28.8
1.5	15	40	6	44.5	22.2	32.7
1.5	20	40	6	46.2	23.1	33.9
1.5	20	40	6	46.4	23.2	34.1
1.5	20	60	6	52.2	26.1	38.4
1.5	20	80	6	58.3	29.1	42.9
1.5	20	100	6	55.1	27.6	40.5
1.5	20	110	6	52.2	26.1	38.4
1.5	20	80	6	58.6	29.3	43.1
1.5	20	80	8	60.2	30.1	44.3
1.5	20	80	10	63.5	31.8	46.7
1.5	20	80	12	67.2	33.6	49.4
1.5	20	80	14	71.2	35.6	52.4
1.5	20	80	16	72.3	36.1	53.2
1.5	20	80	18	72.5	36.2	53.3

Table 4. Crude protein analysis by Kjeldhal method (acid-enzyme hydrolysis), sample weight 50 gm

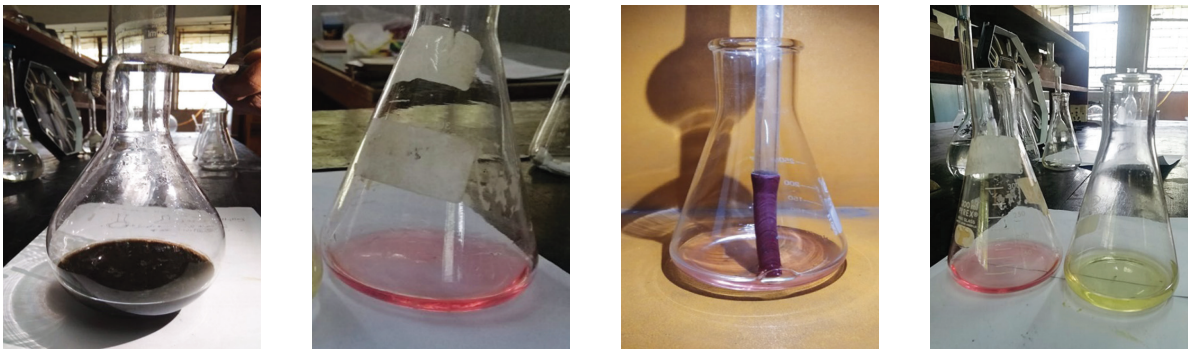
Acid conc. (M)	Enzyme ratio (%)	Enzyme Solution (ml)	1st stage Temp (°C)	1st stage time (h)	2nd stage Temp (°C)	2nd stage time (h)	Crude protein (%)	Crude protein (g/50gm)	Yield (%)
1.5	1	10	50-60	3	80	1	45.7	22.8	33.6
1.5	2	10	50-60	3	80	1	48.3	24.2	35.5
1.5	3	10	50-60	3	80	1	53.2	26.6	39.1
1.5	4	10	50-60	3	80	1	55.3	27.7	40.7
1.5	5	10	50-60	3	80	1	55.1	27.6	40.5
1.5	6	10	50-60	3	80	1	55.1	27.6	40.5
1.5	4	5	50-60	3	80	1	48.3	24.2	35.5
1.5	4	10	50-60	3	80	1	55.3	27.7	40.7
1.5	4	15	50-60	3	80	1	61.6	30.8	45.3
1.5	4	20	50-60	3	80	1	63	31.5	46.3
1.5	4	25	50-60	3	80	1	63.2	31.6	46.5
1.5	4	30	50-60	3	80	1	63.4	31.7	46.6
1.5	4	20	50-60	3	80	1	63.5	31.8	46.7
1.5	4	20	50-60	5	80	1	85.2	42.6	62.7
1.5	4	20	50-60	7	80	1	88.2	44.1	64.9
1.5	4	20	50-60	9	80	1	88.2	44.1	64.9
1.5	4	20	50-60	11	80	1	88.4	44.2	64.9

Some pictures of protein determination by Kjeldhal method are shown in Figure 13 and 14.



(a) (b)

Figure 13. Digestion of raw trimmings (a) at the start of digestion and (b) at the end of digestion



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

Figure 14. Distillation of the digested sample: (a) adding NaOH to the digested sample, (b) boric acid with the methyl red indicator, (c) capturing the NH_3 in boric acid (yellowish color), (d) boric acid solution before and after capturing NH_3

Raw trimmings hydrolysis and the final product are shown in Figure 15.



(a) (b)

Figure 15. (a) raw trimmings hydrolysis and (b) final product

APPLICATIONS OF RAW TRIMMIMGS

The collagen / gelatin extracted from untanned raw trimmings of hides and skins has huge application potential in the field of packaging, biomedicine and cosmetics according to recent studies. The gelatin extracted by acetic acid from raw trimmings was blended with polyvinyl alcohol (PVA) and used to prepare a kind of biodegradable packing material. Here, PVA was used to reduce high water-solubility of gelatin. Collagen burn-healing membrane was prepared from the collagen extracted from pig skin by enzymatic hydrolysis. Hybrid films can be prepared by using collagen extracted from the trimmed skin by acetic acid and then blending with starch/soy protein [18].

Collagen can bind large fat quantities as an emulsifier in meat products. Gelatin is playing a major role in stabilizing ice-cream and other frozen foods. The most common use of gelatin in pharmaceutical industry is covering the outer layer of capsules [19]. Tissue adhesive, vascular grafts, aortic heart valves, drug delivery matrices, wound dressing, and tissue engineering scaffold can be made from collagenous materials derived from leather trimmings [3].

CONCLUSION

Solid wastes generated by the tanning industry pose a great threat to the environment if not properly managed. Untanned raw trimmings do not get any treatment in the tanning process and they can be used as a cheap source of collagen, which has various industrial applications. In this study, the addition of enzyme to the process of hydrolysis with acetic acid accelerated the extraction of protein hydrolysate when compared to hydrolysis with acetic acid alone. About 88% of protein hydrolysate was obtained from enzyme-assisted acid hydrolysis of raw trimmings from tanneries.

Acknowledgments

The authors acknowledge the financial support provided by the Leather Research Institute, Bangladesh Council of Scientific and Industrial Research (BCSIR) under Research and development (R&D) activities.

REFERENCES

[1] Barbe J. The History of Leather Tanning [Internet]. Maharam. Available from: https://www.maharam.com/stories/barbe_the-history-of-leather-tanning.

[2] Bajza Z, Vrcek V. Thermal and enzymatic recovering of proteins from untanned leather waste. Waste Management. 2001; 21(1):79–84. Doi: 10.1016/s0956-053x(00)00039-8

[3] Sundar VJ, Gnanamani A, Muralidharan C, Chandrababu NK, Mandal AB. Recovery and utilization of proteinaceous wastes of leather making: A review. Reviews in Environmental Science and Biotechnology. 2011; 10:151–163. Doi: 10.1007/s11157-010-9223-6

[4] Nur-E-Alam M, Abu Sayid Mia M, Ahmad F, Mafizur Rahman M. Adsorption of chromium (Cr) from tannery wastewater using low-cost spent tea leaves adsorbent. Applied Water Science. 2018; 8:129. Doi: 10.1007/s13201-018-0774-y

[5] Kanagaraj J, Velappan KC, Babu NC, Sadulla S. Solid wastes generation in the leather industry and its utilization for cleaner environment. Cheminform. 2006; 37(49). Doi: 10.1002/chin.200649273

[6] Damrongsakkul S, Ratanathammapan K, Komolpis K, Tanthapanichakoon W. Enzymatic hydrolysis of rawhide using papain and neutrase. Journal of Industrial and Engineering Chemistry. 2008; 14(2):202-206. Doi: 10.1016/j.jiec.2007.09.010

[7] Chakarska I, Goshev I, Idakieva K, Apostolov G. Isolation and characterization of phosphoric acid-soluble collagen from leather wastes of pig breed Bulgarian white. Journal of the Society of Leather Technologists and Chemists. 2006; 90(6):260–263.

[8] Khatoon M, Kashif SR, Saad S, Umer Z, Rasheed A. Extraction of amino acids and proteins from chrome leather waste. Journal of Waste Recycling. 2017; 2:1-4.

[9] Masilamani D, Madhan B, Shanmugam G, Palanivel S, Narayan B. Extraction of collagen from raw trimming wastes of tannery: A waste to wealth approach. Journal of Cleaner Production. 2016; 113:338–344. Doi: 10.1016/j.jclepro.2015.11.087

[10] Simeonova LS, Dalev PG. Utilization of a leather industry waste. Waste Management. 1996; 16(8):765–769. Doi: 10.1016/S0956-053X(97)00020-2

[11] Sathish M, Madhan B, Rao JR. Leather solid waste: An eco-benign raw material for leather chemical preparation – A circular economy example. Waste Management. 2019; 87:357–367. Doi: 10.1016/j.wasman.2019.02.026

[12] Jian S, Wenyi T, Wuyong C. Ultrasound-accelerated enzymatic hydrolysis of solid leather waste. Journal of Cleaner Production. 2008; 16(5):591–597. Doi: 10.1016/j.jclepro.2006.12.005

[13] Chang SK, Zhang Y. Protein Analysis. In: Nieise S, editor. Food Analysis. Food Science Text Series. Springer International Publishing, 2017. p. 315–331.

[14] PanReac AppliChem. Nitrogen Determination by Kjeldahl Method. Available from: https://www.itwreagents.com/uploads/20180114/A173_EN.pdf

[15] ISO 5983-1:2005. Animal Feeding Stuffs-Determination of Nitrogen Content and Calculation of Crude Protein Content-Part 1: Kjeldahl Method. 2005. Available from: https://www.iso.org/standard/39145.html

[16] Zhao W, Yang R, Zhang Y, Wu L. Sustainable and practical utilization of feather keratin by an innovative physicochemical pretreatment: High density steam flash-explosion. Green Chemistry. 2012; 14:3352–3360. Doi: 10.1039/C2GC36243K

[17] Oliveira LCA, Goncalves M, Oliveira DQL, Guerreiro M, Guilherme LRG, Dallago RM. Solid waste from leather industry as adsorbent of organic dyes in aqueous-medium. Journal of Hazardous Materials. 2007; 141(1):344–347. Doi: 10.1016/j.jhazmat.2006.06.111

[18] Li Y, Guo R, Lu W, Zhu D. Research progress on resource utilization of leather solid waste. Journal of Leather Science and Engineering. 2019; 1:1–17. Doi: 10.1186/s42825-019-0008-6

[19] Jayathilakan K, Sultana K, Radhakrishna K, Bawa AS. Utilization of byproducts and waste materials from meat, poultry and fish processing industries: A review. Journal of Food Science and Technology. 2012; 49:278–293. Doi: 10.1007/s13197-011-0290-7

Experimental Techniques
https://doi.org/10.1007/s40799-020-00411-7

RESEARCH PAPER

Estimation and Experimental Validation of Mean-Field Homogenised Effective Properties of Composite

A. Sharma[1] · V. Rastogi[1] · A.K. Agrawal[1]

Received: 27 December 2019 / Accepted: 26 October 2020
© The Society for Experimental Mechanics, Inc 2020

Abstract

For composite materials, the system response changes abruptly with a change in the properties of the material. Therefore, attaining significant knowledge about the effect of the material composition on the material properties is crucial. The researchers are looking for new computational methods which can predict these alterations so that the effort in experimental testing can be reduced. In this direction, this paper presents a robust and novel methodology of validating the estimation of the composite's effective through a multi-scale approach by a set of standardized experimentation. These effective properties are estimated through the mean-field homogenization technique whose parameters are driven from the image analysis of Scanning Electron Microscopy (SEM) images. The predicted results are validated with the results obtained by the experimentation as per ASTM E1876 standard. The estimated error between the predicted properties and the experimental values increased with the increase of alumina particle fraction in the matrix. The mean-field homogenization lags behind the experiments for the parameters defined by the image analysis method than the experimental results. The upper bounds of the mean-field homogenization can be used for the composites with higher reinforcement volume fraction.

Keywords Multi-scale analysis · Mean-field homogenization · Image analysis method · ASTM E1876 · Impact hammer test

Introduction

With the advent of composite materials, industries have started focusing on the use of lighter weight materials with the same mechanical properties. Metal matrix composites have the edge over the parent metals for rotor applications, as it has a higher specific modulus, tensile strength and other mechanical properties. Aluminium /alumina MMCs have shown prominent growth in the composite material market because of their compatibility to the rotor systems. [1–3]. Many manufacturers such as Duralcan, G.K.N., Toyota, GM and Nissan have used MMC materials to develop various components with both static and dynamic applications [4].

Metal matrix composites can be manufactured by various methods from liquid and semi solid-state of matrix mixed with solid reinforcement particles. There are various methods such as powder metallurgy, diffusion bonding, infiltration, stir casting and spray forming to manufacture MMC as per requirements. Stir casting is one of the most inexpensive methods which has been used for the fabrication of metal matrix composites. For complicated and extensive size design of metal matrix composites, stir casting is very convenient and useful [5].

A computation tool that provides results validated by experimental or real-time results saves a lot of time and energy for industries. Researchers seek for more tools that can provide accurate and reliable results for metal matrix composites properties. Multi-scale computation of composites is a two-phase process, where the effective macroscopic properties are derived through micro-mechanics. Analysis of composites with discrete reinforcements with random orientation has been found in archival literature [6–9]. There are various approaches through which these effective properties can be predicted efficiently [10]. For composites with a low volume fraction of inclusions, Eshelby's method [11] with a dilute scheme can be used to evaluate the effective properties. For composites with a higher volume fraction of inclusions, aggregate model [12, 13] can be used, which utilizes the known properties of constituents. Mori-Tanaka [14] have proposed a theory by assuming that the average strain in the inclusion is

✉ A. Sharma
anujsharma_phd2k16@dtu.ac.in

[1] Department of Mechanical, Production, Industrial and Automobile Engineering, Delhi Technological University, Delhi 110042, India

Published online: 23 November 2020

related to the average strain in the matrix by a fourth-order tensor. The relation between the uniform strains in the inclusion in the matrix is provided by this fourth-order tensor. It was further extended by Benveniste [15] to investigate the stress and strain concentration tensors along with the overall elastic modulus of composites. This mean-field homogenization theory has been adopted by various researches [16–20] to determine the effective properties such as modulus, density and thermal expansion and is used in this work.

Significant increase of tensile strength, yield strength and hardness on the addition of alumina particles in the aluminium matrix is validated experimentally in archival literature [21–23]. Properties such as young's modulus, poisson's ratio and shear modulus are computationally derived successfully by various researchers. However, experimental validation of these effective properties with standard practice and technique is equally important. These properties can be experimentally determined by impact hammer test based on ASTM E1876 standard [24].

Recently, *Praveen* et al. [25] performed an impact hammer test based on ASTM E1876 to evaluate the dynamic elastic properties of composites at ambient temperatures. However, the comparison of multi-scale computation of the effective properties and the respective experimental results have not been explored in detail. Daramola et al. [26] compared the mean-field homogenization technique with the experiment and estimated an error of 4.2% between the predicted and experimental values. However, the microscopic parameters such as volume fraction were assumed to be same as the fraction of fibres added during manufacturing. For metal matrix composites fabricated by stir casting method, there is much difference between the added alumina particles and the particles mixed with the matrix. Therefore, a unified and novel approach is adopted in this paper, where the effective properties of the composite are derived computationally using parameters derived through an image analysis method. Moreover, these computationally driven values are compared with the experimental results as per ASTM E1876 standard, evaluated by the impact hammer test.

Material and Method

The composite shaft samples (800 mm length and 22 mm diameter) are fabricated using a stir casting process, as shown in Fig. 1. The aluminium 6061 metal matrix is mixed through stirring with a different weight percentage of alumina (0 wt.%, 6 wt.% and 12 wt.% of alumina) supplied by Loba Chemie Pvt. Ltd. In order to increase the wettability of the alumina particles in the aluminium matrix, the alumina particles were preheated, and magnesium was added during the stir casting process. Magnesium forms an interface of MgO and $MgAl_2O_4$ between the alumina and aluminium, resulting in better wettability and strong bonding [27]. As shown in Fig. 2, the bonding between the reinforcement and the aluminium matrix is strong for both dispersed and agglomerated particles. The Scanning Electron Microscopy (SEM) images are taken for these samples, and then these images are analyzed through an image analysis method. The average aspect ratio and average volume fraction are determined through image analysis. These parameters are used to predict the effective properties of the composite through Mori-Tanaka mean-field homogenization. These properties are then compared with the experimental response of the composite specimen.

Estimation of Effective Mechanical Properties through Mean-Field Homogenization by Using Parameters Extracted through Image Processing

Cubic samples of 5 mm length, 5 mm width and 5 mm height were processed by EDM wire cutting and Scanning Electron Microscopy (SEM) micrographs at the magnification of 500X

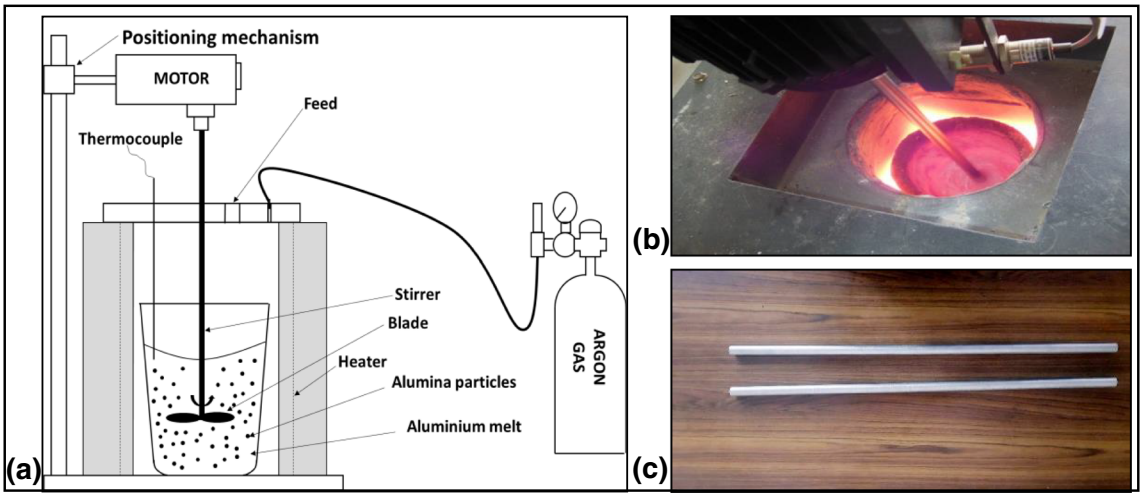


Fig. 1 (**a**) Stir casting set up (**b**) Vortex creation by a four-blade stirrer (**c**) Stir casted 800 x 22 mm shaft

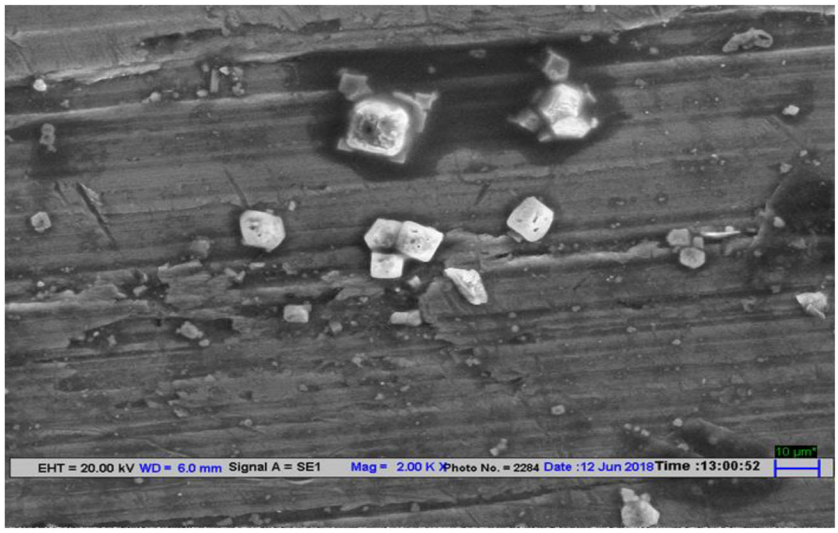


Fig. 2 2000X image showing strong bond between alumina particles and aluminium matrix

and 1000X were obtained using a Zeiss EVO 50 machine. Morphology of the micrographs depicts uniform distribution of reinforcements in the aluminium matrix. For each composition, three samples were scanned, and the fraction and size of alumina particles were evaluated by image processing and averaged for the computation of effective properties.

The open-source image processing software ImageJ was used to generate binary images of the composites with ellipsoid form, assumed for homogenization of properties. SEM micrographs were converted to binary images by using threshold of 61 foreground pixels and 255 background pixels. This transformed the different luminous pixels into white/black bifurcating with respect to the threshold values. At this value of threshold, particles were easily differentiated from the matrix material. Parameters such as the size of the reinforcement, aspect ratio and volume percentage for the composite were derived from the image analysis, as shown in Fig. 3. These parameters were incorporated in the mean-field homogenization of the metal matrix composite material where a Representative Volume Element (RVE) is assumed to be the building block of the macroscopic structure.

RVE is a statistical representation that connects microscopic structure and constituents with the macroscopic properties of composite materials. Therefore, the size of the RVE should

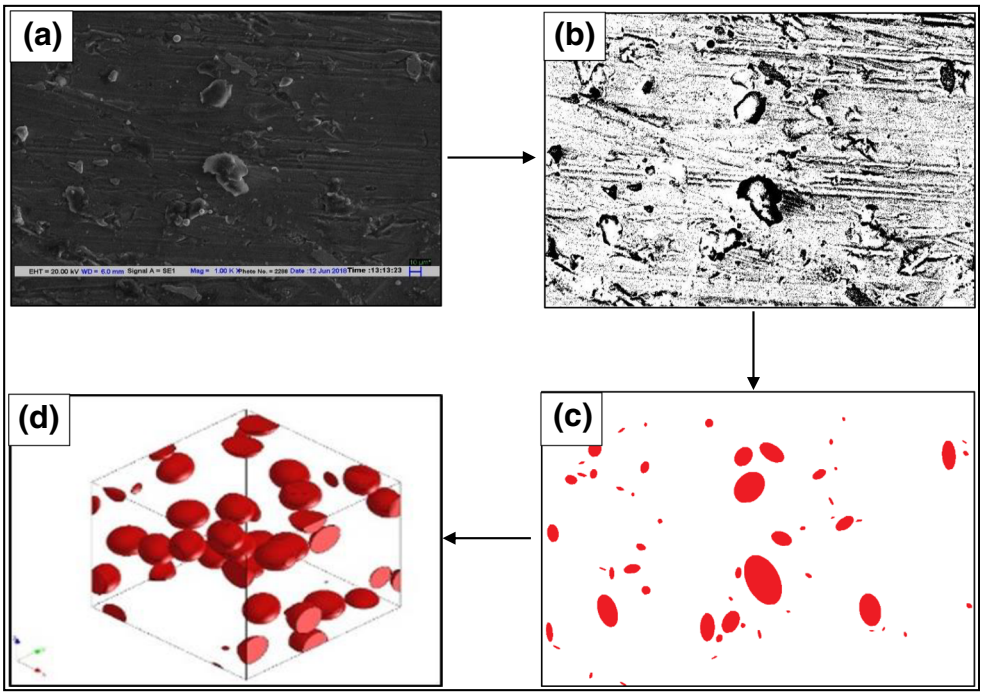


Fig. 3 Conversion of an SEM image into an RVE (**a**) SEM Image of a composite sample (**b**) Cropped SEM image converted into a binary image (**c**) Particle analysation of the binary image (**d**) RVE with mean aspect ratio, size and volume fraction

be such, that it is aptly smaller than the macroscopic structural dimensions and simultaneously large enough to contain an ample amount of information on the microstructure [28, 29].

Digimat is a composite modelling platform which has been used for the generation of RVEs with standard size and threshold. It consists of a Digimat-FE tool, which generates realistic RVEs with the finite element method by using the RVE generation algorithm. This algorithm is based on Random sequential adsorption (RSA) technique which randomizes the orientation of the particles [30]. Particles position themselves in random directions sequentially, and each particle blocks nearby locations from being occupied by another particle, therefore avoiding coagulation. This platform has been used for the development of RVEs in many types of researches. RVEs with randomly distributed spherical inclusions [31, 32], perfectly aligned fibres [33] or random fibre distributions by using a fibre randomization algorithm [34] are generated by using this technique. These RVEs can be exported as standard CAD formats. The effective properties of these RVEs were calculated and have been used for further analysis.

For mean-field homogenization, Mori-Tanaka method [14] has been adopted to determine the effective Young's modulus of the composite material. Therefore, it assumed that the metal matrix composite is comprised of N number of phases. Matrix phase is denoted by subscript m, and there are remaining N-1 inclusion phases in the control volume. The volume fractions of the matrix and the i_{th} inclusion are denoted by ϑ_m and ϑ_i. Similarly, stiffnesses of the matrix and the i_{th} inclusion are denoted by K_m and K_i. These stiffness quantities are generally represented as elasticity tensors of fourth-order which possess specific symmetrical properties. The overall elastic stiffness of the composite [35] is defined in Eq. (1).

$$\bar{\bar{K}}_{av} = \left(\vartheta_m * K_m + \sum_{i=1}^{N-1} \vartheta_i \langle K_i * D_i^{dil} \rangle\right) + \left(\vartheta_m * \boldsymbol{I} + \sum_{i=1}^{N-1} \vartheta_i \langle D_i^{dil} \rangle\right)^{-1} \tag{1}$$

Where $\boldsymbol{I}$ is the identity tensor and the dilute strain concentration factor D_i^{dil} is generally expressed in Eq. (2).

$$D_i^{dil} = \left[\boldsymbol{I} + S_i\, K_m^{-1}(K_i - K_m)\right]^{-1} \tag{2}$$

Here S_i is the Eshelby tensor [6] for the spheroidal inclusion in the infinite matrix.

Since the alumina particles are randomly oriented, orientation averaging has to be taken into account. Therefore, the terms in the angle parentheses $\langle\rangle$ denote the averaged value of the parameters over all possible orientations. This orientationally averaged tensor is a fourth-order tensor in the 3D space [35] and can be expressed as in Eq. (3).

$$\langle T_{klmn} \rangle = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{2\pi} T_{klmn}(\alpha, \beta) sin\beta \, d\alpha \, d\beta \tag{3}$$

where α, β are the generalized angular orientations. For the randomization, the local coordinates should be transformed to the global coordinates. For overall random transformation, the matrix may be expressed as sown in Eq. (4).

$$a_{kl} = \begin{bmatrix} \cos\alpha & \cos\beta \sin\alpha & \sin\beta \sin\alpha \\ -\sin\alpha & \cos\beta \cos\alpha & \sin\beta \cos\alpha \\ 0 & -\sin\beta & \cos\beta \end{bmatrix} \tag{4}$$

Moreover, the transformation may be written, as shown in Eq. (5).

$$T_{klmn}(\alpha, \beta) = a_{ko} a_{lp} a_{mq} a_{nr}\, T'_{opqr} \tag{5}$$

The effective Young's Modulus(E_{efv}) [35] can be now calculated by the general equation (Eq. (6)) with shear modulus and bulk modulus.

$$E_{efv} = E_m * \frac{\kappa_{efv} * \mu_{efv}(3\kappa_m + \mu_m)}{3k_{efv} * \kappa_m + \mu_{efv} * \mu_m} \tag{6}$$

Effective Shear modulus can be derived using Eq. (7).

$$\frac{\kappa_{efv}}{\kappa_m} = \frac{1}{1 + z\, p} \tag{7}$$

Effective Bulk modulus can be derived using Eq. (8).

$$\frac{\mu_{efv}}{\mu_m} = \frac{1}{1 + z\, q} \tag{8}$$

Where E_{efv} is the effective Young's modulus, E_m is Young's modulus of the matrix, κ_{efv} and μ_{efv} are the effective bulk and shear moduli of the composite, z is the volume concentration/fraction, κ_m and μ_m are the bulk and shear modulus of the matrix. Also, p and q are the explicit expressions, and these expressions and details of the expression can be seen in Tandon and Weng [36] for spheroidal inclusions for detailed reference.

Also, effective density can be easily calculated through the rule of the mixture, as shown in Eq. (9).

$$\rho_{efv} = \rho_m \vartheta_m + \rho_i \vartheta_i \tag{9}$$

Where ρ_{efv} is the effective density, ρ_m and ϑ_m are the density and volume fraction of matrix and ρ_i and ϑ_i are the density and volume fraction of inclusions.

Experimental Evaluation of Elastic Properties

For the experimental validation of the estimated/predicted properties of the composites, ASTM E1876 impact hammer vibration analysis was employed. The analysis is performed on a standard experiment setup which conforms the prerequisites for the ASTM E1876 standard, as shown in Fig. 4.

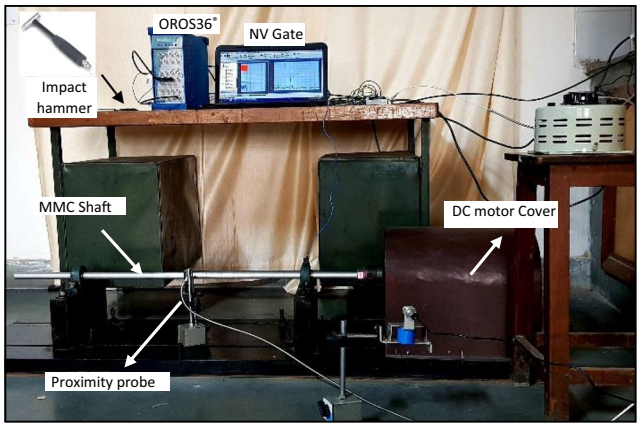


Fig. 4 Experimental set up for impact hammer test

The inductive proximity sensor (LJ12A3–4-Z /BY) having a sensitivity of 1000 V/m and a detection range of 4000 μm is used. The test specimens are energized through a force transducer with a plastic tip. Impact Hammer (Model: - PCB-086C03) with sensitivity 2250 μ(V)/(N) and range 4440 N is used for transient excitation. The response from the transducers is collected and analyzed by using a 32 channel, vibration analyzer OROS36®. The channels of these analyzers are handled in real-time: FFT, 1/3rd Octave. The fundamental natural frequencies of the specimen are observed from the peaks of the frequency response function obtained by using NV Gate® (version 10.1.1) interface.

The prime objective of the experimentation was to extract the values of natural frequencies for the first two bending (flexural) modes of the shaft. A range of 0–600 Hz was targeted as per the dimensions for experimental analysis to confirm that primarily these modes are excited. For excitation, the plastic tip of suitable hardness was used such that the input spectrum doesn't prominently excite the frequencies beyond the working range. The experimental samples were selected carefully such that the impact force input spectrum excites the required range of frequency at amplitude high enough to avoid noise interceptions. The sampling rate of 2.04 kS/s (kilo samples per second) was chosen, which was much higher than the Nyquist frequency according to the range excited by the input spectrum and the range of observation. This high rate of sampling prevented the problem of leakage and aliasing, and no anti-aliasing filters were required. The motive of the experimentation was also to observe the damping of the material relative to each other. The windowing technique would have altered the peak width and thus, the estimation of damping. Therefore, the time range was taken in such a way that windowing was not required for sample selection. The high resolution of the sampling ensured that the natural frequency is confidently determined by Gaussian peak fitting method. The position of the sensor has to be chosen in such a way that it should detect the first and second mode excitation. At the centre of the shaft, the first mode was highly excited, but the second mode was not excited (Since the antinode of the second mode is at the centre for the simply supported system). Therefore, a non-contact inductive proximity transducer was roved, and a suitable position was chosen where the excitation of first and second peaks was obtained successfully (At a distance of 50 mm from the centre as shown in Fig. 5. At this position, the amplitude of first bending natural frequency peak is the highest followed by second bending natural frequency peak.

The Young's modulus expression for the known fundamental flexural natural frequency is shown in Eq. (10).

$$E = 1.6067\left(\frac{l^3}{d^4}\right)\left(mf_f^2\right)T_1 \tag{10}$$

where E = Young's modulus, l = Length of the specimen, d = diameter of the shaft, m = mass of the shaft, f_f = fundamental

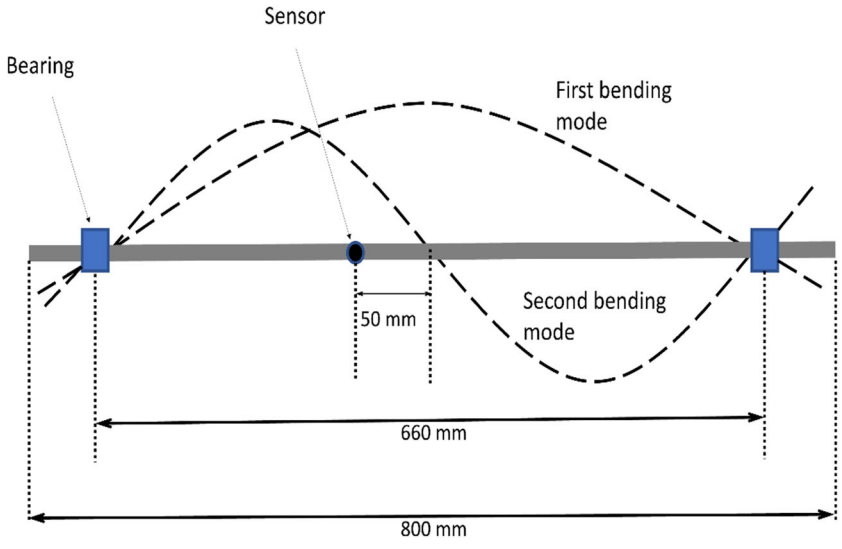


Fig. 5 First and second bending mode with sensor positioning

Table 1 Mean volume fraction, aspect ratio and particle size for various SEM images

Shaft Material		Mean volume fraction			Mean aspect ratio			Mean particle size (μm)		
		S1	S2	S3	S1	S2	S3	S1	S2	S3
Al 6061/6 wt.% Al_2O_3	Max	4.0	4.8	4.7	3.1	2.9	3.1	8.7	8.9	8.8
	Min	2.9	3.4	3.1	2.6	2.6	2.7	7.3	7.2	7.1
	Mean	3.4	4.1	3.8	2.9	2.7	3.0	8.0	8.1	8.1
Al 6061/12 wt.% Al_2O_3	Max	8.3	9.4	9.9	3.4	3.2	3.3	8.9	8.8	9.0
	Min	6.9	7.4	7.9	2.8	2.7	2.9	7.3	7.5	7.4
	Mean	7.3	8.1	9.1	3.2	3.0	3.1	8.3	8.2	8.4

flexural frequency of the shaft and T_1 = correction factor for the fundamental flexural mode to account for finite diameter of the rod, and Poisson's ratio.

Since the aspect ratio (l/d) is higher than 20, the simplified form for T_1 can be used as shown in Eq. (11).

$$T_1 = \left[1 + 4.939\left(\frac{d}{l}\right)^2\right] \tag{11}$$

The expression for Dynamic shear modulus for known fundamental torsion resonant frequency of a cylindrical is shown in Eq. (12).

$$G = 16mf_t^2\left(\frac{l}{\pi d^2}\right) \tag{12}$$

where G = Dynamic shear modulus, l = Length of the specimen, d = diameter of the shaft, m = mass of the shaft, f_t = fundamental torsional frequency of the shaft.

Correspondingly, Poisson's Ratio can be determined using Eq. (13).

$$\mu = \left(\frac{E}{2G}\right) - 1 \tag{13}$$

Where μ is the Poisson's ratio.

Results and Discussion

Since there might be inhomogeneous zones within the sample which lead to interference of the results, measures are taken to avoid these uncertainties. Three samples were taken for each material for SEM analysis. The SEM images were randomly taken for different positions within the sample and all six surfaces of the cuboid sample with 1000x magnifications. For each sample (S1, S2 and S3), maximum, minimum, and mean values of volume fraction, aspect ratio and particle size are taken for various SEM micrographs as shown in Table 1.

The aspect ratio and particle size doesn't vary much for different samples and materials. The mean volume fraction increased for the Al 6061/12 wt.% Al_2O_3 as compared to the Al 6061/6 wt.% Al_2O_3. The variation of the microstructural parameters within the sample were not resulting in interference with the microstructural parameters of different composition. For example, the mean volume fraction ranges from 2.9 to 4.8 for Al 6061/6 wt.% Al_2O_3 and 6.9 to 9 for Al 6061/12 wt.% Al_2O_3. The data is therefore categorical, and we can classify the results for different composition of materials respectively.

The modulus of elasticity is highest for the largest volume fraction, highest aspect ratio and smallest mean particle size and vice versa. Therefore, the overall highest young's modulus can be predicted by taking mean volume fraction as 4.8, aspect ratio as 3.1 and mean particle size as 7.1 μm and vice versa for Al 6061/6 wt.% Al_2O_3. Also, overall highest young's modulus can be predicted by taking mean volume fraction as 9.7, aspect ratio as 3.4 and mean particle size as 9.0 μm and vice versa for Al 6061/12 wt.% Al_2O_3. The consolidated results for the estimated microscopic properties through image analysis and mean-field homogenization are shown in Table 2.

On the addition of 6 wt.% of alumina, the overall mean volume fraction of 3.8% is observed by image analysis of

Table 2 Image analysis overall mean results

Material	Mean volume fraction	Mean aspect ratio	Mean particle size (μm)
Al 6061	–	–	–
6 wt.% Al_2O_3/Al 6061	3.8	2.9	8.1
12 wt.% Al_2O_3/Al 6061	8.166	3.1	8.3

Table 3 Mean-field homogenization results

Engineering constants	Al 6061	6 wt.% Al_2O_3/Al 6061			12 wt.% Al_2O_3/Al 6061		
		Mean	Max	Min	Mean	Max	Min
In-Plane Young's modulus (GPa)	70	73.73	74.84	72.9	78.66	80.1	77.4
Out -Plane Young's modulus (GPa)	70	73.61	74.71	72.78	78.32	79.75	77.06
In-Plane Poisson's ratio	0.328	0.326	0.325	0.326	0.322	0.321	0.323
Transverse Poisson's ratio	0.328	0.326	0.325	0.326	0.322	0.321	0.323
In-Plane shear modulus (GPa)	26.35	27.8	28.21	27.48	29.75	30.29	29.27
Transverse shear modulus (GPa)	26.35	27.76	28.17	27.44	29.2	29.73	28.73
Global Density (Kg/m^3)	2710	2748	2772	2723	2798	2838	2753

SEM images for different samples. The observed volume fraction of alumina in fabricated samples is low as compared to the volume fraction of alumina added during stir casting. The limitation of stir casting method that all alumina particles are not mixed with the matrix is the reason for this deviation. Although stir casting is flexible enough and large-sized samples used for analysis can be easily manufactured through this method. Similarly, for the addition of 12 wt.% of alumina particles, the mean volume fraction of 8.166% is observed. The mean aspect ratio and the average particle size are also in the conformance with the alumina particles manufacturer specification.

The effective Young's Modulus, Poisson's ratio and the shear modulus of in-plane and transverse plane for a particular RVE is estimated through mean-field homogenization as shown in Table 3.

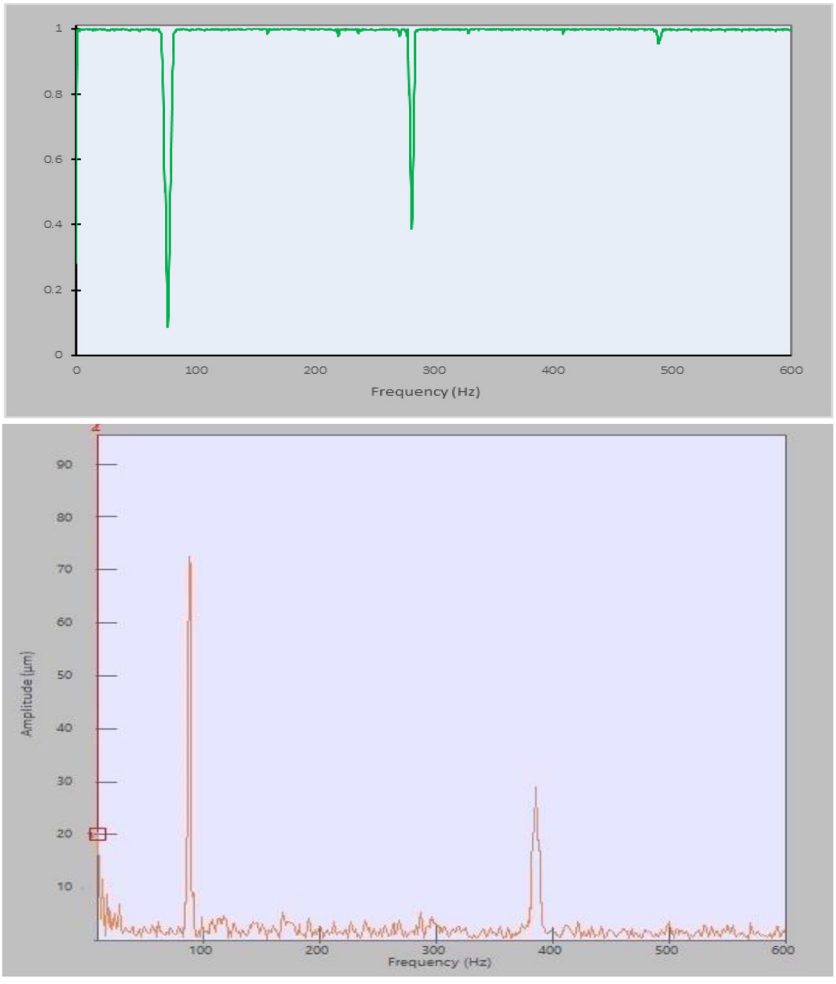


Fig. 6 FRF and coherence for Al6061

The estimated mean In-plane Young's modulus is increased by 5.32%, on the addition of 6 wt.% of alumina due to enhanced dislocation density and precipitation hardening. Since the orientation of the particles is randomized, the value of out of plane Young's modulus is approximately the same as values of In-plane Young's modulus. The global density of the composite is also estimated to be increased by the addition of the particles since the density of the alumina particles is more than the aluminium matrix. The mean In-Plane shear modulus is also increased by 5.5% with transverse shear modulus value approximately similar to the In-plane shear modulus. Moreover, the effective In-plane Young's Modulus is estimated to be increased by 12.37% on addition of 12 wt.% of alumina. The highest value of density and shear modulus is observed for the composite with 12 wt.% of alumina reinforcements.

As shown in Table 3, there is no interference in respective values of Young's modulus, Poisson's ratio and shear modulus for different materials. For example, the minimum value of In-plane young's modulus of 12 wt.% Al_2O_3/Al 6061 is 77.4 GPa, whereas the maximum value of In-plane young's modulus of 6 wt.% Al_2O_3/Al 6061 is 74.84 GPa. So, this data can be classified by the composition of the material.

The strengthening effect of reinforcement is not directionally dependent as in the case of fibre reinforced composites due to the random orientation of particles. By adding 12 wt.% of alumina, In-Plane Young's modulus is increased by 12.37%, but the global density is increased by 3.2%. It implies that the addition of reinforcement improves the specific modulus of the material and therefore improves the modal response of the material. Therefore, impact hammer test is performed to obtain the properties of composites through modal response.

The frequency response function and the coherence for Al6061, Al6061/6 wt.% Al_2O_3 and Al6061/ 12 wt.% Al_2O_3 composite in the impact hammer test is shown in Figs. 6, 7 and 8. The input and output spectrums were coherent for all shafts with the resonance value dropped at anti-nodes. The high coherence between the input and output spectrum, except at the anti-nodes confirms the replicability of the experiments. The high resolution of the frequency spectrum was obtained because of the high sampling rate. The amplitude and the natural frequency were calculated by determining the peak using the Gaussian peak fitting method, and the damping ratio is obtained by the half-power bandwidth method.

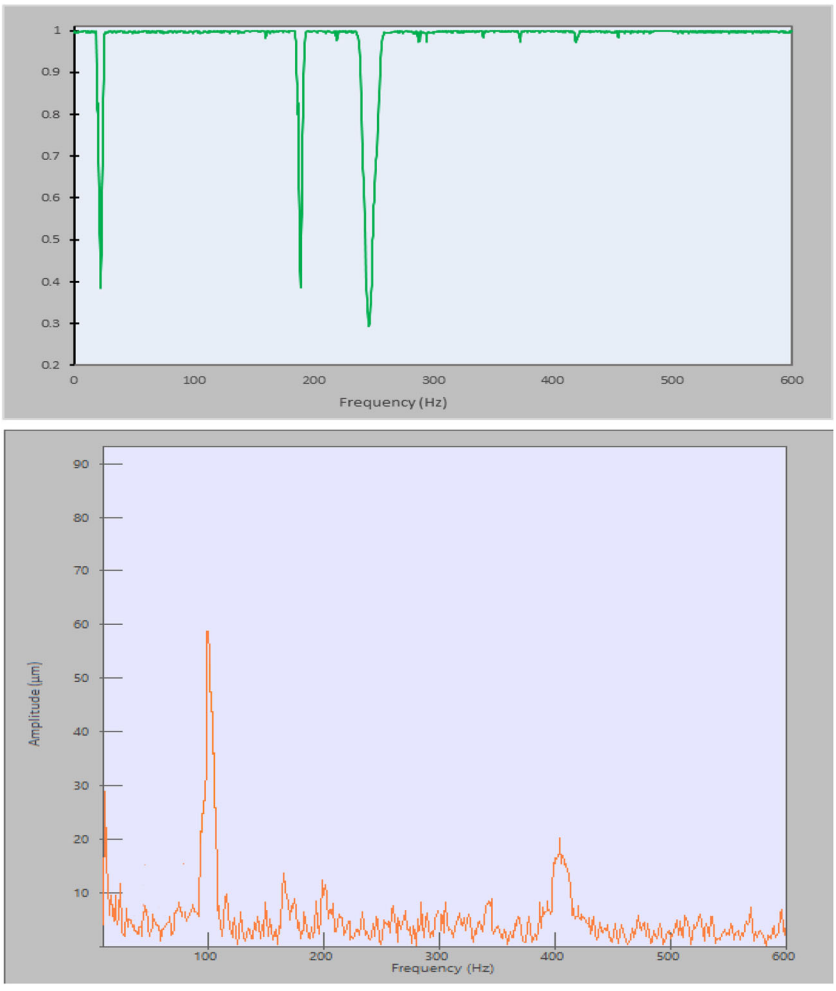


Fig. 7 FRF and coherence for Al6061/ 6 wt. % Al_2O_3

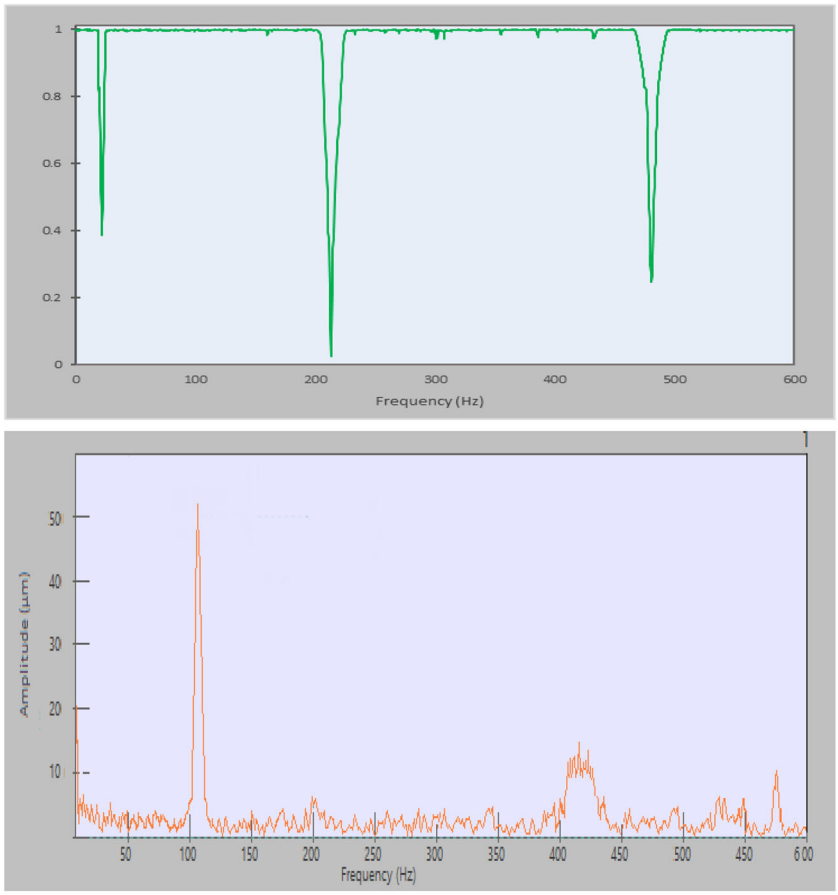


Fig. 8 FRF and coherence for Al6061/ 12 wt. % Al_2O_3

For each shaft with different composition, thirty experiments were taken for the determination of natural frequencies, damping ratio and amplitudes. The mean, standard deviation and range of the natural frequencies for different sets of experimentations are shown in Table 4.

The low standard deviation was obtained for the values of natural frequencies for both first and second mode. Also, the range of natural frequency values for different composition doesn't interfere with each other. The natural frequency increases significantly with the increase of reinforcement particles in the shafts. This is due to the reason that the addition of alumina particles in the Al6061 matrix leads to a substantial improvement in the bending stiffness as compared to the material density. The mean, standard deviation and range of the damping for different sets of experimentations are shown in Table 5.

The higher value of damping was observed for the specimen with higher wt.% of alumina particles in the aluminium matrix. The damping was significantly increased for the second mode of vibration. It suggests that the damping increases with the increase of alumina particles in the aluminium matrix. This is due to the grain refinement and formation of secondary phases which increases elastic strain energy dissipation in the materials.

The mean, standard deviation and range of the amplitude of vibrations for different sets of experimentations are shown in Table 6.

Table 4 Experimental natural frequency mean, standard deviation and range

Shaft Material	Natural frequency (first mode)			Natural frequency (second mode)		
	Mean	Standard deviation	Range	Mean	Standard deviation	Range
Al6061	98.62	0.65	97.6–99.1	388.93	1.92	386.1–391.6
Al 6061/6 wt.% Al_2O_3	101.65	0.72	100.1–102.8	408.43	2.42	406.3–411.7
Al 6061/12 wt.% Al_2O_3	108.91	1.01	107.1–110.7	420.21	3.02	416.2–425.1

Table 5 Experimental damping mean, standard deviation and range

Shaft Material	Damping (first mode)			Damping (second mode)		
	Mean	Standard deviation	Range	Mean	Standard deviation	Range
Al6061	0.0097	0.0001	0.0095–0.0099	0.021	0.0003	0.0207–0.026
Al 6061/6 wt.% Al_2O_3	0.0123	0.0004	0.0119–0.0129	0.0233	0.0005	0.0225–0.0242
Al 6061/12 wt.% Al_2O_3	0.0146	0.0006	0.0137–0.155	0.026	0.0008	0.0252–0.0262

Table 6 Experimental amplitude mean, standard deviation and range

Shaft Material	Amplitude (first mode) (μm)			Amplitude (second mode) (μm)		
	Mean	Standard deviation	Range	Mean	Standard deviation	Range
Al6061	69.1	2.11	67.1–70.9	30.21	2.01	31.2–27.1
Al 6061/6 wt.% Al_2O_3	62.7	2.03	60.7–64.2	21.62	2.42	22.5–19.1
Al 6061/12 wt.% Al_2O_3	52.1	2.81	50.2–54.8	14.29	2.65	12.31–17.01

The amplitude reduced significantly with the increase in alumina particles for both the modes. The range and the standard deviation are well within the limit that is required for the data to be categorical about the composition of the material. This is due to the increase in bending stiffness of the shafts because of the reinforcing effect of the alumina particles in the matrix.

These values are utilized to obtain the value of In-plane Young's modulus, shear modulus and Poisson's ratio experimentally, as per ASTM E1876 standard. The comparison of predicted values obtained through mean-field homogenization and the experimental values obtained is shown in Table 7.

The experimental value of In-plane Young's modulus of Al 6061 is observed to be 2.88% lower than that of initial prediction. This lower value can be attributed to the defects induced during the casting process of aluminium alloy. However, these defects might have prevailed for the case of 6 wt.% Al_2O_3/Al 6061 also, but the experimental In-plane Young's modulus has increased by 1.19% as compared to the predicted value. Similarly, experimental In-plane Young's modulus has increased by 3.75% as compared to the predicted value for the case of 12 wt.% Al_2O_3/Al 6061. From literature [26], it is observed that the addition of 6 wt.% micro-sized kaolinite particle in epoxy polymer matrix composite increased the effective elastic modulus from 3100 MPa to 3107.5 MPa. The mean-field homogenization method predicted the value 4.2% higher than the experimental value. Whereas in the present study, the mean-field homogenization predicted the value 1.19% lower than the experimental values. The reason for predicting a lower value by the homogenization process may be that the actual alumina particle volume fraction is higher than the volume fraction estimated by image analysis.

The range of the experimental and the predicted values for the In-plane Young's modulus is shown in Fig. 9. For Al 6061, the predicted value is the standard value for Young's modulus, and the experiment values are of the cast material. Therefore, the values are lower due to the introduction of some porosity. For, Al 6061/ 6 wt.% Al_2O_3 the experimental range intersects with the predicted value range. For, Al 6061/ 12 wt.% Al_2O_3 the experimental range exceeds the predicted

Table 7 Comparison of Predicted and experimental value

Engineering constants	Al 6061		6 wt.% Al_2O_3/Al 6061		12 wt.% Al_2O_3/Al 6061	
	Predicted	Experimental	Predicted	Experimental	Predicted	Experimental
In-Plane Young's modulus	70 GPa	67.98 GPa	73.73 GPa	74.61 GPa	78.66 GPa	81.61 GPa
In-Plane shear modulus	26.35 GPa	25.55 GPa	27.82 GPa	29.01 GPa	29.75 GPa	30.91 GPa
In-Plane Poisson's ratio	0.328	0.327	0.326	0.323	0.322	0.320



Fig. 9 Comparison of range of predicted and experimental value of In-plane Young's modulus

value range at a few points. Therefore, it can be concluded that the upper bounds of mean-field solutions should be used for the higher reinforcement compositions.

The similar kind of trend is observed for In-plane shear modulus, and the mean-field homogenization analysis does not deviate significantly from the experimental results.

Conclusion

The composite specimens have been successfully fabricated by stir casting method as per ASTM E1876 standards. The microstructure of the composites was taken through SEM and particles average size; aspect ratio and fraction were obtained through image analysis method. The effective properties were then evaluated by mean-field homogenization method by using these parameters. These properties were then evaluated experimentally as per ASTM E1876 standards. Following conclusions can be driven from the analysis in this paper:

- The volume fraction of the alumina particles observed by the image analysis method was significantly less as compared to the added volume fraction during fabrication. It suggests that the volume fraction added initially during fabrication should not be used directly for predicting the effective properties.
- The addition of alumina particles results in the strengthening of composite specimens due to enhanced dislocation density and precipitation hardening.
- The difference between the In-Plane and transverse effective properties is negligible since the strengthening effect of reinforcement is not directionally dependent due to the random orientation of particles.
- The addition of alumina particles in the Al6061 matrix leads to a substantial improvement in the bending stiffness as compared to the material density. Due to this reason, the natural frequency for various modes of vibration increased on the addition of particles.
- Damping ratio also increased on the addition of alumina particles due to the grain refinement and formation of secondary phases.
- The estimated error for the case of 6 wt.% Al_2O_3/Al 6061 is 1.19%, whereas the estimated error for the case of 6 wt.% Al_2O_3/ Al6061 is 3.75% for the mean values.
- The range of mean-field values lags behind the experimental values as the material of high reinforcement composition is used. It can be concluded that the upper bound of mean-field homogenization can be used for the material of high reinforcement composition and can be the future scope of work. However, mean-field homogenization predicts a low higher than experimental value while assuming that all alumina particles added for process are mixed into the matrix.
- Also, samples were taken from the end of the shaft for SEM analysis; this might be the low/high composition area of the shaft during casting. This variation can be addressed in future work.
- The variation of the experimental and the predicted values was such that categorization of the values was possible in reference to the composition of the shaft.

Author's Contribution All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Mr. Anuj Sharma, Prof. Vikas Rastogi and Prof. Atul kumar agrawal The first draft of the manuscript was written by Mr. Anuj Sharma and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they do not have any conflict of interest.

References

1. Chawla N, Shen YL (2001) Mechanical behavior of particle reinforced metal matrix composites. Adv Eng Mater 3(6):357–370. https://doi.org/10.1002/1527-2648(200106)3:6<357::AID-ADEM357>3.0.CO;2-I
2. Rawal S (2001) Metal-matrix composites for space applications. Jom 53(4):14–17. https://doi.org/10.1007/s11837-001-0139-z
3. Sarmasti AB, Yazdanirad M, Nouri Khezrabad M, Karbalaie M (2011) Effect of alumina particle size and thermal condition of casting on microstructure and mechanical properties of stir cast Al-Al2O 3 composites. Mater Sci Technol 27(11):1653–1656. https://doi.org/10.1179/1743284710Y.0000000020
4. Koczak MJ, Khatri SC, Allison JE, Bader MG (1993) Metal-Matrix Composites for Ground Vehicle, Aerospace, and Industrial Applications. Fundamentals of Metal-Matrix Composites, 297–326. https://doi.org/10.1016/b978-0-08-052371-2.50020-1
5. Singh S, Singh I, Dvivedi A Design and development of novel cost-effective casting route for the production of metal matrix composites (MMCs) (2017). Int J Cast Metals Res 6(30):356–364. https://doi.org/10.1080/13640461.2017.1323605

6. Chaturvedi SK, Tzeng GY (1991) Micromechanical modeling of material damping in discontinuous fiber three-phase polymer composites. Compos Eng 1(1):49–60. https://doi.org/10.1016/0961-9526(91)90025-N
7. Guden M, Hall IW (1998) Dynamic properties of metal matrix composites: a comparative study. Mater Sci Eng A 242(1–2): 141–152. https://doi.org/10.1016/s0921-5093(97)00516-9
8. Lesieutre GA, Yarlagadda S, Yoshikawa S, Kurtz SK, Xu QC (1993) Damped structural composite material using resistively-shunted piezoelectric ceramic short fibers. Collect Tech Papers - AIAA/ASME Structures, Structural Dynamics Mater Conf (pt 6): 3238-3243
9. Vaidya UK, Dadzie P-E, Haque A, Mahfuz H, Jeelani S (1997) Nondestructive evaluation and characterization of microfibers modified textile carbon-phenolic and carbon-carbon composites. J Reinf Plast Compos 16(11):968–1001. https://doi.org/10.1177/073168449701601101
10. Mura T (2013) Micromechanics of defects in solids. Mechanics of Elastic and Inelastic Solids. Springer Science & Business Media. https://doi.org/10.1007/978-94-009-3489-4
11. Eshelby JD (1957) The determination of the elastic field of an ellipsoidal inclusion, and related problems. Proc Royal Soc A: Math Phys Eng Sci 241(1226):376–396. https://doi.org/10.1098/rspa.1957.0133
12. Christensen RM, Waals FM (1972) Effective stiffness of randomly oriented fibre composites. J Compos Mater 6(3):518–535. https://doi.org/10.1177/002199837200600307
13. Halpin JC, Jerine K, Whitney JM (1971) The laminate analogy for 2 and 3 dimensional composite materials. J Compos Mater 5(1):36–49. https://doi.org/10.1177/002199837100500104
14. Mori T, Tanaka K (1973) Average stress in matrix and average elastic energy of materials with misfitting inclusions. Acta Metall 21(5):571–574. https://doi.org/10.1016/0001-6160(73)90064-3
15. Benveniste Y (1987) A new approach to the application of Mori-Tanaka's theory in composite materials. Mech Mater 6(2):147–157. https://doi.org/10.1016/0167-6636(87)90005-6
16. Abaimov SG, Trofimov A, Sergeichev IV, Akhatov IS (2019) Multi-step homogenization in the Mori-Tanaka-Benveniste theory. Compos Struct. https://doi.org/10.1016/j.compstruct.2019.03.073
17. Imani SM, Goudarzi AM, Rabiee SM, Dardel M (2018) The modified Mori-Tanaka scheme for the prediction of the effective elastic properties of highly porous ceramics. Ceram Inter 44(14):16489–16497. https://doi.org/10.1016/j.ceramint.2018.06.066
18. Kamarian S, Shakeri M, Yas MH (2016) Natural frequency analysis and optimal design of CNT/fiber/polymer hybrid composite plates using Mori-tanaka approach, GDQ technique, and firefly algorithm. Polym Compos 39(5):1433–1446. https://doi.org/10.1002/pc.24083
19. Tran VP, Brisard S, Guilleminot J, Sab K (2018) Mori–Tanaka estimates of the effective elastic properties of stress-gradient composites. Int J Solids Struct 146:55–68. https://doi.org/10.1016/j.ijsolstr.2018.03.020
20. Sobhaniaragh B, Batra RC, Mansur WJ, Peters FC (2017) Thermal response of ceramic matrix nanocomposite cylindrical shells using Eshelby-Mori-Tanaka homogenization scheme. Compos Part B 118:41–53. https://doi.org/10.1016/j.compositesb.2017.02.032
21. Yılmaz O, Buytoz S (2001) Abrasive wear of Al_2O_3-reinforced Aluminium-based MMCs. Compos Sci Technol 61:2381–2392
22. Kok M (2005) Production and mechanical properties of Al_2O_3 particle-reinforced 2024 Aluminium alloy composites. J Mater Process Technol 161:381–387
23. Suthar J, Patel KM (2018) Processing issues, machining, and applications of aluminium metal matrix composites. Mater Manuf Process 33(5):499–527. https://doi.org/10.1080/10426914.2017.1401713
24. Standard Test Method for Dynamic Young's Modulus, Shear Modulus, and Poisson's Ratio by Impulse Excitation of Vibration. ASTM E1876, ASTM International, United States, 1–17
25. Praveen AP, Rajamohan V, Mathew AT (2019) Material and mechanical characterization of multi-functional carbon nanotube reinforced hybrid composite materials. Exp Tech 43(3):301–314. https://doi.org/10.1007/s40799-019-00316-0
26. Daramola OO, Olajide JL, Adediran AA, Adewuyi BO, Ayodele TT, Desai DA, Sadiku ER (2020) Multiscale analysis and experimental validation of the effective elastic modulus of epoxy-dioctahedral phyllosilicate clay composite. Heliyon 6(6):e04008. https://doi.org/10.1016/j.heliyon.2020.e04008
27. Sangghaleh A, Halali M (2009) Effect of magnesium addition on the wetting of alumina by aluminium. Appl Surf Sci 255:8202–8206. https://doi.org/10.1016/j.apsusc.2009.05.044
28. Bouhfid N, Raji M, Boujmal R, Essabir H, Bensalah M-O, Bouhfid R, Qaiss AeK. (2019) Numerical modeling of hybrid composite materials. Modelling of Damage Processes in Biocomposites, Fibre-Reinforced Composites and Hybrid Composites, 57–101. https://doi.org/10.1016/b978-0-08-102289-4.00005-9
29. Gitman IM, Askes H, Sluys LJ, Valls OL (2004) The concept of representative volume for elastic, hardening and softening materials. International Summer School - IPME conference. XXXII International Summer School - IPME Conference: Advanced Problems in Mechanics. p 1-5
30. Babu KP, Mohite PM, Upadhyay CS (2018) Development of an RVE and its stiffness predictions based on mathematical homogenization theory for short fibre composites. Int J Solids Struct 130-131:80–104. https://doi.org/10.1016/j.ijsolstr.2017.10.011
31. Klusemann B, Svendsen B (2010) Homogenization methods for multi-phase elastic composites: comparisons and benchmarks. Tech Mech 30(4):374–386
32. Yi I, Wiedmaier J, Schmauder S (2015) Modeling and simulation of mechanical properties of nano particle modified polyamide 6. J Mater Sci Chem Eng 3(1):80–87
33. Ogierman W, Kokot G (2016) A study on fiber orientation influence on the mechanical response of a short fiber composite structure. Acta Mech 227(1):173–183
34. Soni G, Singh R, Mitra M, Falzon BG (2014) Modelling matrix damage and fibre matrix interfacial decohesion in composite laminates via a multi-fibre multilayer representative volume element (M2RVE). Int J Solids Struct 51(2):449–461
35. Chang CY, Chang MY, Huang JH (2004) Vibration analysis of rotating composite shafts containing randomly oriented reinforcements. Compos Struct 63(1):21–32
36. Tandon GP, Weng GJ (1986) Average stress in the matrix and effective moduli of randomly oriented composites. Compos Sci Technol 27(2):111–132. https://doi.org/10.1016/0266-3538(86)90067-9

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

Experimental and computational analysis of mechanical properties of RGD840 material manufactured through PolyJet process

Nagendra Kumar Maurya
Department of Mechanical Engineering, GL Bajaj Institute of Technology and Management, Greater Noida, India, and
Vikas Rastogi and Pushpendra Singh
Department of Mechanical Engineering, Delhi Technological University, Delhi, India

Abstract

Purpose – Nowadays, the PolyJet technique is used to fabricate low volume functional parts in engineering and biomedical applications. However, the mechanical properties of the components fabricated through this process are inferior in comparison to components fabricated through the traditional manufacturing process. This paper aims to attempt to investigate the influence of process parameters, i.e. raster angle, orientation and type of surface finish on mechanical properties of RGD840 material manufactured by the PolyJet process.

Design/methodology/approach – Initially, this study focuses on experimental evaluation of elastic modulus, ultimate tensile strength and percentage elongation of the material. Further detailed experimental study of true stress, true strain, and plastic strain are conducted. Computational analysis of plastic strain is performed by using finite element analysis (FEA) software ABAQUS. The value of strength coefficient (K) and strain hardening coefficient (n) is calculated by using the graphical method from the true stress-plastic strain curve.

Findings – It is observed that 90° raster angle, flat orientation and glossy surface are the best level of process parameters for the tensile strength, true stress and modules of elasticity of the RGD840 material and the obtained value are 27.88, 30.134 and 2891.5 MPa, respectively. The percentage elongation is maximum at 60° raster angle, flat orientation, and matte finish type and the obtained value is 23.38%. The optimum level of process parameters are 90° raster angle, Flat orientation, with Glossy surface finish. SEM analysis of the fracture surface of the tensile test specimen proves that the fracture surface is brittle in nature.

Originality/value – The novelty of this work lies in the fact that no attempts were made to investigate the computational investigation of mechanical properties of RGD840 material.

Keywords Rapid prototyping, Tensile strength, Elastic modulus, Plastic strain, PolyJet process, Tensile strength, Elastic modulus, Plastic strain

Paper type Research paper

1. Introduction

Nowadays, the rapid prototyping (RP) technique is emerging as an alternate method of product manufacturing for engineering and biomedical applications (Suresh *et al.*, 2018). Several AM technologies are developed, which works on a different principles (Maurya *et al.*, 2019). PolyJet components are used for fit and functional prototypes, over-mold samples, medical development, full multi-colored marketing models, medical models, consumer goods, automotive and electronic industries, etc. (Ainsworth *et al.*, 2000; Singh and Singh, 2009).

PolyJet 3D printing technology was developed by Israeli company Objet geometries Ltd in 2000. It has capability to replace the Stereo Lithography (SLA) technology (Maurya *et al.*, 2020). In this method photopolymers are used as build as well as support material. Inkjet head is use to deposit the build material as well as support material according to the predefined program. A UV flood lamp is mounted on the print head to complete cure after depositing each layer of the build material (Kumar and Kumar, 2015). Support material is removed by washing with pressurized water in the post processing operation. The resolution of parts printed by PolyJet process is equal to the SLA systems (Liu *et al.*, 2020). There are several materials suitable available for the PolyJet technologies. The drawback of this technology is that print heads are very expensive and needed to be replaced regularly. The schematic diagram for PolyJet system is shown in Figure 1 (Pugalendhi *et al.*, 2020).

However, the mechanical properties of additive manufactured components are poorer as compared to the conventional manufacturing process (Singh, 2011). From archival literature, it was observed that mechanical properties of AM printed components depend on process parameters, material and printing technology (Tabi *et al.*, 2015; Sagias *et al.*, 2018). Among all the available AM technology, the PolyJet process has superior dimensional accuracy (DA), flatness, surface roughness (SR) and mechanical properties

The current issue and full text archive of this journal is available on Emerald Insight at: **https://www.emerald.com/insight/1355-2546.htm**



Rapid Prototyping Journal
© Emerald Publishing Limited
[ISSN 1355-2546]
[DOI 10.1108/RPJ-03-2020-0049]

Received 6 March 2020
Revised 5 July 2020
15 October 2020
Accepted 31 October 2020

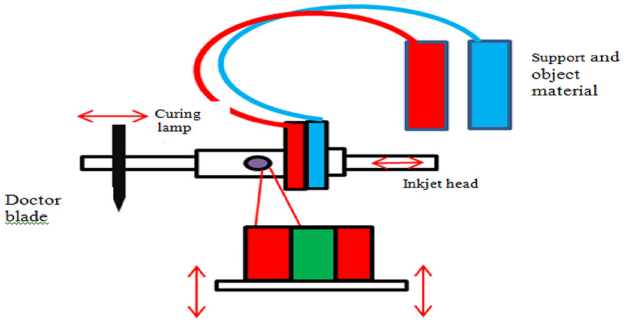
(Gaya *et al.*, 2015; Moore and Williams, 2015). Several studies have been conducted to evaluate the effect of PolyJet process parameters on mechanical properties such as yield strength, ultimate tensile strength, shear strength and modulus of elasticity (Zelený *et al.*, 2014; Cazon *et al.*, 2014).

Gay *et al.* (2015) have experimentally investigated the effect of process parameters on relaxation modulus (Et) of RGD240 material of flat manufactured component through Objet 30 machine. The process parameters, i.e. part spacing along *x*-axis, *y*-axis, within the tray and surface quality were used. Relaxation modulus was reduced about 25% along the measured period of 5 s. Kesy and Kotlinski (2010) have evaluated the influence of orientation on mechanical properties such as tensile strength, percentage elongation and density of PolyJet printed components. Results revealed that components printed along the orientation of *x*-axis had higher tensile strength, hardness and density. Barclift and Williams (2012) have conducted the effect of process parameters, namely, the out-of-build plane part orientation (Z), in-build plane part orientation (X-Y) and the distance between specimens on tensile strength and tensile modulus of the PolyJet printed components. It was observed that part spacing has the largest effect on tensile strength. Pilipović *et al.* (2009) have experimentally investigated the flexural strength, tensile strength, surface roughness and dimensional accuracy of VeroBlack, VeroBlue and FullCure 720 materials. Objet Eden 330 and ZPrinter 310 Plus PolyJet printer were used for the manufacturing of test specimens. Results revealed that dimensional accuracy of the Objet Eden 330 was more precise in manufacturing than ZPrinter 310 Plus. Surface roughness of VeroBlack and VeroBlue materials were superior over the zp 102 powder material. The test specimens made of FullCure 720 material have highest tensile strength.

Ge *et al.* (2020) have applied coordinated multiscale approach for comparing and validating the mechanical properties. A DEM simulation was performed by using TBBM models to evaluate the mechanical properties. FEM simulation was performed under the limit of linear elastic deformation region only. Experimental investigation of quasi-static uniaxial compression tests were conducted and compared to DEM simulations of the agglomerate deformation behavior.

In present study, specimen samples were prepared by using RGD840 (VeroBlue) material. It is a cartridge containing ink, which can be used under normal condition. The substance is released only from the cartridge inside printing system. Hence, external exposure is limited. After printing the component, nontoxic support material is removed from the surface of build geometry through pressurized water. In industries, RGD840 (VeroBlue) material is used for the checking of form and fit testing, moving and assembly parts, fabrication of exhibition models, silica molding, assembly of electronic component, etc. (Kumar and Kumar, 2015). The effects of process parameters such as raster angle, orientation and surface type on tensile strength, elastic modulus and percentage elongation have been evaluated. Further, true stress, true strain and plastic strain have been investigated at optimum level of process parameters. Computational analysis of plastic strain was evaluated through ABAQUS software.

Figure 1 PolyJet 3 D printing system



Source: Pugalendhi *et al.* (2020)

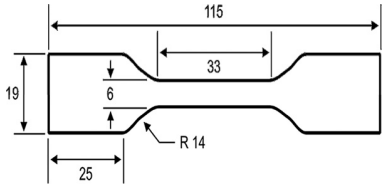
2. Materials and method

In this investigation, mechanical properties, i.e. elastic modulus, ultimate tensile strength, true stress, true strain and plastic strain of RGD840 material, were evaluated. For the evaluation of tensile behavior of the material, specimens were fabricated according to ASTM D638 standards (Pugalendhi *et al.*, 2020). The test specimen for the tensile test has been designed by using CATIA V6 software and converted into an STL file by using 0.0001 convergence tolerance. The dimensional details of the test specimen are presented in Figure 2. Samples were manufactured by using Objet 30 PolyJet machine. In the PolyJet process, two different materials are used. First is the primary material used to fabricate components and second is the support material, which is used to stabilize the hollow portion of the component. Subsequently, the support material is removed through a pressurized water jet. In the present investigation, RGD840 VeroBlue was used as a building material, whereas SUP705 was used as support material. A constant layer thickness of 28 μm was used to fabricate all samples. The details of experimental plan are discussed in the subsequent subsections.

2.1 Process parameters and their levels

The present study deals with the effects of process parameters, namely, raster angle, orientation and surface type on ultimate tensile strength, elastic modulus and percentage elongation of the material. Further, true stress, true strain and plastic strain have been evaluated at the optimum level of process parameters. The complex relation between UV irradiation patterns and manufacturing strategy recommends the presence of non-linear effects. However, the possible interaction between the process parameters should be challenging to avoid. Keeping these facts into account, all the possible process parameters combinations were used for the experimentation. Table 1 illustrates the process variables and their levels. In this study, the raster angle was considered at four levels, i.e. 0°, 30°, 60° and 90°. Process parameter (orientations) was considered

Figure 2 Dimension of the test specimen



at two levels only, due to the height limitation along 'Z' direction. The levels of process parameters (orientations) used in this study were flat orientation and on-edge orientation. Surface type was also considered at two levels only, due to the machine limitation. In this machine, only two types of finish products can be obtained namely glossy and matte finish type.

Table 2 illustrates the experimental plan for the investigation of tensile properties of the material. To check the repeatability of results, four samples were taken for each trial, and the average value of four samples was considered. In this study, 16 runs were used for the conduction of experiments; hence, a total of 64 samples were fabricated. Figure 3(a) displays the raster direction on PolyJet build tray. Angles are measured from the *x*-axis. Raster angle represents the printing direction of the component on the build platform. Figure 3(b) demonstrates the raster angle and Figure 3(c) shows the orientation of fabricated samples on the Objet 30 PolyJet machine tray. Figure 3(d) illustrates the type of surface finish of the manufactured samples through the Objet 30 PolyJet machine.

In glossy surface, support material does not stick on the surface of the fabricated samples. From Figure 3(d), it can be visualized that the matte finish surface contains support material on its surface. In the removal process, support material leaves burrs on the surface of the fabricated sample, which results in poor surface quality in the fabricated component. A computerized UTM Logic Controller machine was used to conduct the tensile test of the specimen as shown in Figure 4. In this study, an instantaneous constant strain rate, i.e. 0.1 mm/min, was used. All the experiments were performed under a controlled environment of 20°C.

Table 1 Factor and their levels

S.N	Factors	Level 1	Level 2	Level 3	Level 4
1	Raster angle	0°	30°	60°	90°
2	Orientation	Flat	on-edge	–	–
3	Surface Finish	Glossy	Matte	–	–

Table 2 Experimental plan for tensile properties of the material

S. N	Fixed process parameters: Orientation	Fixed process parameters: Surface type	Variable process parameter: Raster angle	Ultimate tensile strength (MPa)	Modulus of elasticity (MPa)	% elongation
1	Flat	Glossy	0°	26.87	2658.5	21.65
2	Flat	Matte	0°	23.56	2533.9	15.83
3	On-edge	Glossy	0°	20.58	1878.4	18.55
4	On-edge	Matte	0°	20.78	1628.2	21.98
5	Flat	Glossy	30°	22.54	2219.4	20.00
6	Flat	Matte	30°	22.84	2182.6	21.32
7	On-edge	Glossy	30°	21.34	1752.4	18.42
8	On-edge	Matte	30°	20.82	1672.6	16.32
9	Flat	Glossy	60°	24.86	2232.5	21.32
10	Flat	Matte	60°	24.46	2165.6	23.38
11	On-edge	Glossy	60°	22.55	1584.8	19.74
12	On-edge	Matte	60°	21.86	1534.4	18.42
13	Flat	Glossy	90°	27.85	2891.5	12.08
14	Flat	Matte	90°	26.34	2867.3	16.43
15	On-edge	Glossy	90°	22.25	1965.5	14.64
16	On-edge	Matte	90°	21.56	2025.6	11.62

2.2 Analysis of plastic strain of RGD840 material

The plastic strain was evaluated experimentally as well as computationally. For this purpose, the ASTM D-638 tensile test specimen was used to evaluate the plastic strain. FEA software ABAQUS was used to investigate the plastic strain computationally. A constant displacement of 0.1 mm/min was applied. The details of experimental and computational studies are discussed in the subsequent subsections.

2.2.1 Experimental investigation of plastic strain

The plastic strain behavior of RGD840 material was investigated from the experimental data of the load-deflection diagram. The load-deflection curve of the material

Figure 3 (a) Raster directions on PolyJet platform; (b) raster angle on Objet 30 PolyJet machine tray of the fabricated samples; (c) orientation of fabricated samples on Objet 30 PolyJet machine tray; (d) type of surface finish of fabricated samples on Objet 30 PolyJet machine tray

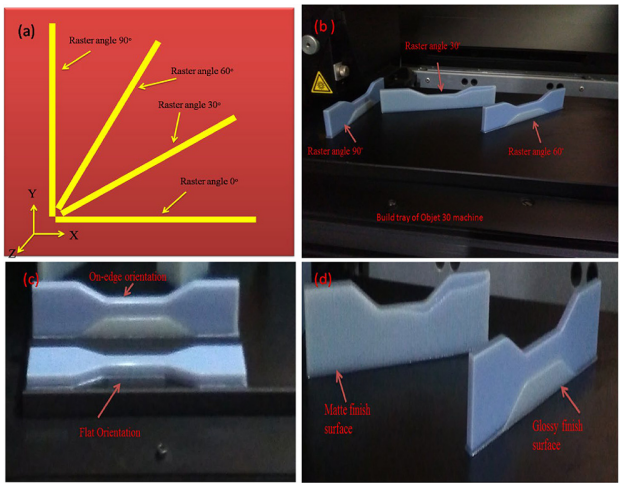
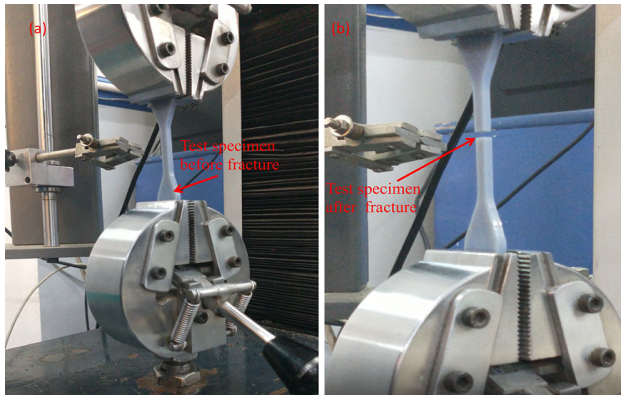


Figure 4 (a) Test specimen before fracture; (b) test specimen after fracture



corresponding to process parameters, i.e. 0° raster angle, flat orientation and glossy type surface finish were used for the instigation of plastic strain. The methodology adopted for the calculation of plastic strain is presented in Figure 5. The value of engineering strain was evaluated by dividing the extension length with a gauge length of the specimen.

The value of true stress and true strain were calculated by using equations (1) and (2) respectively. The plastic strain was calculated by using power law from the true strain data. Equation (3) is used for the calculation of plastic strain. The relation between true stress and plastic strain is shown in Equation (4).

$$Truestress(\sigma_T) = \sigma_E \times ln(1 + \varepsilon_E) \quad (1)$$

$$Truestrain\ (\varepsilon_T) = ln(1 + \varepsilon_E) \quad (2)$$

$$\varepsilon_T = \varepsilon_E + \varepsilon_P \quad (3)$$

$$Truestress(\sigma_T) = K \times \varepsilon_P^n \quad (4)$$

where σ_T is true stress (N/mm^2), σ_E is engineering stress (N/mm^2), ε_E is engineering strain (mm/mm), ε_T is true strain (mm/mm), K is strength coefficient (MPa), n is strain hardening exponent and ε_p is plastic strain.

2.2.2 Computational investigation of plastic strain

Explicit 3 D ASTM D-638 model in the shape of flat dog bone is shown in Figure 6(a). Finite element analysis software ABAQUS was used for the computational analysis of plastic strain.

A continuum 3-D, eight-node linear brick element C3D8R was used to simulate the material in a virtual environment. The mesh density was considered high. The mesh size at the gauge length section was selected as 0.5 mm, and the rest of the section mesh size was used 2 mm. A total of 9368 elements were used in this analysis. Figure 6(b) shows the mesh geometry of the ASTMD638 specimen.

A continuum 3-D, eight-node linear brick element C3D8R was used to simulate the material in a virtual environment. The mesh density was high at the gauge length section. The mesh size at the gauge length section was selected as 0.5 mm, and the rest of the section mesh size was used 2 mm. A total of 9368 elements were used in this analysis. Figure 6(b) shows the mesh geometry of the ASTMD638 specimen.

Figure 5 Flow chart for the calculation of true stress-plastic strain

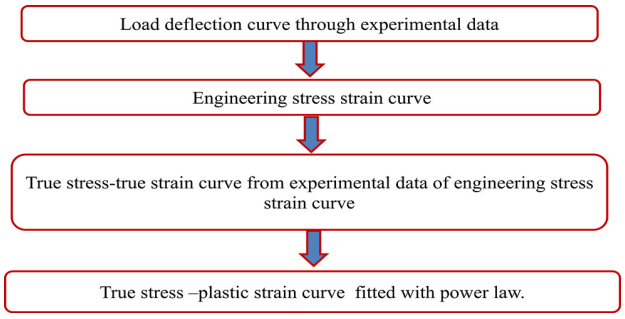
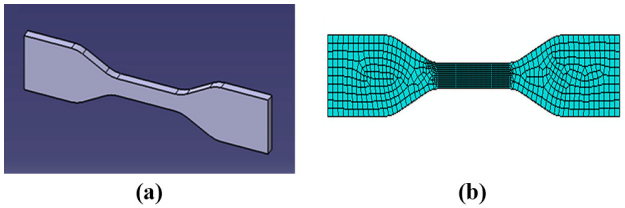


Figure 6 (a) 3D CAD model of the specimen; (b) mesh geometry of ASTM D638 tensile test specimen



3. Results and discussion

This section discusses about the impact of process parameters on mechanical properties such as elastic modulus, ultimate tensile stress and percentage elongation of the RGD840 material. Fabricated samples were immediately sent for the tensile testing. Table 2 shows the experimental measured average value of tensile strength, modulus of elasticity and percentage elongation of the material. Figure 7 shows the effect of process parameters on ultimate tensile strength of the material. It can be observed that process parameters have significant impact on tensile strength of the material. The ultimate tensile strength of 90° raster angle, flat orientation and glossy surface has the peak average value of 27.85 MPa, and on-edge, glossy surface and 0° raster angle has the lowest average value of 20.58 MPa. The components fabricated other than 0° and 90° raster angle have different orientation than loading direction in tensile strength. Due to anisotropic behavior of material, components exhibit different property with variation of raster angle.

Figure 8 shows the influence of process parameters on elastic modulus of RGD840 material. It was observed that elastic modulus of the fabricated samples were also affected by the process parameters. Results of experimental studies show that the value of elastic modulus at raster angle of 90°, flat orientation and glossy surface has average peak value of 2891.5 MPa, and at 60° raster angle, matte finish and edge orientation has the lowest average value of 1534.4 MPa. Figure 9 shows that the average percentage elongation of fabricated samples. It can be observed that percentage elongation was found to be maximum of 23.38% in the component fabricated along 60° raster angle, flat orientation and matte finish. However, the lowest average percentage elongation of 12.08% was observed in the sample of 60° raster angle, edge orientation and matte finish. Figure 10 shows the stress

Figure 7 Ultimate tensile strength (MPa) of the RGD840 material

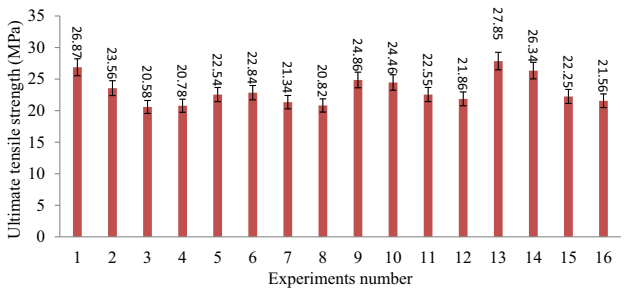
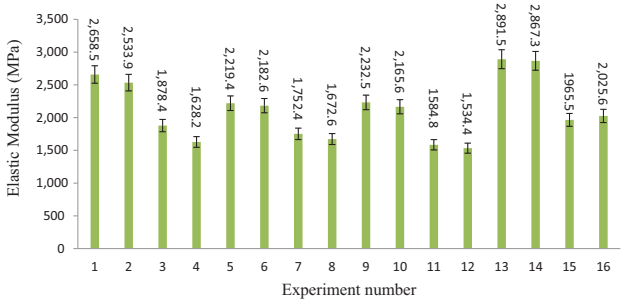


Figure 8 Elastic modulus (MPa) of RGD840 material



strain curve of samples (1, 5, 10 and 15). Results of tensile behavior and percentage elongations of the samples were consistent with the observations made by (Pugalendhi *et al.*, 2020). It was observed that elongation increased with reducing 'E' and vice-versa in the samples printed along raster angle of 0°, 60° and 90°. It follows the stress strain relationship. However, an aberration was observed in the samples fabricated along 30° raster angle.

3.1 Experimental result analysis of plastic strain

Figure 11(a) illustrates the load-deflection relation of the tensile test specimen corresponding to process parameters i.e. raster angle 90°, the flat orientation and glossy type surface. Figure 11(b) shows the true stress-true strain curve of the material. Figure 11(c) demonstrates the experimental true stress-plastic strain curve. The value of strength coefficient (K = 29.26 MPa) and strain hardening coefficient (n = 0.029) were determined by using the graphical method. The value of regression coefficient (R^2) of the fitted curve represents the accuracy of the fitted curve. In curve fitting power law was chosen to fit the experimental data because the calculation of plastic strain was based on power law. The value of the R^2 of the fitted line is 0.991 close to 1. It indicates that the fitted line is good enough to predict the plastic strain. The value of strength coefficient and strain hardening coefficient was used for the computational analysis of plastic strain in ABAQUS software.

3.2 Computational result analysis of plastic strain

Figure 12(a) shows the boundary condition for the tensile test specimen simulation. Boundary condition was applied similarly to the tensile test on UTM. Degree of freedom at the bottom section (Ux = 0,Uy = 0, Uz = 0, Mt/x = 0,Mt/y = 0, Mt/z = 0) was considered as zero. A displacement (Uy = 0.1 mm/min) along the Y-axis load was applied at the top of the specimen. The other boundary condition (Ux = 0, Uz = 0) was applied at the top section. Figure 12(b) represents the contour plot of the plastic strain of the tensile test specimen. In experimental studies, deformation of the test specimen is measured only for the gauge length section. Hence, the FEA analysis of the gauge length section becomes a critical observation.

Figure 9 Percentage elongation of RGD840 material

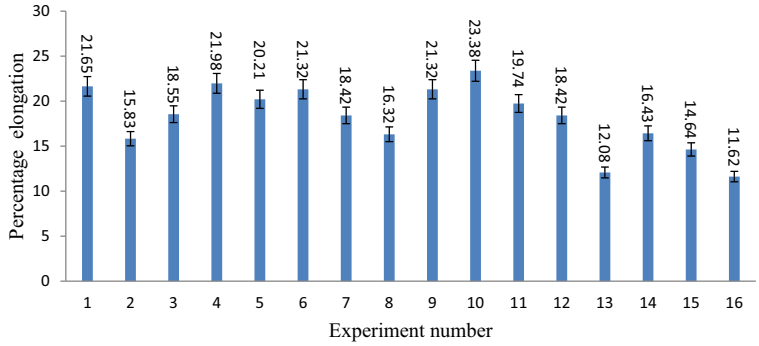
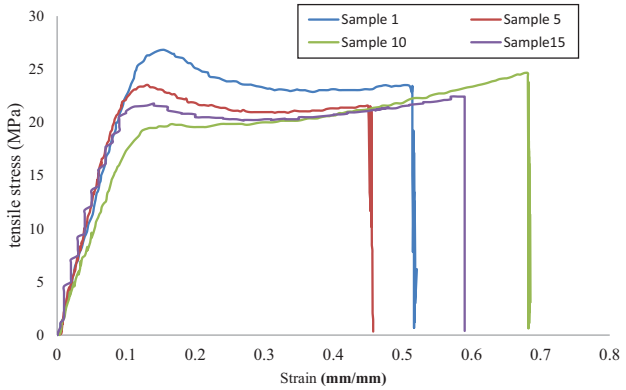


Figure 10 Stress strain curve of samples

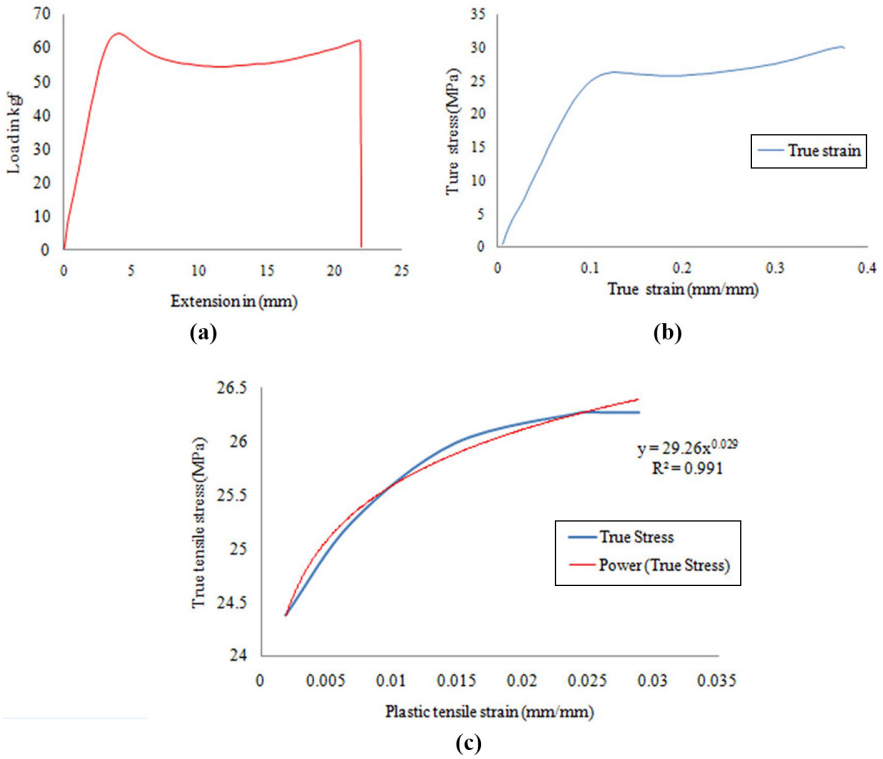


In ABAQUS software, analysis of plastic strain material properties, namely, modulus of elasticity, Poisson's ratio, strength coefficient (K) and strain hardening coefficient (n) were required. The value of elastic modulus E = 2.65 GPa and μ = 0.49, K = 29.26 MPa and n = 0.029 were used for the computational investigation. The value of strength coefficient and strain hardening coefficient was calculated from the experimental value of true stress and plastic strain curve fitting by applying power law. The contour plot of the plastic strain at the gauge length section is shown in Figure 12(c). The true stress-plastic strain curve is shown in Figure 12(d).

3.3 Comparison between experimental and computational results of plastic strain

Figure 13 distinguishes the experimental and computational observation of plastic strain behavior of RGD840 material under tensile loading condition. It can be concluded from the

Figure 11 (a) Load deflection curve of the material; (b) experimental true stress-true strain curve of the RGD840 material; (c) experimental true stress-plastic tensile strain curves



graph that FEA analysis closely approximated the experimental results. Deviations of values of plastic strain obtained from experimental and computational analysis were less than 5%, which is an acceptable limit (Sugavaneswaran and Arumaikkannu, 2014). FEA results deviate from experimental results at some point due to the approximation error of strength coefficient and strain hardening coefficient. In the case of strain analysis, the convergence of FEA results is over predicted. Hence, the result observed from the numerical simulation was correlating to convergence criteria.

3.4 Microstructural analysis

The failure characteristics of rigid opaque materials (VeroBlue RGD840) fabricated through PolyJet process was analyzed using scanning electron microscopy (SEM). VeroBlue RGD840 is a non-conducting material, the electron charge will bombard with the sample and the accumulating charge will create charging effect and the images will be blurred. To avoid accumulation of charging effect, gold coating on broken surface of tensile test specimens were performed to enhance the conductivity of the material.

Raster angle 0°, flat orientation and glossy surface finish fabricated specimen fracture surface is shown in Figure 14 (a). Figure 14(b) shows the SEM image of sample printed along 90° raster angle, edge orientation and glossy finish surface. Figure 14(c) shows the SEM image of the sample fabricated along the raster angle 60°, edge orientation and matte surface finish. Raster angle 0°, flat orientation and matte finish fracture image is shown in Figure 14(d). RGD840 material fracture surfaces are completely flat. The fracture surface is perpendicular to the direction of loading direction of tensile test. It is concluded that failure of material is brittle in nature (Sugavaneswaran and Arumaikkannu, 2014).

4. Conclusions

The effects of PolyJet process parameters on elastic modulus, percentage elongation and ultimate tensile strength on the RGD840 material were investigated experimentally. Further true stress, true strain and plastic strain were investigated. Moreover computational investigation of plastic strain was carried out by using ABAQUS software to validate the experimental results. The major highlights of this research work are given below:

- Elastic modulus of 90° raster angle, flat orientation and glossy surface has average peak value of 2891.5 MPa, whereas 60° raster angles, matte finish and edge orientation has the lowest average value of 1534.4 MPa.
- The ultimate tensile strength of 90°raster angle, flat orientation and glossy surface has peak average value of 27.85 MPa, whereas on-edge, glossy surface and 0° raster angle has lowest average value of 20.58 MPa.
- Percentage elongation was found to be maximum of 23.38% in the component fabricated along 60° raster angle, flat orientation and matte finish, whereas the lowest average percentage elongation of 12.08% was observed in

the sample of 60° raster angle, edge orientation and matte finish:

- To understand the plastic behavior of RGD840 material, true stress, true strain and plastic strain of sample fabricated along 90° raster angle, flat orientation and glossy surface was investigated. Computational analysis of plastic strain was the close approximation (within the range of 5%) of the experimental results.
- The optimum level of process parameters are 90° raster angle, Flat orientation, with Glossy surface finish.
- Failure analysis of RGD840 material was analyzed by using scanning electron microscopy (SEM). The fracture surface of the tensile test specimen indicates the failure is brittle in nature.

Figure 12 (a) Boundary condition and load application on specimen for investigation of plastic strain; (b) contour plot of plastic strain of the test specimen; (c) contour plots at the gauge length section; (d) FEA true stress-plastic strain curve

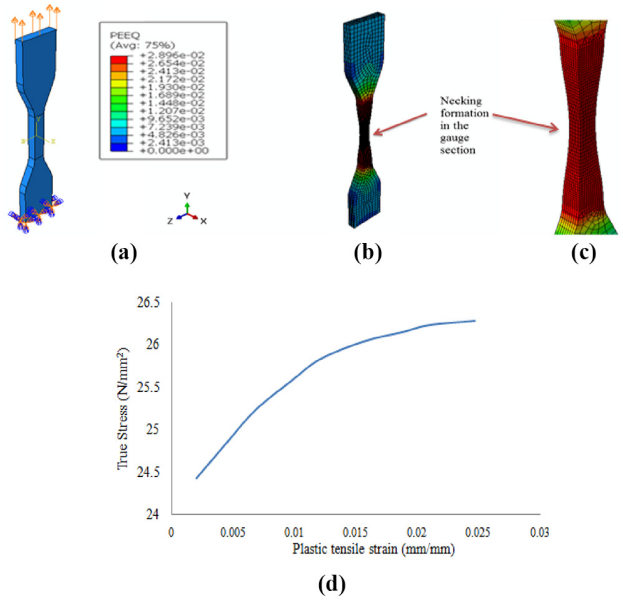


Figure 13 Comparison of experimental and FEA analysis of true stress-plastic strain curve

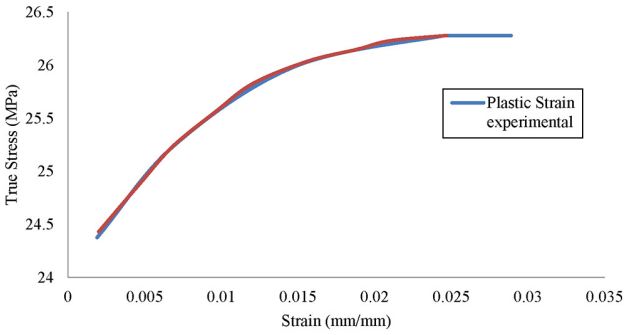
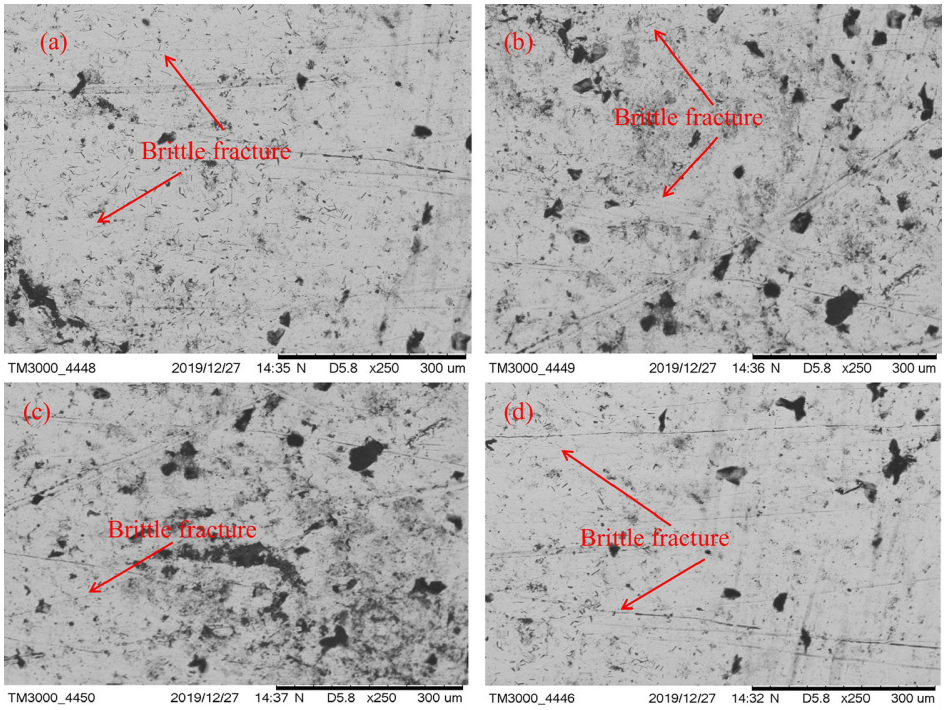


Figure 14 SEM image of fracture surface (a) flat orientation glossy surface type; (b) edge orientation glassy surface; (c) edge orientation matte finish surface; (d) flat orientation matte finish surface



References

Ainsworth, I., Ristic, M. and Brujic, D. (2000), "CAD based measurement path planning for free form shapes using contact probes", *International Journal of Advanced Manufacturing Technology*, Vol. 16 No. 1, pp. 23-31.

Barclift, M.W. and Williams, C.B. (2012), "Examining variability in the mechanical properties of parts manufactured via PolyJet direct rapid prototyping", *International. Solid Freeform Fabrication Symposium Proceedings*.

Cazon, A., Morer, P. and Matey, L. (2014), "PolyJet technology for product prototyping: tensile strength and surface roughness properties", *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, Vol. 228 No. 12, pp. 1664-1675.

Gay, P., Blanco, D., Pelayo, F., Noriega, A. and Fernández, P. (2015), "Analysis of factors influencing the mechanical properties of flat PolyJet manufactured parts", *Procedia Engineering*, Vol. 132, pp. 70-77.

Gaya, P., Blanco, D., Pelayo, F., Noriega, A. and Fernández, P. (2015), "Analysis of factors influencing the mechanical properties of flat PolyJet manufactured parts", *Procedia Engineering* , Vol. 132, pp. 70-77.

Ge, R., Ghadiri, M., Bonakdar, T., Zheng, Q., Zhou, Z., Larson, I. and Hapgood, K. (2020), "Deformation of 3D printed agglomerates: multiscale experimental tests and DEM simulation", *Chemical Engineering Science*, Vol. 217, doi: 10.1016/j.ces.2020.115526.

Kesy, A. and Kotlinski, J. (2010), "Mechanical properties of parts produced by using polymer jetting technology", *Archives of Civil and Mechanical Engineering*, Vol. 10 No. 3, pp. 37-50.

Kumar, K. and Kumar, G.S. (2015), "An experimental and theoretical investigation of surface roughness of poly-jet printed parts", *Virtual and Physical Prototyping*, Vol. 10 No. 1, pp. 23-34.

Liu, W., Song, H. and Huang, C. (2020), "Maximizing mechanical properties and minimizing support material of PolyJet fabricated 3D lattice structures", *Additive Manufacturing*, Vol. 35, p. 101257.

Maurya, N.K., Maurya, Rastogi, V. and Singh, P. (2019), "Comparative study and measurement of form errors for the component printed by FDM and PolyJet process", *Instrumentation Mesure Métrologie*, Vol. 18 No. 4, pp. 353-359.

Maurya, N.K., Rastogi, V. and Singh, P. (2020), "An overview of mechanical properties and form error for rapid prototyping", *CIRP Journal of Manufacturing Science and Technology*, Vol. 29, pp. 53-70.

Moore, J.P. and Williams, C.B. (2015), "Fatigue properties of parts printed by PolyJet material jetting", *Rapid Prototyping Journal*, Vol. 21 No. 6, pp. 675-685.

Pilipović, A., Raos, P. and Šercer, M. (2009), "Experimental analysis of properties of materials for rapid prototyping", *The International Journal of Advanced Manufacturing Technology*, Vol. 40 No. 1-2, pp. 105-115.

Pugalendhi, A., Ranganathan, R. and Chandrasekaran, M. (2020), "Effect of process parameters on mechanical properties of VeroBlue material and their optimal selection in PolyJet technology", *The International Journal of Advanced Manufacturing Technology*, Vol. 108 No. 4, pp. 1049-1059.

Sagias, V.D., Giannakopoulos, K.I. and Stergioua, C. (2018), "Mechanical properties of 3D printed polymer specimens", *Procedia Structural Integrity*, Vol. 10, pp. 85-90.

Singh, R. (2011), "Process capability study of polyjet printing for plastic components", *Journal of Mechanical Science and Technology*, Vol. 25 No. 4, pp. 1011-1015.

Singh, J.P. and Singh, R. (2009), "Comparison of rapid casting solutions for lead and brass alloys using three dimensional printing", *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science)*, Vol. 223 No. 9, pp. 2117-2123.

Sugavaneswaran, M. and Arumaikkannu, G. (2014), "Modelling for randomly oriented multi material additive manufacturing component and its fabrication", *Materials & Design (1980-2015)*, Vol. 54, pp. 779-785.

Suresh, J.A., Kumar, G.S., Ramu, P. and Rengaswamy, J. (2018), "Fatigue life characterization of additively manufactured acrylic like Poly-Jet printed parts", *Advances in Structural Integrity*, doi: 10.1007/978-981-10-7197-3_52.

Tabi, T., Kovacs, N.K., Sajo, I.E., Czigany, T., Hajba, S. and Kova, J.G. (2015), "Comparison of thermal, mechanical and thermomechanical properties of poly(lactic acid) injection-molded into epoxy-based rapid prototyped (PolyJet) and conventional steel mold", *Journal of Thermal Analysis and Calorimetry*, Vol. 123 No. 1, pp. 349-361.

Zelený, P., Safka, J. and Elkina, I. (2014), "The mechanical characteristics of 3D printed parts according to the build orientation", *Applied Mechanics and Materials*, Vol. 474, pp. 381-386.

Further reading

Costa, C.A., Linzmaier, P.R. and Pasquali, F.M. (2013), "Rapid prototyping material degradation: a study of mechanical properties", *6th IFAC Conference on Management and Control of Production and Logistics The International Federation of Automatic Control September 11-13*.

Corresponding author

Nagendra Kumar Maurya can be contacted at: nagendramnnit@gmail.com

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-020-02698-1

ORIGINAL RESEARCH

Hoax news-inspector: a real-time prediction of fake news using content resemblance over web search results for authenticating the credibility of news articles

Deepika Varshney[1] · **Dinesh Kumar Vishwakarma**[1]

Received: 14 March 2020 / Accepted: 16 November 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract
Nowadays social media is one of the important medium of sharing thoughts and opinions of the individual due to its easy access and also it provides an opportunity to the malicious user to post deliberately fabricated false content to influence people for creating controversies, playing with public emotions, etc. The spread of contaminated information such as Rumours, Hoax, Accidental misinformation, etc. over the web is becoming an emergency situation that can have a very harmful impact on society and individuals. In this paper, we have developed an automated system "Hoax-News Inspector" for the detection of fake news that propagates through the web and social media in the form of text. To distinguish fake and real reports on an early basis, we identified prominent features by exploring two sets of attributes that lead to information spread: Article/post-content-based features, Sentiment based features and the mixture of both called as Hybrid features. The proposed algorithm is trained and tested on the self-generated dataset as well as one of the popular existing datasets *Liar*. It has been found that the proposed algorithm gives the best results using the Random Forest classifier with an accuracy of 95% by considering all sets of features. Detecting and verifying news have many practical applications for business markets, news consumers, and time-sensitive services, which generally help to minimize the spread of false information. Our proposed system Hoax News-Inspector can automatically collect fabricated news data and classify it into binary classes Fake or Real, which later benefits further research for predicting and understanding Fake news.

Keywords Fake news detection · Rumour · Death hoax · False information · Social media

1 Introduction

With the invention of the internet and technology, people can easily share information with each other. The primary thing that encourages, the fast-spreading and dissemination of information are online social media platforms. People can share their opinions and thoughts freely over social media. This freely sharing of expression is one of the stepping stones to encourage false information due to which malicious users try to promote false content on web platforms. The spreading of false information over social media may create controversies that may affect human emotions and create a negative impact on society. One of the categories under false information is Fake news. The Fake news can be defined as false information that is disseminated or published as a piece of news for fraudulent or politically motivated purposes or can also be defined as news stories with intentionally false information (Shu et al. 2017). The recent example of false information goes viral during Jamia Millia protest about the Delhi police in India, it was reported that the Delhi police is thrashing a man during Jamila protest as shown in Fig. 1a, but from google reverse image it can be clearly seen in Fig. 1b that the photograph is clicked in Kashmir. This report creates huge controversies during the protest and provokes people towards government.[1] Recently, it was reported that there was a piece of fake news that was in circulation that Delhi Police had destroyed the public

✉ Dinesh Kumar Vishwakarma
dvishwakarma@gmail.com

Deepika Varshney
deepikavarshney06@gmail.com

[1] Biometric Research Laboratory, Department of Information Technology, Delhi Technological University, New Delhi, India

[1] https://thelogicalindian.com/fact-check/delhi-police-jamia-students-protest-citizenship-amendment-act/



(a) (b) (c)

Fig. 1 Example of Fake news

property during the protest of Jamia Millia Islamic University, New Delhi. The sample images and videos[2] that were in circulation is as shown in Fig. 1c.

It has been observed from the study that the trust of the users in mass media has dramatically degraded with different age groups (Swift 2016). Hence, the extensiveness of the false news can break the integrity of the entire news ecosystem. That's why false information detection and mitigation of these negative effects over web platforms is one of the critical issues.

Recent research encountered fake news in different ways. One of the methods is via crawling the web and analyzing the article content. In Horne and Adali (2017), the author focuses on the headline of fake and real news; from the study, it has been found that the fake news has a significant difference compared to real news. Whereas the authors of Pennycook et al. (2019), applied a warning system on news articles, that can help users to better understand the credibility of the news article. In Vishwakarma et al. (2019), authors crawl and analyze the google search results and applied a rule-based approach to predict whether the news is fake or not. These approaches are leading to some promising results. However, no principle study is conducted on paragraph features of an article. From the analysis, it has been found that each paragraph leads to giving important features (content and sentiment-based features) for prediction. Besides, there has been no research that provides a standardized understanding of (i) what possible paragraph features in a news article are; (ii) how discriminative these features are. To give a comprehensive understanding of these aspects, in this paper, we made the following key contributions that are summarized below:

- A classification framework is developed that uses a Content similarity-based agreement approach on web URLs for authenticating the credibility of content, to detect fake news on an early basis.
- A new methodology is outlined that leverages a rumour debunking system (Snopes, Politifact) and Wiki database to build a dataset for early detection of fake news on the web and social media.
- The proposed framework can handle long articles for analysis, and in most of the real scenario, articles are long and descriptive in nature, to have a good representation of an event.
- The performance of the proposed approach is evaluated on public and self-generated datasets of Fake News, and also it is compared with the existing state-of-the-art.

The rest of the paper is structured as follows; Sect. 2 gives an overview of the recent states-of-the-art in the false information detection; while Sect. 3 presents the detailed description of our proposed framework; Sect. 4 describes the experimental analysis and results. Finally, Sect. 5 gives the conclusion and future work.

2 Related work

Detecting the credibility of content is one of the crucial issues over social media. Before the emergence of the research in the field of fake news/rumors, many of the researchers have worked on the credibility evaluation of the information. Due to the fast-spreading of false information over online platforms, detection and analysis of misleading information is one of the challenging problems to work on. From the previous studies, it has been observed that few studies have been done in this area. Hence, detecting

[2] https://www.youtube.com/watch?v=2v1juUg5ZM4

misleading information is one of the emerging areas. Many of the problems that have been taken in the existing literature to detect misleading information face the task as a classification problem, with an intent to classify it as rumour or not rumour, true or false on a particular piece of text by employing a certain set of features. The detailed description of the feature extraction approaches that have been adopted by state-of-the-art is as follows. From the previous study, a broad categorization of features has been proposed by Shu et al.(2017) focusing mostly on fake news detection distinguishing context and content-based approaches to feature extraction. Content-based approaches can be described as approaches that mainly rely on content features, where the information can be directly extracted from text like linguistic features. Many of the existing studies have exploited content-based features at least in some capacity, extracting some basic knowledge from the text (sentence, tokens, linguistics, etc.) and also some of the studies employed various linguistic cues which characterize deception like the use of negative words, swear words. These linguistic cues, including some other features that have been successfully implemented in numerous methods concerning fake news and rumour detection (Castillo et al. 2011; Gupta et al. 2014; Qin et al. 2016; Zhang et al. 2012). The authors of Zubiaga et al. (2016) and Qin et al. (2016) have employed syntactic features such as the number of content words(noun, verb adjectives), frequency of specific keyword has been employed. Whereas, the authors Feng and Hirst (2013), Chakraborty et al. (2016), and V et al. (2016) have exploits bi-grams and tri-grams as a feature. It has also been proven that negation, doubt words, negation, vulgar expression gives some efficient clues in determining the veracity of online information (Vosoughi et al. 2017). Along with these features, sentiment-based features are also adopted based on the opinions and emotions expressed in the text (Castillo et al. 2011; Qazvinian et al. 2011; Yang et al. 2012). The other set of features that have been employed for fake news detection is context-based features. The context-based features mainly rely on the nearby information like user's characteristics, network propagation features, and the reaction of the other users on the news and post. The authors of Castillo et al. (2011), have employed user based characteristics of social media users like the number of posts, age of an account, number of followings, etc. The user characteristics give an efficient clue concerning the reliability of the shared information. Whereas, the age of the user account and the reputation of the source posted an article leads to gives some good measure for analyzing the credibility of posted claims (Castillo et al. 2011; Zubiaga et al. 2016). Like user characteristics, network-based characteristics are also employed concerning the properties of the network where the news is shared. The authors of Ma and Wong (2017), reports that most of the existing studies have adopted network-oriented feature like considering statistics on the diffusion patterns, such as the number of retweets and propagation times (Castillo et al. 2011; Ma et al. 2015; Yang et al. 2012), modelling the temporal characteristics of propagation (Kwon et al. 2017, 2013). Some of the work has incorporates both user and content-based features, including the information from the propagation tree (Wang and Terano 2015; Wu et al. 2015). Along with these features user opinions and expression, stance based features also play an important role to identify rumour and fake news. In this paper we have mainly considered content and sentiment-based features, the proposed approach is distinct from the others in terms that we have developed a classification framework that uses a Content similarity-based agreement approach on web URLs for authenticating the credibility of content, to detect fake news on an early basis, where paragraph based analysis with respect to each article has been performed. To the best of our knowledge none of the previous work have adopted paragraph based analysis with respect to the proposed features to get more insights into the story. For the verification of the posted news, credible sources are one the important key aspect through which we can retrieve significant clues for the prediction of fake news. Now if we talk about in respect to learning algorithm than to counter the problem of false information it has been observed that machine learning algorithm is found to be well efficient in solving several tasks related to false information detection, which provides promising results. Mr. Ryszard S. Michalski was among the most significant contributors to machine learning (Wojtusiak and Kaufman 2010; Cervone et al. 2010). He developed numerous machine learning algorithms, and one of the learning algorithm given by him is AQ learning. The basic version of the AQ learning algorithm takes as an input a set of positive samples(P) and a set of negative samples (N) and a quality measure, where a consistent hypothesis in the form of the attributional ruleset is returned. However, many of the problems related to false information detection has been implemented by employing a supervised machine learning algorithm. In the supervised machine learning algorithm, most of the researcher has reported that SVM outperformed all other classifiers and extensively adopted method for classification in several research areas. The authors of Zhang et al. (2012) showed that for deception detection, the SVM method outperforms all other methods with an f1 score of 0.84. SVM is reported as one of the good classifiers for fake news detection. However, there are some cases where it may not perform well if there is the existence of noise in the dataset selected for training. Nonetheless, it has also been reported that there is significant variation in performance, depending on the dataset selected for training. In SVM based approach most of the authors were exploited content-based features (like linguistic features, visual features etc.) for detecting false information (Afroz et al. 2012; Pérez-Rosas et al. 2015; V et al. 2016; Zhang et al. 2012). Most of the

approaches have exploited a mixture of content and context-based features for the training of classifiers (Qin et al. 2016; Yang et al. 2012). The authors of Wu et al. (2015) have employed a fascinating SVM-based strategy, by proposing a graph-kernel-based SVM classifier for rumour detection using content-based features and propagation structures with an accuracy of 91% was reported on Sina Weibo dataset. For combating rumour, there is another widely considered family of the algorithm for this task is the decision tree (Breiman 2017). In the decision tree, a recursive split on feature value has been performed, for determining the class. C4.5(J48) is an algorithm used to generate a decision tree developed by Ross Quinlan. C4.5 is an extension of earlier ID3 algorithm. In Quinlan (2014), J48(C4.5) algorithm is used to generate a decision tree from data. The improvement of the J48 decision tree, with respect to other algorithms (Aker et al. 2017; Castillo et al. 2011; Giasemidis et al. 2016). The authors of Lafferty et al. (2002) have proposed the set of temporal, structural and linguistic features for rumour analysis by employing a random forest model with an accuracy of 90%. In Zeng and Spiro (2016), the author proposed only content level features of tweet, random forest classifier is used as a classification model and achieved an average accuracy between 0.83 and 0.88. Then authors of Hardalov et al. (2016) exploited a logistic regression classifier for identifying the credibility of Bulgarian true and false news with an accuracy of 75%. In addition to these classifiers, CRF has also been employed for rumour classification problems. The main advantage of the CRF classifier is the capability of modelling sequences. It has been observed from the studies (Zubiaga et al. 2016, 2017) that CRF is efficiently leverage previous context for improving the performance of rumour detection in breaking news events. Like CRF, HMM (Hidden Markov Model) has also able to deal with sequential data. In the rumour detection problem, HMM is mainly exploited in rumour veracity classification(Vosoughi 2015; Vosoughi et al. 2017). Another, sort of classification algorithm that has been exploited for the verification of news is ensemble-based classification, which relies on the prediction of a number of independent classifiers, where the final classification is provided by majority vote schema (Wang et al. 2017).

Most of the researchers have shown their keen interest in deep learning techniques from traditional machine learning algorithms. One of the reasons for doing so is, the existing machine learning algorithms are based on hand-crafted features, which leads to being very time consuming for complex feature extraction. With a deep neural network, it has also been observed that the framework can easily learn the hidden representation from smaller input. The two widely used artificial neural networks exploited for false information detection named as CNN (Li et al. 2019) and RNN. In RNN, the nodes are sequentially attached to one another, to create a directed graph. Like RNN, in some of the research particularly Long Short-Term Memory(LSTM) is exploited for modelling sequential data effectively (Hochreiter and Schmidhuber 1997). For rumour detection Ma et al. (2016) has adopted the RNN method. There are different variants of RNN that have been proposed by Cho et al. (2014), namely: LSTM, GRU, and tanh-RNN. It has been observed that GRU has obtained better results on the mentioned datasets with an accuracy of 0.88 and 0.91. The authors of Chen (2017) have proposed the CNN architecture for countering both stance and veracity classification of tweets with an accuracy of 0.70 and 0.53, respectively. In Yu et al. (2017) authors have exploited paragraph embedding and CNN. The paragraph embeddings are used to learn the representation of the small group of posts in a specific event, which further be used as an input to CNN. Along with this work, some of the authors have exploited the mixture of RNN and CNN (Ajao et al. 2018; Bodnar et al. 2014).

3 Proposed architecture: Hoax news inspector

In this section, we describe our proposed framework in detail. The basic overview of the proposed idea is shown in Fig. 2, where the system is segregated into two units Data Collection and Data Classification. The data collection unit is responsible for collecting samples from .

rumour debunking websites and the Wiki database using the proposed script with respect to the posted claims related to different subjects like politics, death reports, etc. The collected dataset will further be used to retrieve features from the credible web sources and later utilized for classification of claim in real or fake. The detailed flow diagram of the proposed architecture is shown in Fig. 3. It can be observed from Fig. 3 that the proposed algorithm is a build-up of small modules, that are combined together to form an overall system. The main goal of the algorithm is to detect whether the post is fake or not. The proposed framework is broadly composed of two module Data Collection and Data Classification. The Data Classification is further subcategorized into modules as (i) URLs Filtering Unit (ii) Processing Unit (iii) Classification Unit. These modules are deeply discussed in the following section.

3.1 Data collection

The first module in the series of Fake news detection is the Data Collection. The collection of fake or real data is one of the challenging tasks. Many of the authors have used prominent fact-checking websites as one of the sources for the collection of claims; some of them are Politifact, Snopes, Fact check, etc. In this paper, we are mainly focused on the Politifact and Snopes websites for collecting the claims and

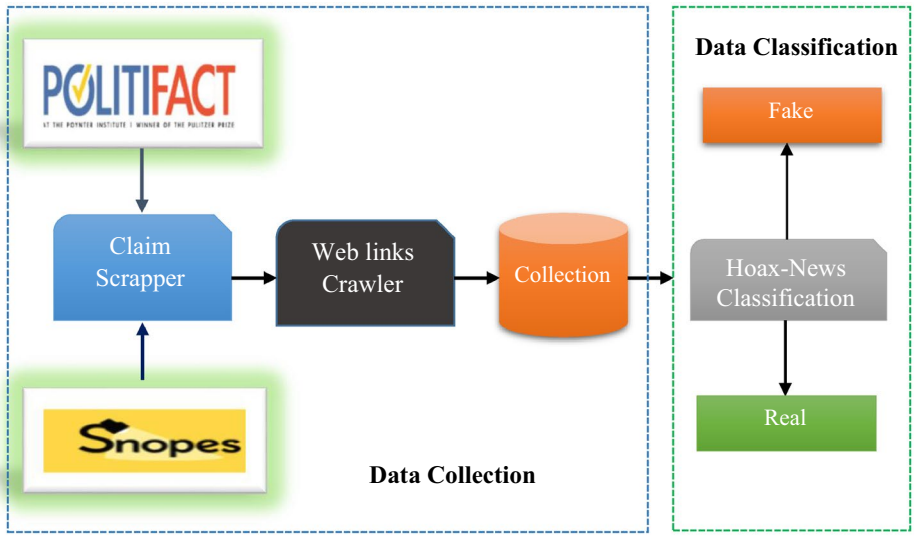


Fig. 2 The flowchart of Hoax-News Inspector system

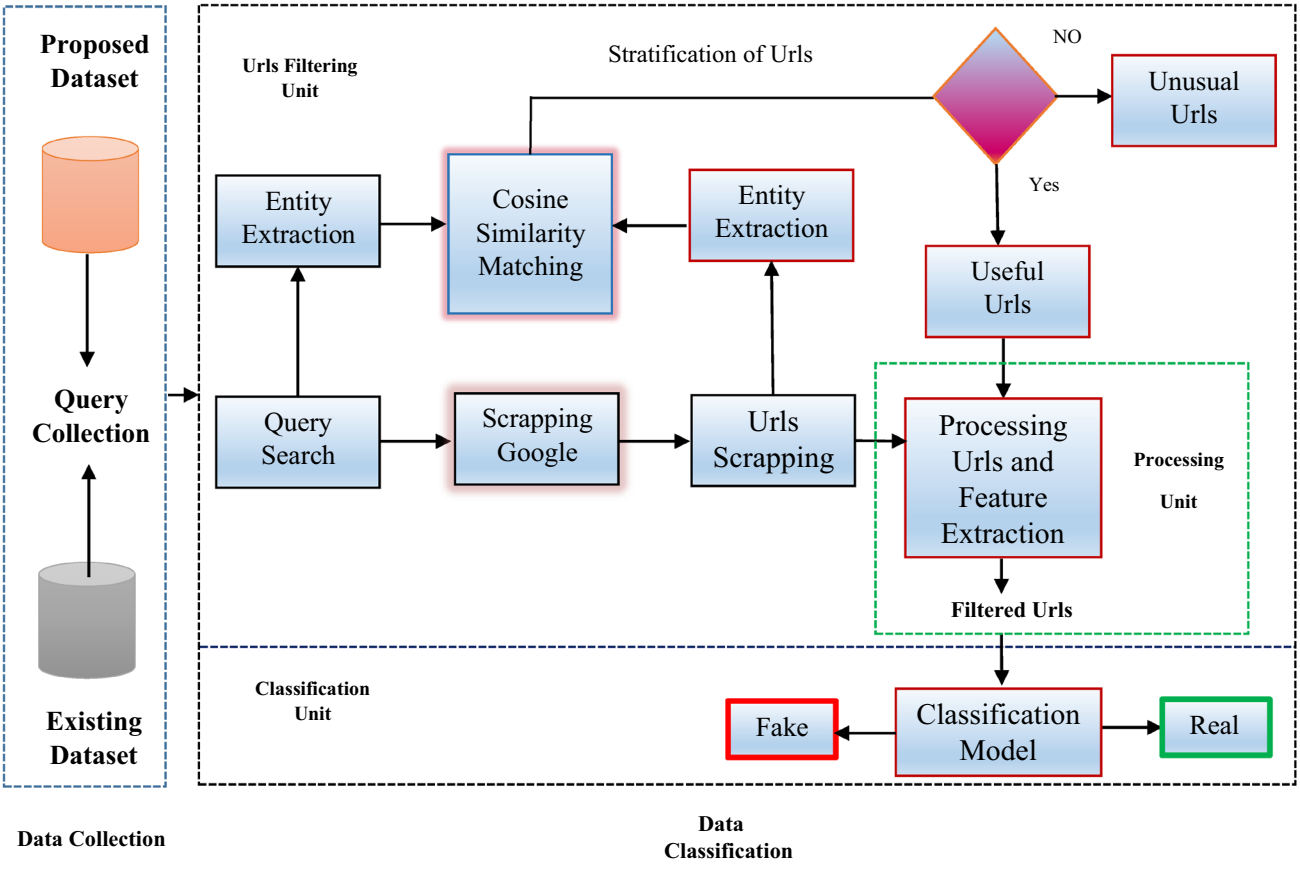


Fig. 3 Proposed Architecture

their veracity for building ground truth data. The data collection module is responsible for collecting ground truth data that further be used to build the model. Two different aspects are employed in the data collection process. Firstly, we have used some popular existing datasets, one of them is "Liar". Liar[3] dataset is one of the publicly available datasets that has been used by Khan et al. (2019).

[3] https://www.cs.ucsb.edu/~william/data/liar_dataset.zip

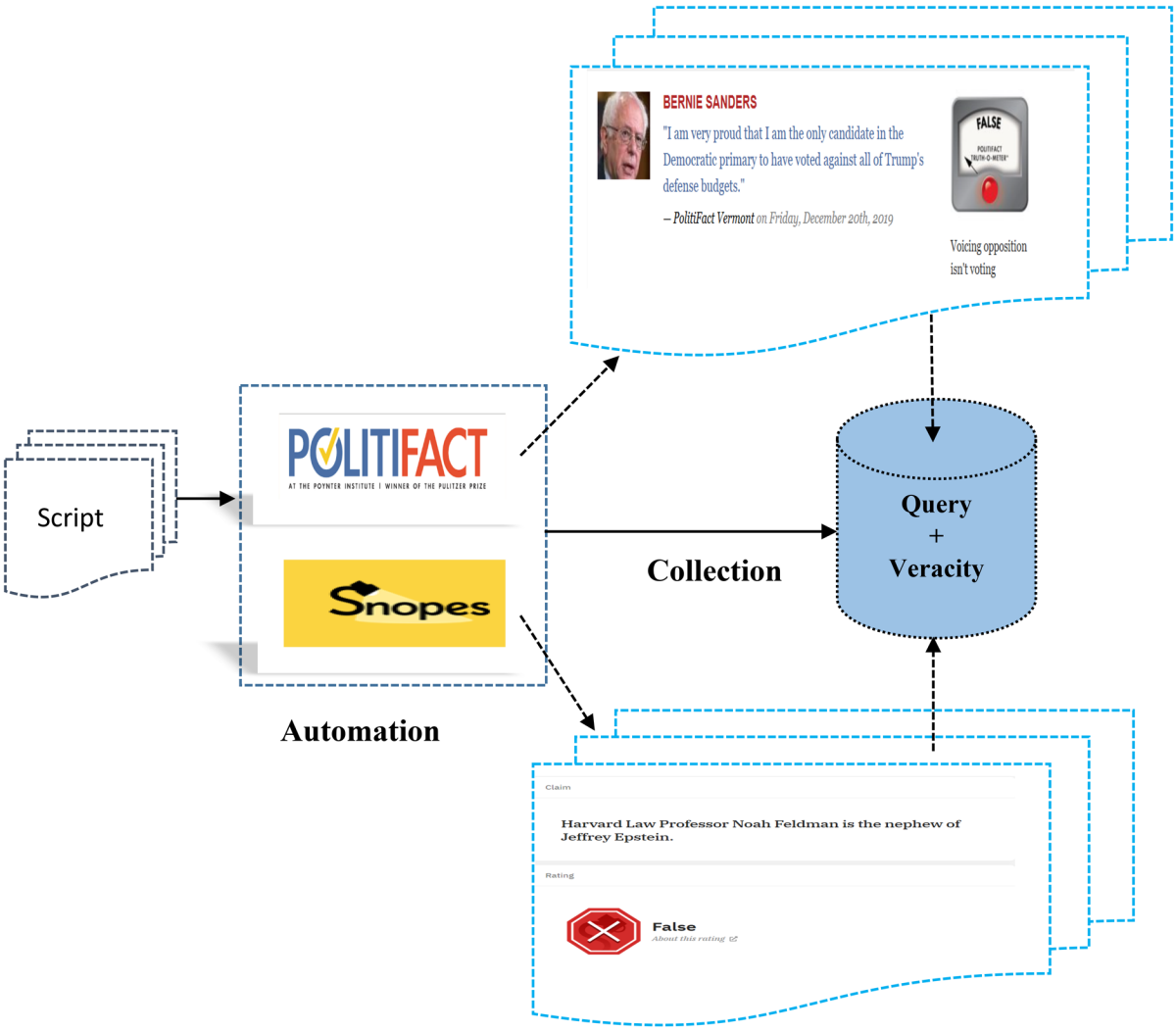


Fig. 4 Data collection process for the Proposed Data set

The dataset contains 12.8 K short statements from Politifact.com that comprise of six labels of truthfulness rating: pants-fire, false, barely-true, half-true, mostly-true, and true. In our work, we mainly focused on categorizing information into two categories fake and real. So, for binary classification of news like (Khan et al. 2019), the same concept has also been applied in our work with respect to fake and real samples(pants-fire, false, barely-true are consider as fake and half-true, mostly-true and true are as real). This dataset is comprised of a significant amount of posts from online social media that mostly deal with political issues, including statements of democrats and republicans. Except for the existing datasets, we have also collected our own dataset (Primary dataset) from rumour debunking websites (Politifact and Snopes). The proposed dataset is the collection of 671 claims related to different categories (politics, death…etc.) containing 255 real claims and 416 false claims. Figure 4 shows the technical view of the data collection process for our dataset. Algorithm 1 shows the process we have adopted for the data collection of our proposed dataset. In this algorithm, we have discussed in detail how claims and respective veracity are extracted from debunking websites (Politifact and Snopes). Selenium web driver has been used to automate the system of collection, where firstly the web driver retrieves the page where the claim has to be extracted. Here n is the number of pages through which we want to collect the claims. By specifying given XPath, we can extract specific attributes (Claim and Veracity) from the page. This set of collected claims as a query is the input for the next unit.

Algorithm 1 Data Collection

```
1. procedure data collection ()
2. driver = web driver. Chrome ()
3. driver.get('https://www.politifact.com/truth-o-meter/statements/')
4. pages= n
5. For k in range (n, page+1):
6.     Scrap(Claims)
7.     Scrap(Veracity)
8.     end procedure
```

3.2 Data classification

The second module in the series of Fake news Detection is the Data Classification. This module is further subdivided into three submodules URLs Filtering, Processing Unit, and Classification unit. This section is covering the technical description of all three modules in detail. The URL filtering unit is the first unit of data classification and responsible for filtering URLs that are needful, later in the second phase the set of useful URLs moves to the processing and feature extraction unit for retrieving efficient clues required for the prediction of fake news from these URLs. Lastly, all sets of features are fed into the classification model for predicting results. The detailed description of each of these phases is shown in the following sections.

3.3 URLs filtering unit

This is the first submodule of the data classification. This module is responsible for filtering useful URLs coming out from the google search results. In the first phase, the query which is coming out from the previous module is given as an input, which further is used to extract google search results. In the next step for each query, entities have been extracted. Rake library is used to extract the entities. Rake[4] is termed as Rapid Automatic Keyword Extraction is one of the well-known and widely used NLP techniques to automatically extract keywords from documents/sentences. Entity extraction is responsible for providing a compact representation of content. As we would like to process only those URLs that relate to our query. For retrieving useful URLs, a Cosine Similarity algorithm has been applied between entities extracted from the search query and the entities extracted from each URL's title, retrieved in the google response results. The filtering of URLs is an important unit as this will gives the set of reliable data that we need to process further. To decide the rule, which URL has to process, the cosine similarity score of the retrieved titles in the google search results is analyzed thoroughly. If a match is above the threshold value and the particular keyword is found in the title, then the URL is considered as a Useful URL, and these sets of URLs further be given as an input to the next unit. For specifying the set of keywords, we have analyzed the news titles and observed that there exists a pattern of specific keywords in the titles encountering false content. One of the rumours debunking website[5] has been analyzed thoroughly and manually extracted the list of ['false', 'fake news', 'falsely', 'mistakenly', 'incorrectly', 'misinformation', 'death hoax', 'false news', 'hoaxes', 'hoax', 'rumour', 'rumour'] frequently used words in the titles encountering fake news and included them in our keyword list. The complete procedure of filtering URLs is described in Algorithm 2. The first decision of the threshold for the stratification of URLs is based on the manual analysis of all the samples value in the feature set. We have manually identified that the relevant titles with respect to the search claim and for all those relevant titles retrieved, we combinedly take an average value that come out to be 0.31 and use as a threshold value for the initial stratification of URLs. It also reveals that the relevant links have a cosine similarity value range lies above 0.30. Hence, if the cosine similarity value is found to be greater than 0.30 and the specific keyword is encountered as mentioned in the keyword list or the cosine similarity value is greater than 0.30 the corresponding Url is returned otherwise not considered for analysis. After this phase of filtering, another phase of filtering is applied by removing URLs not are from authentic sources. Instead of manually building a dictionary of the authentic news channel and make it restricted, we applied the dynamic verification of URLs. The respective URL has been searched over google, and their Wikipedia source has been crawled for retrieving useful attributes like Type, Publisher, Owner, Founded, Country, Circulation. These attributes are responsible for identifying the reliability of the respective URLs, as shown in Fig. 5. In the last phase of URLs Filtering, we manually analyzed all the collected attributes and removed any sort of irrelevant data. The final set of filtered titles are given as an input to the next unit.

Algorithm 2 URLs Filtering

```
1. procedure URLs Filtering (title, query, keywords)
2. CS ← Cosine Similarity (query, title)
3. If CS>0.30 and the title has "keywords" or CS>0.30
4.     return URLs
5. Else
6.     return false
7. end procedure
```

[4] https://pypi.org/project/rake-nltk/

[5] https://www.snopes.com/fact-check/celebrity-death-hoaxes/

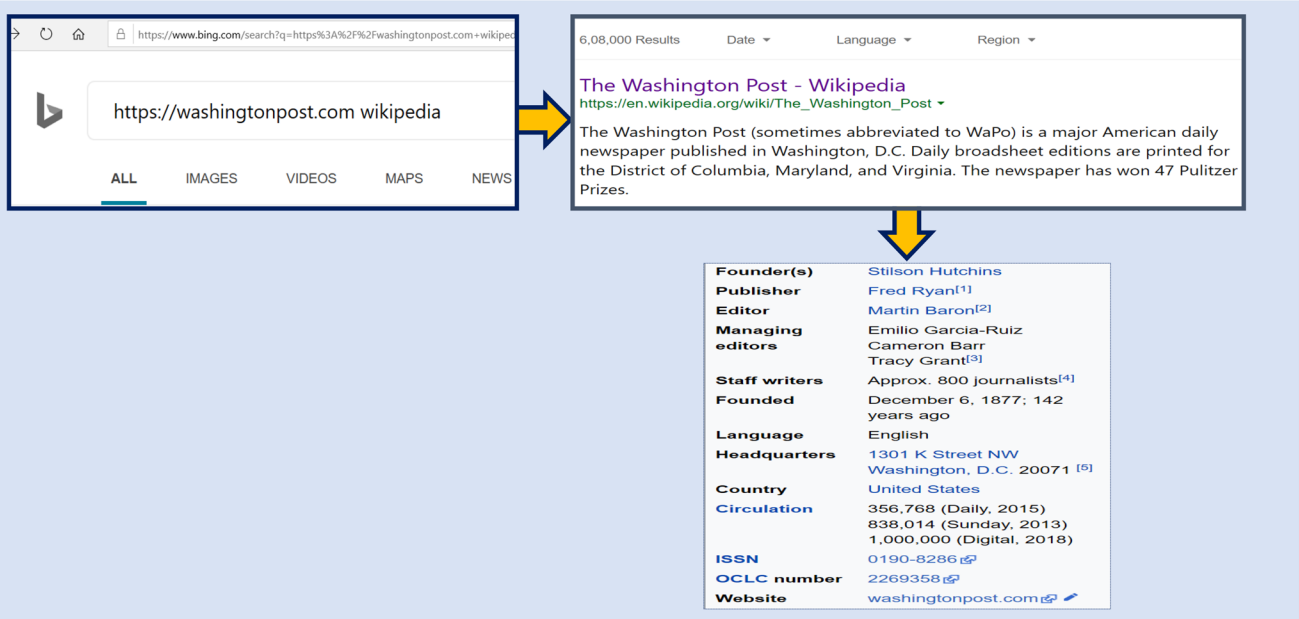


Fig. 5 Process of verifying the source of Useful URLs

3.3.1 Processing unit and feature extraction

This is the second submodule of the data classification. This module is responsible for the processing of useful URLs that are given as input from the previous phase. In this phase, the content of the useful URLs are processed. As we have discussed earlier that we are going to consider long articles. So for processing, articles having at least 8 paragraphs are considered for analysis, it has been observed from[6] that on an average the length of good articles lies within 4 to 8 paragraphs, so we have chosen the max value for analysis. The first 8 paragraphs of each title are extracted and processed. A beautiful soup library is used for the extraction of paragraphs from the webpage, and the Rake library of python is utilized for the entity extraction. One of the examples of encountered keyword from the extracted entities having a false claim is shown in Fig. 6. In this unit, another set of filtering is applied. The filtering is relying on the average cosine similarity match value between the entities extracted from the title and the entities extracted from each paragraph. If the Average Total Cosine Similarity Score(ATCS) ≥ 0.45, then the URL is considered to be the useful URL, otherwise not. The Average Total Cosine Similarity Score (ATCS) ≥ 0.45 is calculated by applying an experiment on the dataset considered for analysis. The threshold value of 0.45 is decided based on analyzing the ATCS value with respect to all fake and real samples. To decide the threshold, the average value of ATCS for all the fake and real samples has been analyzed, where it has found that the average value of ATCS concerning the total number of samples in the feature set is come out to be below 0.45. So, in this way, we have considered the threshold value to be ATCS ≥ 0.45 to focus on the relevant set of URLs for the reliable and accurate prediction that further goes to the classification unit. The entire process of this phase is explained in the following steps and Fig. 7. In the first step titles coming from the previous unit are further be processed and filtered concerning the article length. The title will only be processed also if the article associated with the title have at-least 8 paragraphs. In the second step, the first 8 paragraphs have been extracted and processed for each given title. For each given paragraph concerning the given title, entities are extracted, and features have been analyzed. In the third step, to find the similarity between the title concerning each paragraph in the article, the cosine similarity score is calculated. The total cosine similarity score can be calculated by summing up the cosine similarity for every 8 paragraphs; later the average value is taken into consideration for the 8 paragraphs, represented in the Algorithm 3 termed as ACS (Average of Total Cosine similarity of all 8 paragraphs). In the last step to decide the threshold value for further stratification of titles with respect to the content of the article, the AVG function is applied over an attribute ACS considering all samples in the feature set concerning this attribute that further is used to calculate ATCS

[6] https://ezinearticles.com/?Article-Writing---How-Many-Paragraphs-Should-My-Article-Be?&id=4019539

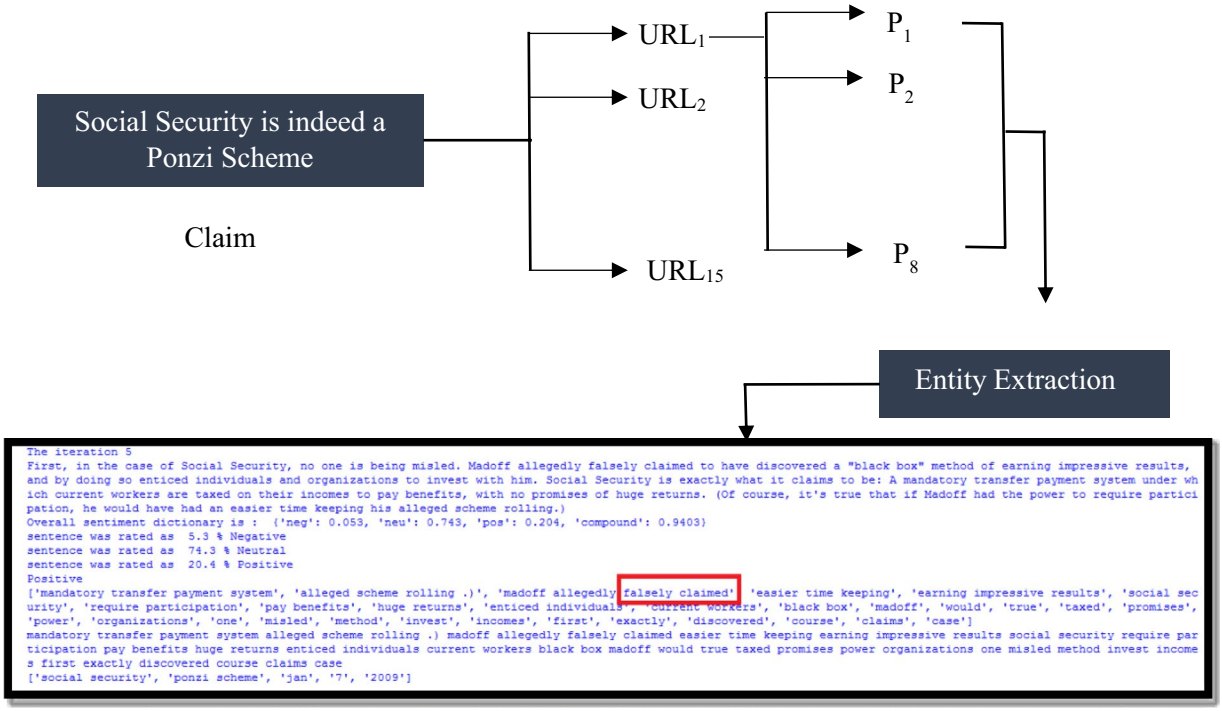


Fig. 6 Example of encountered false keyword from the extracted entities having false claim

(Average of Total Cosine Similarity) that comes out to be 0.45. So, when ATCS < 0.45, URL is not considered for analysis. The filtered set of titles is considered as useful titles and goes for the classification in the next unit for reliable prediction.

Algorithm 3 Processing Unit

```
procedure Processing Unit (URLs)
for k in range (1,9):
    if(soup.find_all("p") [k].get_text ()):
        paragraph=soup. find_all("p") [k]. get_text ()
        CS1 ← Cosine_Similarity (paragraph, title)
        simi_score_total= simi_score_total+CS1
        Feature_extraction(paragraphs)
ACS= ((simi_score_total)/8)
Average_Total_Cosine Similarity (ATCS)= AVG(ACS)
If ATCS>=0.45
    return URL
Else
    return False
end procedure
```

3.3.2 Classification unit

The last submodule of this system is the classification unit. In this unit, the final set of filtered URLs are going to process further and is considered as useful URLs. The extracted features with respect to each useful URL, as discussed in the processing unit have been used to build the classification model and to classify whether a post is fake or real. Table 1 describes all sets of features we have utilized for analysis.

4 Experiment and results

In this section, we first illustrate the experimental setup and the baselines. Then, we showed the observed results and compared various models. The experiment has been performed on two sets of data. The first one is on the publicly available dataset for fake news analysis: Liar dataset and another one on the self-generated dataset. We have four baselines and among four baselines majority baseline, are logistic regression classifier(LR), a support vector machine(SVM). In this section, we provide an in-depth performance analysis of our traditional machine learning classifiers. We have also highlighted the best performance for each dataset and each matrix in bold. The evaluation measures that have been taken for analysis are accuracy, tp-rate, precision, recall, and f1-score to distinguish news, fake and real.

4.1 Liar dataset

Liar dataset was created by Wang (2017), which is the collection of 12,836 short-statements along with its veracity. The dataset is labelled with a discrete set of values from 1 to 6 corresponding to pants-fire, false, mostly-false, half-true,

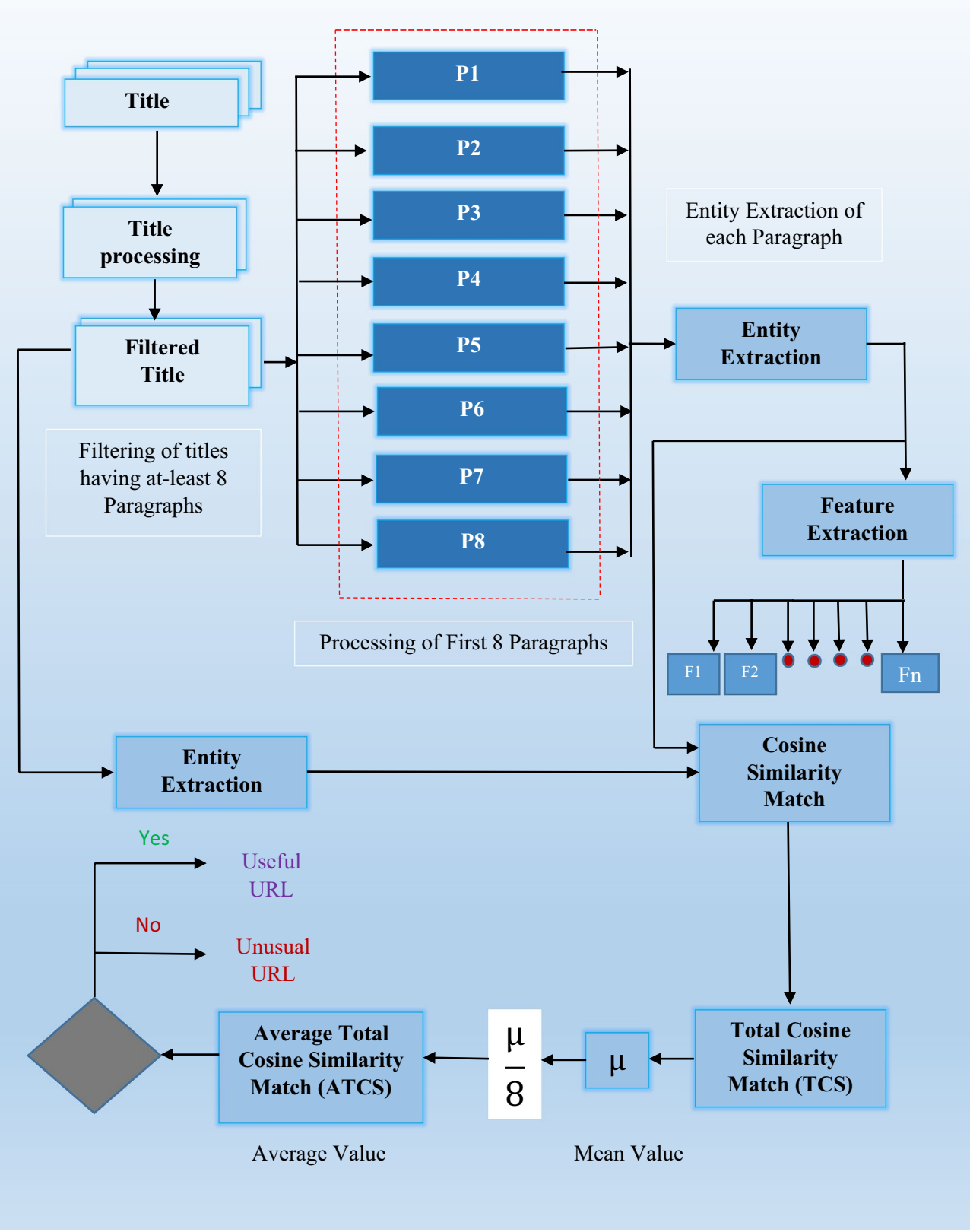


Fig. 7 Flow Diagram of Processing Unit

Table 1 Proposed features

Feature Type	Feature	Feature description
Content-Based	Fake_count	The total number of fake keywords in each paragraph text of an article
Content-Based	Min_fake_count	Min count of the fake keyword in each paragraph of an article
Content-Based	Max_fake_count	Max count of the fake word in each paragraph of an article
Content-Based	TS	Title Cosine similarity score with the given query
Content-Based	TCS	The total cosine similarity score between title heading and each paragraph text
Content-Based	ACS	Average of Total Cosine similarity Score of all 8 paragraphs
Sentiment Based	TCPS	Total Count of Positive Sentiments in all paragraphs
Sentiments Based	TCN_ES	Total Count of Negative Sentiments in all paragraphs
Sentiment Based	PTN	Positive to Negative ratio
Sentiment Based	NTP	Negative to Positive ratio
Sentiment Based	TCNS	Total Count of Neutral Sentiments

Table 2 Performance of the Model on the Liar dataset

Feature	Classifier	Accuracy	TP-rate	FP-rate	Precision	Recall	F-measure
Content based	LR	0.79	0.79	0.21	0.84	0.79	0.78
Content based	Random Forest	0.95	0.95	0.04	0.96	0.95	0.95
Content based	Naïve- Bayes	0.78	0.78	0.21	0.82	0.78	0.77
Content-based	Decision Tree	0.81	0.81	0.18	0.85	0.81	0.80
Content based	k-NN	0.94	0.94	0.05	0.94	0.94	0.94
Content based	SVM	0.63	0.63	0.36	0.78	0.63	0.57
Sentiment based	LR	0.83	0.83	0.16	0.83	0.83	0.83
Sentiment based	Random Forest	0.86	0.86	0.13	0.87	0.86	0.86
Sentiment based	Naïve-Bayes	0.83	0.83	0.16	0.83	0.83	0.83
Sentiment based	Decision Tree	0.87	0.87	0.12	0.89	0.87	0.86
Sentiment based	k-NN	0.82	0.82	0.17	0.83	0.82	0.82
Sentiment based	SVM	0.53	0.53	0.46	0.54	0.53	0.51
Hybrid feature (Content + sentiment)	LR	0.79	0.79	0.20	0.85	0.79	0.78
Hybrid feature (Content + sentiment)	Random-Forest	0.93	0.93	0.06	0.93	0.93	0.93
Hybrid feature (Content + sentiment)	Naïve-Bayes	0.78	0.78	0.21	0.83	0.78	0.77
Hybrid feature (Content + sentiment)	Decision-Tree	0.81	0.81	0.18	0.85	0.81	0.80
Hybrid feature (Content + Sentiment)	k-NN	0.89	0.89	0.10	0.89	0.89	0.89
Hybrid feature (Content + sentiment)	SVM[RBF]	0.64	0.64	0.35	0.79	0.64	0.59

mostly-true, and true. As our problem is based on binary class classification so the data with the first, three labels are labelled as fake news and the others are labelled as real news. The proposed solution has been applied on the given dataset, and the detailed evaluation measure like Accuracy, TP-rate, FP-rate, Precision, Recall, F-Measure has been shown in Table 2.

From Table 2 it has been observed that the evaluation is performed using various machine learning classifiers (Logistic Regression (LR), Random-Forest, Naïve Bayes, K-NN, and SVM) with respect to individual proposed features. Weka tool is used to evaluate the performance of the classifier. Weka is one of the open-source machine learning software that gives transparent access to well-known toolboxes such as sci-kit-learn, R, and Deep-learning. From Table 2, it is clearly depicted that the Random-Forest classifier outperforms all others. From the previous studies also, it has been seen that Random-Forest and SVM were found to be the best classifier for detecting fake news. The analysis has been performed for the given classifier with respect to each type of feature (Content-based, Sentiment-based, and Hybrid-based) exploited for fake news detection. It can be seen that random forest outperforms others in each of the given types of feature, with having an accuracy of 95% while considering only content-based features, 86% when considering sentiment-based features, and 93% while considering both the features together. Naive Bayes and Decision Tree classifiers perform best when considering only Sentiment

Table 3 Prediction comparison with the existing literature on Liar dataset

References	Model	Accuracy	Precision	Recall	F1score
Khan et al. (2019)	SVM	0.56	0.57	0.56	0.48
	LR	0.56	0.56	0.56	0.51
	Decision-Tree	0.51	0.51	0.51	0.51
	Ad boost	0.56	0.56	0.56	0.54
	Naïve-Bayes	0.60	0.59	0.60	0.59
	k-NN	0.53	0.53	0.53	0.53
Wang (2017)	SVM	0.25	-	-	-
	LR	0.24	-	-	-
Rubin and Lukoianova (2015)	SVM	0.57	0.59	0.53	0.54
Janze and Risius (2017)	SVM	0.73	0.64	0.70	0.67
Our method	Random Forest	**0.95**	0.96	0.95	0.95
	LR	0.83	0.83	0.83	0.83
	Naïve Bayes	0.83	0.83	0.83	0.83
	Decision Tree	0.87	0.89	0.87	0.86
	k-NN	0.94	0.94	0.94	0.94
	SVM[RBF]	0.64	0.79	0.64	0.59
	SVM[Linear]	0.93	0.94	0.93	0.93

Bold signifies the highest accuracy achieved by our method through random forest classifier

based features, while K-NN gives an accuracy of 94% when considering only content-based features. The comparison analysis has also been performed with the existing literature on the Liar dataset. The detailed evaluation measure, like Accuracy, Precision, Recall, and F-Measure, is considered for analysis, as shown in Table 3.

From the given Table 3, it can be clearly observed that the proposed method outperforms other methods (Khan et al. 2019; Wang 2017; Rubin and Lukoianova 2015; Janze and Risius 2017) applied to the same dataset. Here also, it can be observed that the Random Forest classifier outperforms all other with an accuracy of 95%. Most of the authors have applied the SVM classifier for evaluation, as we already discussed previously. It has been observed that among all given methods our method performs best on SVM, with an accuracy of 93% when applying Linear kernel, while (Wang 2017) performs worst using an SVM classifier.

4.2 Proposed dataset

The fake news prediction is one of the new fields, due to which very few standard datasets are available for evaluation. So, we have also included one of the key contributions in this paper by building our dataset and testing the proposed architecture on it. The dataset is the collection of 671(416 False statements and 255 True statements) short statements along with its veracity, collected from two well-known rumour debunking websites "Politifact" and "Snopes" related to the topics of politics and death hoaxes, along with it we have also used wiki death reports data.[7] The selenium Web driver has been used to automate the rumour debunking website for the collection of facts and veracity (True/False). The performance analysis of the model on the proposed dataset is described in Table 4.

Logistic regression is also found to be one of the good classification models in the previous studies, for the detection of false information (Bondielli and Marcelloni 2019). From the above table, it has been observed that the Logistic Regression classifier outperforms others with an accuracy of 96% by considering only content-based features, whereas Random forest gives the best results with an accuracy of 95% by considering only sentiment-based features, while it has been found that k-NN gives the best accuracy by considering only content-based features instead of using sentiment and hybrid-based features.

5 Conclusion and future work

In this paper, we provide a model called "Hoax News Inspector" which provides general solutions for data collection and data classification towards fake news detection. Data collection strategy is provided to collect data, and a machine learning solution is provided to detect fake news. We have used a content-based, sentiment-based, and Hybrid feature for the detection task. It has been observed that random forest classifier outperforms others in each of the given types of features, with an accuracy of 95%. Fake news detection is

[7] https://en.wikipedia.org/wiki/Wikipedia:Database_reports/Recent_deaths

Table 4 Performance of the Model on the proposed dataset

Feature	Classifier	Accuracy	TP-rate	FP-rate	Precision	Recall	F-measure
Content based	LR	0.96	0.96	0.039	0.962	0.96	0.96
Content based	Random Forest	0.95	0.95	0.04	0.95	0.95	0.95
Content based	Naïve- Bayes	0.68	0.68	0.31	0.74	0.68	0.66
Content based	k-NN	0.95	0.95	0.04	0.95	0.95	0.95
Content based	SVM	0.59	0.59	0.39	0.66	0.59	0.55
Sentiment based	LR	0.95	0.95	0.04	0.95	0.95	0.95
Sentiment based	Random- Forest	0.95	0.95	0.04	0.95	0.95	0.95
Sentiment based	Naïve-Bayes	0.94	0.96	0.03	0.96	0.96	0.96
Sentiment based	k-NN	0.94	0.94	0.05	0.94	0.94	0.94
Sentiment based	SVM	0.56	0.56	0.43	0.56	0.56	0.56
Hybrid feature(Content + sentiment)	LR	0.93	0.93	0.06	0.93	0.93	0.93
Hybrid feature(Content + sentiment)	Random-Forest	0.94	0.94	0.05	0.94	0.94	0.94
Hybrid feature(Content + sentiment)	Naïve-Bayes	0.68	0.68	0.30	0.74	0.68	0.66
Hybrid feature(Content + sentiment)	k-NN	0.93	0.93	0.07	0.93	0.93	0.93

one of the emerging topics, and there are several interesting options for future work. One of the important works that can be done to build an Image dataset that is embedded with textual data as very few datasets are available related to this. Also, our proposed framework could be extended to detect false information in real-time by implementing the plugin/extension. In the future, we are extending this work with respect to multimedia as well. The work can be extended with respect to multimedia data where a voice in a video can be processed via translating the speech content using google speech API into text and apply the analysis over it. We can extend this work with an image as well via extracting the text claim embedded in an image using an OCR technique, and extracted text can be processed that further be utilized in the verification of images attached with a false claim.

References

Afroz S, Brennan M, Greenstadt R (2012) Detecting hoaxes, frauds, and deception in writing style online. In: Proceedings of the 2012 IEEE symposium on security and privacy, pp 461–475

Ajao O, Bhowmik D, Zargari S (2018). Fake news identification on twitter with hybrid cnn and rnn models. In: Proceedings of the 9th international conference on social media and society, pp 226–230

Aker A, Derczynski L, Bontcheva K (2017) Simple open stance classification for rumour analysis. arXiv Preprint. arXiv:1708.05286

Bodnar T, Tucker C, Hopkinson K, Bilén S G (2014) Increasing the veracity of event detection on social media networks through user trust modeling. In: Proceedings of the 2014 IEEE international conference on big data (Big Data), pp 636–643

Bondielli A, Marcelloni F (2019) A survey on fake news and rumour detection techniques. Inf Sci 497:38–55. https://doi.org/10.1016/j.ins.2019.05.035

Breiman L (2017) Classification and regression trees. Routledge, Abingdon

Castillo C, Mendoza M, Poblete B (2011) Information Credibility on Twitter. In : Proceedings of the 20th international conference on world wide web, 675–684. https://doi.org/10.1145/1963405.1963500

Cervon G, Franzese P, Keesee APK (2010) Algorithm quasi-optimal (AQ) learning. Wiley Interdisciplinary Reviews: Computational Statistics, 2(2):218–236

Chakraborty A, Paranjape B, Kakarla S, Ganguly N (2016) Stop clickbait: Detecting and preventing clickbaits in online news media. In: Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp 9–16

Chen YC, Liu ZY, Kao HY (2017) IKM at SemEval-2017 Task 8: Convolutional neural networks for stance detection and rumor verification. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), pp 465–469

Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv Preprint. arXiv:1406.1078

Feng VW, Hirst G (2013) Detecting Deceptive Opinions with Profile Compatibility. In: Proceedings of the Sixth International Joint Conference on Natural Language Processing, pp 338–346. https://www.aclweb.org/anthology/I13-1039

Giasemidis G, Singleton C, Agrafiotis I, Nurse JRC, Pilgrim A, Willis C, Greetham DV (2016) Determining the Veracity of Rumours on Twitter. CoRR, abs/1611.0. http://arxiv.org/abs/1611.06314

Gupta A, Kumaraguru P, Castillo C, Meier P (2014) TweetCred: {A} Real-time Web-based System for Assessing Credibility of Content on Twitter. CoRR, abs/1405.5. http://arxiv.org/abs/1405.5490

Hardalov M, Koychev I, Nakov P (2016) In search of credible news. In: Proceedings of the International Conference on Artificial Intelligence: Methodology, Systems, and Applications, pp 172–180

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780

Horne BD, Adali S (2017). This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In: Proceedings of the Eleventh International AAAI Conference on Web and Social Media

Janze C, Risius M (2017) Automatic Detection of Fake News on Social Media Platforms. PACIS, p 261

Khan JY, Khondaker M, Islam T, Iqbal A, Afroz S (2019). A Benchmark Study on Machine Learning Methods for Fake News Detection. arXiv Preprint. arXiv:1905.04749

Kwon S, Cha, M, Jung K, Chen W, Wang Y (2013) Prominent Features of Rumor Propagation in Online Social Media. In: Proceedings of the 2013 IEEE 13th International Conference on Data Mining, pp 1103–1108

Kwon S, Cha M, Jung K (2017) Rumor detection over varying time windows. PLoS ONE 12:1–19. https://doi.org/10.1371/journal.pone.0168344

Lafferty J, Mccallum A, Pereira F (2002) Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. Proc ICML

Li Q, Hu Q, Lu Y, Yang Y, Cheng J (2019) Multi-level word features based on CNN for fake news detection in cultural communication. Personal and Ubiquitous Computing, 1–14. https://doi.org/10.1007/s00779-019-01289-y

Ma J, Gao W, Wei Z, Lu Y, Wong, KF (2015) Detect Rumors Using Time Series of Social Context Information on Microblogging Websites. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. https://doi.org/10.1145/2806416.2806607

Ma J, Gao W, Mitra P, Kwon S, Jansen B J, Wong K F, Cha M (2016). Detecting Rumors from Microblogs with Recurrent Neural Networks. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, 3818–3824. http://dl.acm.org/citation.cfm?id=3061053.3061153

Ma J, Gao W, Wong K F (2017) Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers): 708–717. https://doi.org/10.18653/v1/P17-1066

Pennycook G, Bear A, Collins E, Rand D G (2019) The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings. Management Science

Pérez-Rosas V, Mihalcea R (2015) Experiments in open domain deception detection.In: Proceedings of the 2015 conference on empirical methods in natural language processing, pp 1120–1125

Qazvinian V, Rosengren E, Radev D R, Mei Q (2011) Rumor Has It: Identifying Misinformation in Microblogs. In: Proceedings of the conference on empirical methods in natural language processing, pp 1589–1599. http://dl.acm.org/citation.cfm?id=2145432.2145602

Qin Y, Wurzer D, Lavrenko V, Tang C (2016) Spotting Rumors via Novelty Detection. CoRR, abs/1611.0. http://arxiv.org/abs/1611.06322

Quinlan J R (2014) C4. 5: programs for machine learning. Elsevier

Rubin VL, Lukoianova T (2015) Truth and deception at the rhetorical structure level. J Assoc Inf Sci Technol 66(5):905–917

Rubin V, Conroy N, Chen Y, & Cornwell S (2016) Fake news or truth? using satirical cues to detect potentially misleading news. In: Proceedings of the Second Workshop on Computational Approaches to Deception Detection: 7–17

Shu K, Sliva A, Wang S, Tang J, Liu H (2017) Fake news detection on social media: a data mining perspective. ACM SIGKDD Explor Newsl 19(1):22–36

Swift A (2016) Americans' trust in mass media sinks to new low. Gallup News 14

Vishwakarma DK, Varshney D, Yadav A (2019) Detection and veracity analysis of fake news via scrapping and authenticating the web search. Cogn Syst Res. https://doi.org/10.1016/j.cogsys.2019.07.004

Vosoughi S (2015) Automatic detection and verification of rumors on Twitter. PhD diss Massachusetts Institute of Technology

Vosoughi S, Mohsenvand MN, Roy D (2017) Rumor Gauge: predicting the veracity of rumors on twitter. ACM Trans Knowl Discov Data 11(4):50. https://doi.org/10.1145/3070644 (**(1–50:30)**)

Wang W (2017) Liar, Liar pants on fire: a new benchmark dataset for fake news detection. arxiv:1705.00648

Wang S, Terano T (2015). Detecting rumor patterns in streaming social media. In: Proceedings of the 2015 IEEE International Conference on Big Data (Big Data): 2709–2715

Wang F, Lan M, Wu Y (2017). ECNU at SemEval-2017 Task 8: Rumour evaluation using effective features and supervised ensemble models. In : Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017): 491–496

Wojtusiak J, Kaufman KA (2010) Ryszard S. Michalski: the vision and evolution of machine learning. Adv Mach Learn I:3–22

Wu K, Yang S, Zhu K Q (2015) False rumors detection on Sina Weibo by propagation structures. In: Proceedings of the 2015 IEEE 31st International Conference on Data Engineering, pp 651–662

Yang F, Liu Y, Yu X, Yang M (2012) Automatic Detection of Rumor on Sina Weibo. In: Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics, 13:1–13:7. https://doi.org/10.1145/2350190.2350203

Yu F, Liu Q, Wu S, Wang L, Tan T (2017) A Convolutional Approach for Misinformation Identification. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, {IJCAI-17}, 3901–3907. https://doi.org/https://doi.org/10.24963/ijcai.2017/545

Zeng L, Starbird K, Spiro ES (2016) unconfirmed: Classifying rumor stance in crisis-related social media messages. In: Proceedings of the Tenth International AAAI Conference on Web and Social Media

Zhang H, Fan Z, Zheng J, Liu Q (2012) An improving deception detection method in computer-mediated communication. Journal of Networks, 7(11):1811

Zubiaga A, Liakata M, Procter R (2016) Learning Reporting Dynamics during Breaking News for Rumour Detection in Social Media. arxiv:1610.07363

Zubiaga A, Liakata M, Procter R (2017) Exploiting context for rumour detection in social media. In: Proceedings of the International Conference on Social Informatics, pp109–123

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

Check for updates

Article

Impact of IRDA Guidelines on Consumer Confidence in Life Insurance Market: Then and Now

Jindal Journal of Business Research
9(2) 117–134, 2020
© 2020 O.P. Jindal Global University
Reprints and permissions:
in.sagepub.com/journals-permissions-india
DOI: 10.1177/2278682120968985
journals.sagepub.com/home/brj



Shikha Khera[1] and Divya[2]

Abstract

This study has been undertaken in order to estimate the impact on IRDA guidelines on customer satisfaction in the life insurance sector from the time of its inception in 1999 to the IRDA Amendment (The Insurance Laws Act) of 2015. In order to accomplish the research aim and objectives appropriately, a thematic framework has been implemented that aspires to conduct quantitative analysis of the primary data collected in relation to the research topic through questionnaire-based survey. Further, comparative survey of customer confidence on IRDA guidelines pertaining to pre- and post-IRDA regulation 2017 has been made in the study. The study findings confirm that IRDA guidelines, that is, transparent and make clear specifications on aggressive online selling by insurance companies, provide outsourcing guideline for them, direct the customers on timely settlement of claims without delay, and are abided by the mandate of Government of India to link PAN and Aadhaar card of the consumers to their policies show significantly positive impact on consumer confidence. Simply put, there is significant increase in confidence post-2015 Amendment Act.

Keywords

IRDA Act 1999, Amendment Act 2015, life insurance, customer confidence, impact, India

Introduction

Regulatory Architecture in Insurance Sector: Emphasis on Life Insurance Market

The Insurance Regulatory and Development Authority (IRDA) of India is an autonomous, statutory body that regulates the insurance industry in India. The Government of India established it for regulating, directing, and controlling the insurance sector to ensure its smooth functioning. Constituted by a Parliament of India Act called Insurance Regulatory and Development Authority Act, 1999 following the recommendations of Malhotra Committee (1994) and passed by the Government of India in April 2000, IRDA's

[1] Delhi School of Management, Delhi Technological University, Delhi, India.
[2] Department of Management Studies, Central University of Haryana, Mahendergarh, Haryana, India.

Corresponding author:
Divya, Department of Management Studies, Central University of Haryana, Jant-Pali, Mahendergarh, Haryana 123031, India.
E-mail: divya7500@gmail.com

Table 1. List of Registered Insurers (IRDA, 2019)

Type of Insurer	Public Sector	Private Sector	Total
Life	1	23	24
General	6	17	23
Health	0	6	6
Reinsurers	1	8	9
Total	8	54	62

Source: IRDA Annual report, various issues.

objectives include promoting competition to enhance customer satisfaction with increased consumer choice and lower premiums while ensuring financial security of the insurance market. IRDA also regulates investment of funds by insurance companies, regulating maintenance of margin of solvency, adjudication of disputes between insurers and intermediaries or insurance intermediaries. The IRDA Act, 1999 allows private players to enter the insurance sector besides a maximum of foreign equity of 26 percent in a private insurance company operating in India. In December 2000, the subsidiaries of the General Insurance Corporation of India were restructured as independent companies and it was converted into a national reinsurer. IRDA was amended in 2002. At the end of financial year 2016–2017, there are 62 insurers operating in India out of which 24 are life insurance companies, 23 are general insurers, 6 are health insurers involved in health insurance business exclusively and 9 reinsurers including foreign reinsurers' branches and Lloyd's India. As far as the classification of insurers into public and private sector is concerned, 8 are operating in the public sector and the remaining 54 in the private sector (Table 1).

Clearly the number of life insurers is more than the number of insurers serving other products or purposes. Life insurance industry earned a premium income of 41,84,768.2 million during 2016–2017 compared to 36,69,432.3 million in the previous financial year 2015–2016 (IRDAI, 2018). This translates to a growth of 14.04 percent and an increase in the growth by 2.2 percent. The growth of the private sector in terms of the premium income of its insurers which was 17.40 percent (as compared to 13.64% in the previous year) was more than the growth of the Life Insurance Corporation of India given by 12.78 percent (as compared to 11.17% in the previous year). Growth in the private sector can be attributed to the rise in the FDI limit in the insurance sector which was raised to 49 percent in June 2016 from its previous level of 26 percent through the automatic route.

As per section 4 of the IRDAI Act 1999, IRDA of India specifies the architecture as:

> The authority consists of a 10-member team of which 1 is Chairman, 5 are whole time members, and 4 are part time members. All the members are appointed by the Government. According to the IRDAI Amendment Act, 2015, with respect to life insurance market, an insurer will be eligible to be appointed as an Appointed Actuary provided some conditions are met. Some of these conditions are: first, the person is a resident of India; second, the person has passed specialization subject in life insurance; third, the person has the relevant experience of at least 10 years in life insurance industry of which minimum of 5 years will be post-fellowship experience; and fourth, the person has at least 3 years of post-fellowship experience in annual statutory valuation of a life insurer.

Customer Growth Picture in Life Insurance Market of India

Growth trends in registered life insurers have been significant since the establishment of IRDA. Increase in the private insurer is more than the public sector insurer represented solely by LIC. Based on the total premium income, the market shares of LIC went down from 72.61 percent in 2015–2016 to 71.81 percent

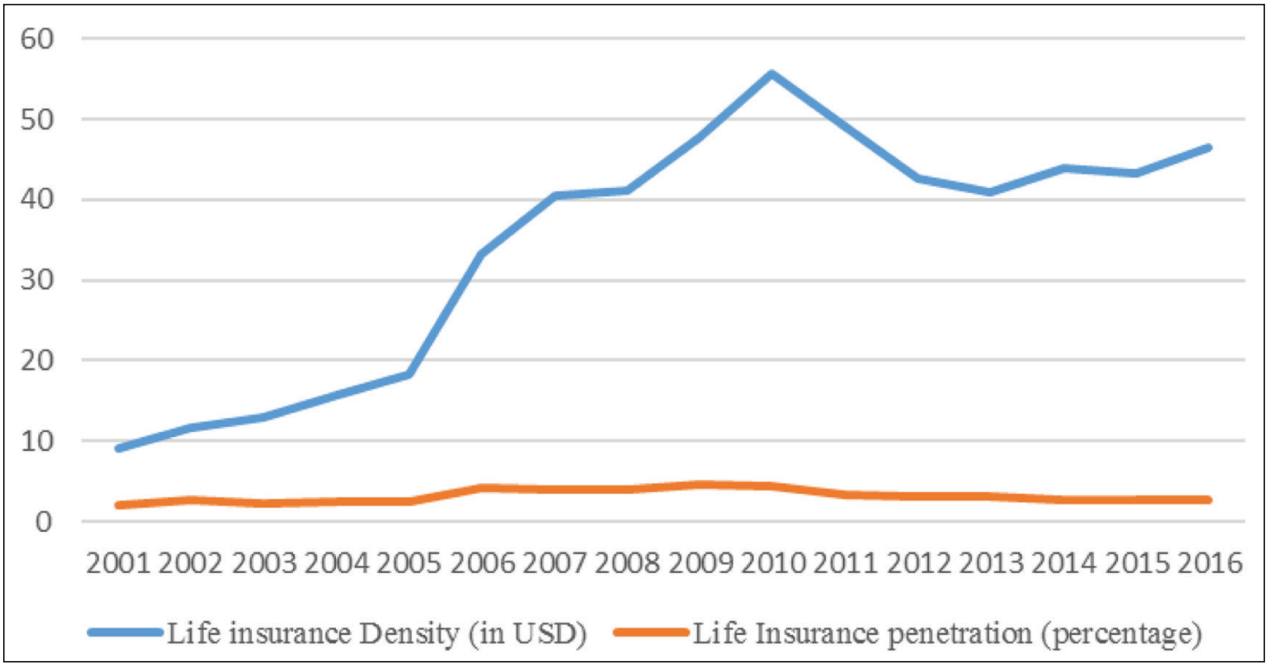


Figure 1. Life Insurance Penetration and Life Insurance Density in India

Source: IRDA Annual Report 2016–2017.

in 2016–2017. On the contrary, market share of private insurers increased to 28.19 percent in 2016–2017 from 27.39 percent in 2015–2016 (IRDAI, 2018).

Customer growth picture in the life insurance market is captured by indicators such as insurance penetration and insurance density (Bharadi, 2012). Insurance penetration is defined as the percentage of insurance premium to GDP and insurance density is measured as the ratio of premium to population or per capita premium. The first decade after the insurance sector liberalization witnessed a steady increase in insurance penetration. It reached its peak at 4.6 percent in 2009. The insurance density also experienced a growth reaching its peak at $55.7 in 2010. Figure 1 represents the insurance penetration and insurance density in India for the time period 2001–2016 (IRDAI, 2018).

Also private life insurers issued 6,324 new policies in 2016–2017 and LIC issued 20,132 new policies. The former reflected a growth of 2.13 percent in the number of new policies issued whereas LIC faced a decrease of 2.02 percent. Much of the growth in the private life insurance market happened due to the establishment of IRDA.

Need for Study

The facts and figures presented in Figure 1 indicate that IRDA has been playing a crucial role in the Indian life insurance sector. The growth of the insurance market also indicates that there is increasing demand in the market with respect to insurance product, marketing, competition and customer's awareness (Rewadikar, 2013). Introduction of large number of new life insurance policies every year proves that consumers are confident about investing in the insurance products. Therefore, there is a definite scope to study the phenomenon of consumer confidence in the life insurance market in India since the

inception of IRDA Act in 1999 to 2015 IRDA Amendment Act, which made notable revisions in the facilities, notably protection of interests of insurance policyholders; insurers, distribution channels, and other regulated entities fulfill the promises they made to the policyholders; and standard procedures and best practices in sales and services of insurance policies are maintained (IRDAI, 2017). However, since this study purports to observe the consumers' behavior against the amendment in the act, capturing the behavioral change of the past 20 years (1999–2018) was practically difficult. Therefore, a period of 10 years (2007–2017) has been considered to conduct the study and establish observations, covering the pre-2015 situation (then) and post-2015 situation (now).

Aim and Central Question

The aim of the study is to investigate the impact of most recent 2015 IRDA amendment act vis-à-vis the IRDA guidelines of 1999 on the behavior of the consumers by doing a comparative analysis of confidence of those consumers who have been purchasing life insurance products for 10 years and more. This study tries to answer the following question:

Is there a change in consumer confidence in the life insurance market from 1999 IRDA guidelines vis-à-vis the recent IRDA Amendment Act of 2015?

Literature Review

IRDA Guidelines in Life Insurance Market: Critical Analysis of the Past 10 years

After being constituted in 1999, IRDA focused on increasing competition in the life insurance market as at that time, there was monopoly enjoyed by the Life Insurance Corporation of India. Along with bringing competition in the market, the other objectives were to increase customer satisfaction by providing them with more choice and to allow them to pay lower premiums. It also aimed that the insurance market becomes more secure financially. The market was opened up for registration applications in August 2000 where foreign companies could have ownership up to 26 percent. As specified in Section 114A of the Insurance Act, 1938, IRDA possesses the power to form guidelines and since 2000, these guidelines included company registrations and the protection of policyholders' interests. The regulatory changes also included minimum death benefit structure and structure to surrender value payment (Iyer, 2014). IRDA revised its guidelines from time to time in order to increase policyholders' interests and welfare. In the year 2010, guidelines towards Unit Linked Insurance Policies were revised. Lock in period was increased from 3 to 5 years. Minimum payment term was also increased to 5 years. Mandatory insurance cover was made for all insurance products. IRDA also introduced maximum limit on expenses. The year 2015 also brought significant changes in the IRDA guidelines (Iyer, 2014). Insurers were directed to launch products that complied with these revised guidelines. It also required that the existing life insurance products which were noncompliant could not be brought to the market beyond the stipulated deadline of December 31, 2015. All these guidelines were made to ensure growth of insurance business and increase the life insurance penetration. The FDI limit in the insurance sector was increased from 26 to 49 percent in June 2016 in accordance with the attempt of IRDA in 2015.

Regulation of Investment

In IRDA Act of 1999, prohibition was imposed on investment of funds outside India. Under section 27C of the act, it was specifically mentioned that no insurer will be able to invest the funds of the policy

holders directly or indirectly outside India. However, under section 27D certain guidelines were specified on the manner and conditions on investment. It was stated that without prejudice to any term contained in any other sections of the act, the authority may specify the guidelines made by the act regarding the time, manner and conditions of investment of assets totality in the interest of the policy holders. Moreover, such investments can be made in infrastructure and social sector. In addition, it was also stated that such guidelines on investment will be uniformly applicable to all the insurers carrying on the business of life insurance, general insurance, or re-insurance in India on or after the commencement of the Insurance Regulatory and Development Authority of India Act, 1999 (Insurance Regulatory and Development Authority of India, 2007). The Insurance Act previously allowed foreign investment up to 26 percent of paid up equity capital in Indian insurance companies (CUTS, 2016). However, while discussing the Union Budget of 2004–2005, it was mentioned by the-then finance minister of India that the IRDA projects to increase foreign direct investment cap in insurance to 49 percent from 26 percent. Finally, in the Insurance Laws Amendment bill of 2015, it was stated that the cap of foreign direct investment and portfolio investment in the insurance sector would be raised to 49 percent (Pandey, 2017). Guidelines were also made since IRDA Act of 1999 regarding increment of capital from domestic sources. Thus, while in 1999 the insurance companies were only allowed to have capital in their form of shares, by 2015, guidelines were issued by IRDA that allowed the companies to o have capital in form of preference shares and subordinate debt. It has additionally been specified that the insurance companies can raise their capital shares by means of public issue on completion of ten years from the date of commencement of business, or any other period prescribed by the central government (CUTS, 2016).

Expenditure by Insurance Companies

The IRDA Act of 1999 made some guidelines regarding funds and expenditure made by insurance companies. In this, it was specified that a fund shall be maintained by insurance companies by the name The Insurance Regulatory and Development Authority of India Fund where all the governmental grants, fees and charges received by the authority shall be deposited, This fund should be exclusively used for incurring expenses like salaries, allowances and other remuneration of the members, officers and other employees of the Authority along with meeting other expenses that are related to the discharge of its functions and for the purposes of this Act (Insurance Regulatory and Development Authority of India, 2007). While previously IRDA Act capped commissions or remunerations paid to the insurance agents in any form in life insurance business, according to the type of insurance, term, and year of premium, in the Act of 2014 this clause was eliminated. Instead, it was stated that while making guidelines with respect to remuneration of intermediaries, factors such as nature and tenure of the policy and in particular the interest of the agents and other intermediaries concerned shall be taken into consideration. Thus, presently, the commissions have been linked with the premium collected on the relevant insurance products. It has further been stated that in case of single premium products commissions or remunerations meant for the procurement of all individual policies in respect of all distribution channels, excluding the direct marketing shall not exceed 2 percent of the single premium. Guidelines in management expenses have also been brought in the IRDA Act of 2015. In this, it has been stated that certain exemptions in management expenses would be provided based on factors such as the size, age of the insurance company as well as the type of business segment. For instance, private insurance firms incurring high initial setup costs shall be exempted from paying this expense for a five years from the commencement of business operations, from compliance of the mentioned rules (CUTS, 2016).

Retention of Customers

One of the core objectives of IRDA Act of 1999 had been to work towards customer retention. For this, the Authority formulated a plan of action to ensure that customers receive clear, precise, and correct

information about the products a services they buy. Additionally, the authority also considered it as a part of this plan to make the customers aware of their subsequent duties and responsibilities related to specific insurance plans. Further, it was included in the action plan to IRDA to work towards speedy settlement of genuine claims, prevent insurance frauds and other malpractices, and set up effective grievance Redressal machinery through which customers can lodge their complains. As guidelines on premium and tariff also determine the rate of customer retention, so IRDA Act of 1999 laid out that the annual premium that every insurer is supposed to make to the Advisory Committee should not exceed 1 percent of the total premiums in respect of facultative insurance accepted by the customer in India. It was also specified that in case of any other insurance business, this premium should not exceed more than 1 percent of the total gross premium written direct by the customer in India (Singh, 2010; Yadav & Mohania, 2014). As part of the customer retention policy, IRDA amendments of 2015 have reconsidered the important aspects such as persistency rate, orphan policy and deferred commissions. While the average persistency rate for each agent for financial year 2011–2012, 2012–2013, and 2013–2014 was 50 percent, in 2014–2015 it became 75 percent. Further guidelines were made from July 1, 2014 onward that stated that life insurers were required to have their own company specific persistency criterion for renewal of individual and corporate agency for removing the requirement of maintaining 50 percent persistency rate. In order to agent credibility for the customers, it was also made a mandate in IRDA Act of 2015 that agents must maintain a record of the policies sold as well as their persistency on a year to year basis and the insurer was required to endorse by the same was at the end of the year (CUTS, 2016).

Impact of IRDA on Consumer Confidence: Systematic Review with Instances from Past 10 Years

One of the key themes identified in correspondence to impact of IRDA guidelines on customer confidence is customer satisfaction. In this context, the research conducted by Siddiqui and Sharma (2010) focused on the factors that determine satisfaction of customers on the service quality of insurance sector. The findings suggest that cardinal features in insurance products in the form of assurance, personalized financial planning, competence, corporate image, tangibles, and technology operate as primary determiners of customer satisfaction. Similar research was conducted by (Kathirvel & Radhamani, 2013) on the factors influencing the satisfaction of customers with the insurance sector. The findings presented additional factors in addition to those specified by Siddiqui and Sharma in the form of comfortable environment in the insurance office, transparency with insurance officials and exposure to new strategies undertaken by the insurance companies to make the services offerings more effective and efficient as factors that determine the satisfaction of insurance policy holders towards their insurance products.

The next theme identified is identification of risks and significance of risk management strategy on the insurance sector. In relevance to this theme identified, Pandey and Rao (2013) conducted research presents that factors like entry of private players, corresponding policy changes and the present day fact of unprofitable books, erosion of capital resulting from unmanageable claim ratios have increased risks in the insurance sector. Hence, implementation of risk management strategy has become integral in the sector because it safeguards the investors as well as investment. Research on the similar topic conducted by Eikenhout (2015) further suggests that identification of risk management strategy helps an insurance company in developing a systematic approach by virtue of which the unprecedented risks and uncertainties could be effectively identified. The scholar further states the accurate identification of these risks and timely implementation of risk management strategies shows positive impact on the performance as well as the value of the insurance company. In context to identification of risks in the insurance sector, the

study conducted by Eling and Pankoke (2016) suggest that traditional insurance activity in the life, nonlife, and reinsurance sectors do not contributes to systemic risk. These activities do not increase the insurers' vulnerability to impairments of the financial system either. Rather, nontraditional insurance activities in the form of credit default swap underwriting usually contribute to systematic risks in the insurance sector and make the financial system of the insurers susceptible to impairment. Such activities even make the life insurers more vulnerable than nonlife insurers due to higher leverage.

Based on the above exploration of literature highlighting the change in IRDA Act since its inception to till date and the resultant impact on customer confidence, this study proposed the following hypothesis for further investigation:

H_0: *There is no difference in impact of IRDA guidelines on consumer confidence from then (pre-2015) to now (post-2015).*

Research Methodology

Research Design

This study aimed to measure the impact of IRDA guidelines on consumer confidence, a comparative analysis of pre- and post-IRDA amendment act of 2015, based on a positivist belief, which lies in understanding the of real world objects existing beyond human knowledge (Kaboub, 2008). Based on the positivist paradigmatic approach this study applied the descriptive as well as explanatory methods. Under descriptive, the demographic distribution of the respondents and their general perception were presented (Bhattacharjee, 2012), while under explanatory, the causality of the IRDA Act and amendments on consumer confidence was measured (Morgan, 2014). Following the application of method, a quantitative approach of data collection was entailed which was essentially carried out using survey strategy.

Survey Participants

In order to determine whether there is any impact of IRDA guidelines on consumer confidence, the survey respondents needed to be familiar with life insurance market for a long period of time. The sampling method in the study was selective sampling where inclusion criteria was that the customers invested in their life insurance products for more than 10 years.

Initially a pool of 500 respondents was targeted for the survey ($N = 500$). But many of them were not willing to participate in the survey and among those who were willing, many did not turn in their responses within due time or their responses were incomplete. Finally, responses from 300 participants ($N = 300$) were collected and analyzed. Therefore, a response rate of 60 ([300/500]×100) was obtained for the study. These participants answered the questions asked in the survey was valid enough to proceed with the analysis. The survey was conducted in Delhi NCR with the only inclusion criteria mentioned above.

Measurement Instruments and Measures

The questionnaire—a structured and close-ended in nature—was designed in such a way that customers were able to reveal whether there has been a difference in their confidence level towards increasing their investment cap on life insurance and increasing purchase of new life insurance policies online.

A pre- and post-test design based experimental survey was implemented wherein the respondents were asked whether factors related to IRDA amendment act 2015 were able to cause a rise in their confidence level.

Procedure of Data Collection and Analysis

Data collection was carried using a purposive sampling approach, wherein the respondents were selected with the criterion that they are insurers for at least 10 years. This inclusion criterion was adopted to ensure that the respondents are aware of the changes occurred in IRDA act for a prolonged time and hence will be able to capture the impact on their overall confidence appropriately. Respondents' information was obtained through personal visit to select insurance offices and explaining the Branch Manager about the study aim. Utmost care on the part of the researcher was taken to maintain ethical consideration in the data collection. All the respondents were informed about the motive of the study and its aim and objectives. They were not forced any way to participate and were allowed to withdraw any time. Furthermore, complete anonymity of the respondents were promised and maintained throughout the study.

Findings

Demographic Profile

The demographic profile consisted of information collected on gender, age, educational qualification, job category of employment, annual income of the respondents, and the insurers from whom they bought their life insurance.

Frequency distribution of demographic profile of participants in Figure 2 reveals that out of 300 respondents, percentage of male was higher than the percentage of female. Also that most of the participants who purchased life insurance, belonged to the age group 45–54. Contrarily, respondents who are aged 65 years and above were the least to have insurance. Further, it was noticed that most of the insurance holders earn 800,000–1 million per annum whereas people from lower earning category like 400–600 thousand per annum buy the least number of insurance products. The annual income level distribution of the respondents reveals that since all of them had been earning at least ₹400,000 per year, they were in a good position financially to invest in the life insurance market.

General Background

The general background of the survey participants revealed information about their behavioral pattern regarding life insurance market.

While analyzing through the frequency distribution of general background of the respondents in Figure 3, it was found that more than 78 percent of the participants have been investing in insurance for more than 11 years. And most commonly used reasons for keeping an insurance were to save taxes, to incur regular retirement income and futureproofing for risks. However, when they were asked the type of insurance they buy, most of them did not have preferences on any particular sector and instead they

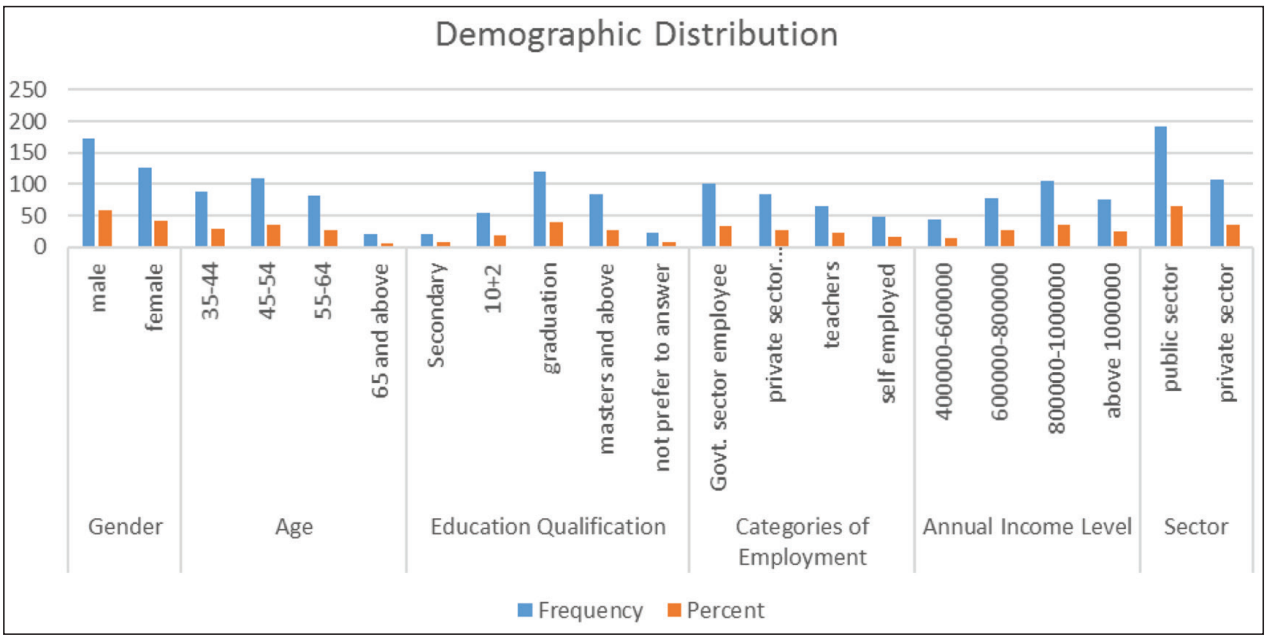


Figure 2. Frequency Distribution of Demographic Profile

Source: The authors.

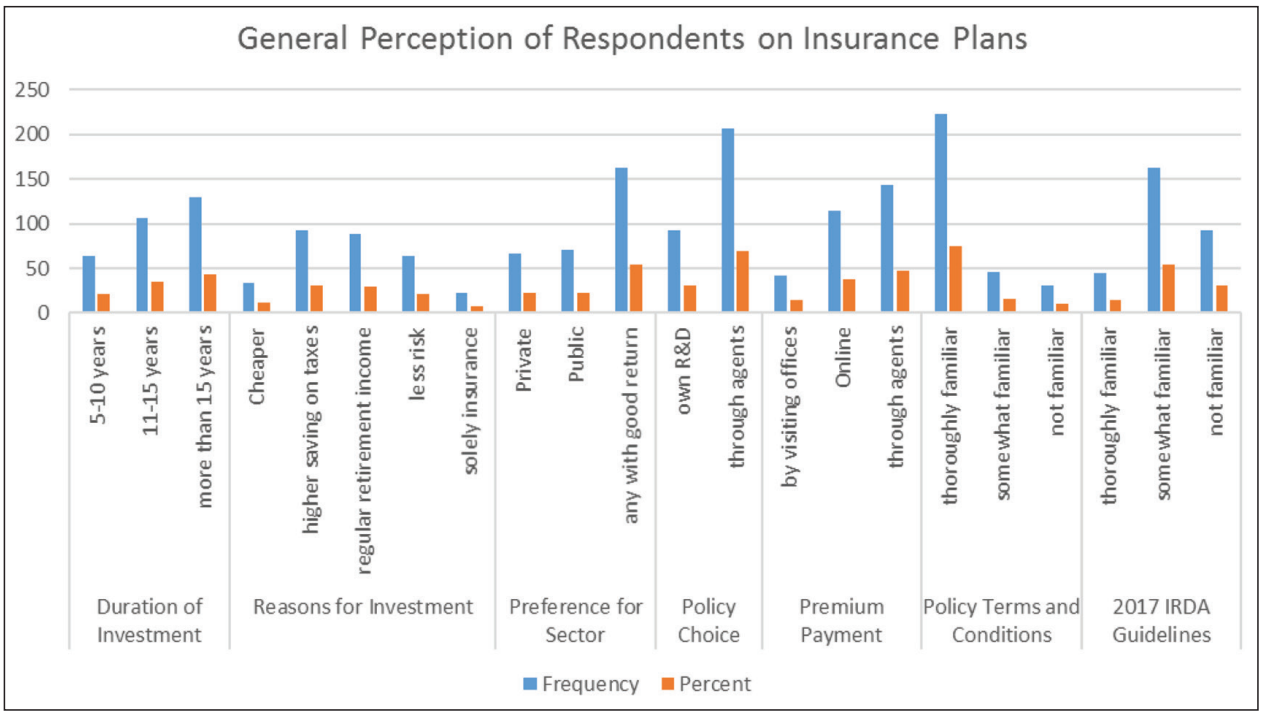


Figure 3. Frequency Distribution of General Background

Source: The authors.

were inclined to buy any insurance package with good returns. Moreover, almost 70 percent of the respondents have agreed that they did not personally research about the type and specifications of the insurance, instead relied on the recommendations from agents and buy from them. With the thorough advancement in technology some of the insurance buyers started to pay their insurance premium online, but the majority of them still depends on their agents for paying insurance premiums. It can be inferred that since most of the survey participants started investing in life insurance more than 10 years ago when there was less scope of obtaining information online compared to today, they had to rely on agents and they continued doing so since then. When the respondents were asked whether they were familiar with the terms and conditions of the policies they purchased. It was found that three fourth of the total participants were thoroughly familiar with the T&Cs. Together, almost 90 percent of them were familiar to some degree implying that buyers of life insurance in India are generally aware of the terms and conditions of the policies at the time of investing. Lastly, when the participants were asked whether they were familiar with the IRDA amendment act of 2015. Almost 70 percent of them revealed that they had some degree of familiarity with these guidelines implying that majority of the respondents were well aware of the changes made by the regulatory authority that is IRDA in the insurance sector to safeguard and protect their interests.

Wilcoxon Signed Rank Test Analysis

The Wilcoxon signed rank test was performed in two steps. In the first step, it was analyzed whether there was an increase in consumer confidence because of the IRDA Amendment Act (Table 2), 2015 pertaining to five factors which were restrictions on aggressive online selling activities, clear outsourcing guidelines, timely settlement of claims, mandatory Aadhaar linkage, and provision of free look-up period. In the second step, it was analyzed whether there was an increase in consumer confidence reflected by their investment decisions which were increase in investment cap and increase in online purchase of life insurance.

Tables 4 and 5 show the results of the rank test in the first step. Pre-test and post-test represent responses of the participants in the pre-2015 Amendment Act (then) and post-2015 Amendment Act (now). Pre-test 1 and post-test 1 are the responses due to the factor of restrictions on selling activities, pre-test 2 and post-test 2 are those due to the factor of outsourcing, pre-test 3 and post-test 3 are those due to the factor of timely settlement, pre-test 4 and post-test 4 are responses due to the factor of Aadhaar linkage and pre-test and post-test 5 are those due to the factor of free lookup period. Before moving further, it is clear from Table 2 that for four out of five factors, mean positive rank is higher than the mean negative rank indicating that there is a definite increase in consumer confidence due to these recent amendments. Timely settlement in claims have helped the consumer to trust their respective insurers. As timely and effective communication and resolution helps the customers in buying again with confidence (Nebo & Victor O, 2016). In other words, it confirms that the consumers have started to put more faith in insurance companies while purchasing insurances after the changes in the IRDA guidelines. This is further validated by the value of the respective test statistics presented in Table 5.

For the first three factors, the null hypothesis has been rejected at 1 percent level of significance (Table 3). The fourth factor rejected the null hypothesis at 5 percent level of significance. Only the fifth factor showed a nonsignificant impact on consumer confidence, for which the null hypothesis was accepted. This shows that the amendments like restricting aggressive online selling activities, clearing outsource guidelines, settling claims on time and mandating Aadhaar linkage have strongly influenced the confidence and purchasing behavior of the customers in a positive way. These changes in the IRDA

Table 2. Wilcoxon Signed Rank Test Results for Factors of IRDA Guidelines, 2015

		N	Mean Rank	Sum of Ranks
Post-test 1–pre-test 1	Negative ranks	20[a]	83.30	1666.00
	Positive ranks	188[b]	106.76	20070.00
	Ties	92[c]		
	Total	300		
Post-test 2–pre-test 2	Negative ranks	47[d]	71.30	3351.00
	Positive ranks	130[e]	95.40	12402.00
	Ties	123[f]		
	Total	300		
Post-test 3–pre-test 3	Negative ranks	26[g]	26.62	692.00
	Positive ranks	43[h]	40.07	1723.00
	Ties	231[i]		
	Total	300		
Post-test 4–pre-test 4	Negative ranks	74[j]	76.85	5687.00
	Positive ranks	92[k]	88.85	8174.00
	Ties	134[l]		
	Total	300		
Post-test 5–pre-test 5	Negative ranks	86[m]	90.39	7773.50
	Positive ranks	85[n]	81.56	6932.50
	Ties	129[o]		
	Total	300		

[a]Post-test 1 < pre-test 1
[b]Post-test 1 > pre-test 1
[c]Post-test 1 = pre-test 1
[d]Post-test 2 < pre-test 2
[e]Post-test 2 > pre-test 2
[f]Post-test 2 = pre-test 2
[g]Post-test 3 < pre-test 3
[h]Post-test 3 > pre-test 3
[i]Post-test 3 = pre-test 3
[j]Post-test 4 < pre-test 4
[k]Post-test 4 > pre-test 4
[l]Post-test 4 = pre-test 4
[m]Post-test 5 < pre-test 5
[n]Post-test 5 > pre-test 5
[o]Post-test 5 = pre-test 5

Source: Authors' calculations based on the IRDA reports.

Table 3. Wilcoxon Test Statistic Values for the Factors of IRDA Guidelines, 2017

	Post-test 1–Pre-test 1	Post-test 2–Pre-test 2	Post-test 3–Pre-test 3	Post-test 4–Pre-test 4	Post-test 5–Pre-test 5
Z	−11.037[a]	−6.927[a]	−3.184[a]	−2.163[a]	−0.701[b]
Asymp. Sig. (2-tailed)	0.000	0.000	0.001	0.031	0.484

[a]Based on negative ranks.
[b]Based on positive ranks.

Source: Authors' calculations based on IRDA report, 2017.

Table 4. Wilcoxon Signed Rank Test Results for Consumer Confidence

		N	Mean Rank	Sum of Ranks
Post-test 6–pre-test 6	Negative ranks	13[a]	63.15	821.00
	Positive ranks	249[b]	135.07	33632.00
	Ties	38[c]		
	Total	300		
Post-test 7–pre-test 7	Negative ranks	5[d]	93.00	465.00
	Positive ranks	262[e]	134.78	35313.00
	Ties	33[f]		
	Total	300		

[a]Post-test 6 < pre-test 6
[b]Post-test 6 > pre-test 6
[c]Post-test 6 = pre-test 6
[d]Post-test 7 < pre-test 7
[e]Post-test 7 > pre-test 7
[f]Post-test 7 = pre-test 7

Source: The authors.

Table 5. Wilcoxon Test Statistic Values for Consumer Confidence

	Post-test 6–pre-test 6	Post-test 7–pre-test 7
Z	−13.970[b]	−14.461[b]
Asymp. Sig. (2-tailed)	0.000	0.000

[a]Wilcoxon Signed Ranks Test.
[b]Based on negative ranks.

Source: The authors.

guidelines have certainly helped the consumers while buying the insurances. Though, fifth factor, the provision of free look-up period did not aid in boosting the confidence of consumers in any way. The reason behind could be the short length of look-up period provided by different insurers which usually lasts for 10 days or more. There are the type of consumers exist in market who only seeks information and takes opportunity for self-gratifying reasons when it comes to free sampling (Carrie et al., 2004). Furthermore, the short time involved in free look-up period is not enough to understand the advantages and disadvantages of insurance products. Overall it shows, now that the IRDA has changed some of the guidelines in favor of consumers, people feel more secured and satisfied while buying insurance products.

The implication of the first step of the analysis is that consumer confidence has increased due to four out of five factors that were part of IRDA amendments. Due to this increase, they were expected to reflect this in their investment decisions. This led to the second step of the rank test analysis where the survey participants were asked whether their life insurance investment cap increased and whether they purchased new life insurance policies online rather than buying through insurance agents compared to the pre-2015 amendment situation.

Similar to the first step, pre-test and post-test represent responses for before and after 2015 Amendment Act. Pre-test 6 and post-test 6 represent consumer confidence in terms of increase in investment cap and pre-test 7 and post-test 7 represent consumer confidence reflected by an increase in online purchase.

Both the increases measured by the positive ranks as the difference between post-test responses and the pre-test responses are found to be statistically significant at 1 percent level. In these cases, the null hypothesis is rejected. It implies that consumer confidence in the life insurance market has significantly increased post-IRDA guidelines indicating an increase in investment cap and online purchases. The search of information is an important factor that helps the customers to find the suitable products or services according to their needs (Mittal, 2013). Put otherwise, these guidelines have had a positive and significant impact on consumer confidence as it lets them take their own time to research and compare insurance policies online to make a responsible decision, whereas the increase in cap has helped assuring the consumers to take a more confident decision.

Conclusion

Recommendations

In this section, perspective of three different entities are involved which are IRDAI, insurance companies and consumers. Based on these results, it is recommended that IRDAI should come up with such guidelines in the future that not only help consumers to keep their faith in the market, but also increase the growth rate of the insurance industry as a whole. For the insurance companies, either they should increase the time involves in free look-up period or should rather focus on some more evident factors mentioned above that wins consumer confidence to see a significant rise in their consumer base and retaining them. Further, they should be more transparent to the consumers by providing complete and comprehensible details of the insurance products to their consumers beforehand. Consumers can simply start doing their own research understand the insurance T&Cs before investing in insurance instead of relying on their agents.

Limitations and Future Scope

Despite the fact this study has established some proven realities, it still suffered some of the limitations like availability of sufficient data, and relatively new topic to find the supporting studies. The volume of total researches on the impact of changes in IRDA guidelines on consumer behavior is very limited and researchers had to rely upon the documentation and information available from other countries to develop the conceptual foundation for this study. Along with that, this study was conducted taking the survey sample data from one part of India which restrained the study to understand the behavior and confidence level of consumers buying insurance from different part of India. It would have been possible if more time was granted to collect and analyze the data in depth.

Findings of this study can be taken forward in future by collecting data from more sources and areas which help in a better understanding of consumer behavior in different places or all over India. An adequate size of sample data can be collected in order to derive more close to reality judgements.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

Appendix I: Questionnaire

This questionnaire aims to help the researcher to determine if there has been any boost in the confidence of customers of selected life insurance companies of India over the period 2008–2017. Your participation in the survey will be highly appreciated. In doing so, the researcher ensures the anonymity of the participant and complete confidentiality of data collected.

Section A: Demographic Profile

1. Gender
   - Male
   - Female

2. Age
   - 35–44
   - 45–54
   - 55–64
   - 65 and above

3. Educational Qualification
   - Secondary
   - 10+2
   - Graduation
   - Masters and above
   - Not prefer to answer

4. Job category of employment
   - Government sector employee
   - Private sector employee
   - Teachers
   - Self-employed

5. Annual Income Level (Taking into consideration the entire time period 2008–2017)
   - ₹4,00,000–₹6,00,000
   - ₹6,00,000–₹8,00,000
   - ₹8,00,000–₹10,00,000
   - Above ₹10,00,000

6. Which of the following sectors do you buy your life insurance from?
   - Public sector
   - Private sector

Section B: General Background

7. What is the duration of your investing in life insurance schemes?
   - 5–10 years
   - 10–15 years
   - More than 15 years

8. Why do you think investing in life insurance is beneficial?
   - It's cheaper if investment is made at a young age.
   - Higher savings on taxes.
   - Provides regular income after retirement.
   - Less risky than other investment options.
   - Solely for insurance purpose.

9. Preference for sector?
   - Private
   - Public
   - Any with good returns/schemes.

10. How do you choose which life insurance policies are most beneficial for you?
    - By doing own R&D
    - I rely on my life insurance agents

11. How do you pay your life insurance premium?
    - By visiting the insurance offices
    - Online
    - Through life insurance agents

12. Are you familiar with the terms and conditions of the policy you purchase?
    - Thoroughly familiar
    - Somewhat familiar
    - Not familiar

13. Are you familiar with the IRDA regulations in the insurance sector?
    - Thoroughly familiar
    - Somewhat familiar
    - Not familiar

Situation: In 2015, in order to boost consumer confidence in the insurance market, IRDA came up with IRDA (The Insurance Law Amendment) Act. The following questions purports to understand your perception before and after this amendment regarding your confidence in investing in the insurance sector.

Section C: BEFORE: Perception on IRDA Guidelines and Consumer Confidence Then

14. Please mention whether the following factors contributed in changing your confidence on the life insurance products you buy on a 3-point Likert scale with the degrees like (YES) = 3, (NO) = 2, (NOT SURE) = 1. Please tick in the appropriate box on your opinion.

Factors			
	Yes (3)	No (2)	Not Sure (1)
There was aggressive selling online by the insurers.			
IRDA did not have clear outsourcing guidelines for insurers.			
Denial of claims for delayed filing.			
Any valid ID such as Passport, Voter ID, and Driving License needed to be linked with the insurance policy.			
No free look period for cancellation.			

15. Was there any change in your confidence in the pre-2017 IRDA Act amendment (e.g. Protection of Policyholders' Interests), in the following factors? Please mention on a 3-point Likert scale with (YES) = 3, (NO) = 2, (NOT SURE) = 1. Please tick in the appropriate box on your opinion.

Factors			
	Yes	No	Not Sure
Was there an increase in your investment cap in life insurance policies compared to other investment options?			
Did you invest directly through online purchase of policies instead of going to life insurance agents?			

Section D: AFTER: Perception on IRDA Guidelines and Consumer Confidence Now

16. Please mention whether the following factors contribute in changing your confidence on the life insurance products you buy on a 3-point Likert scale with the degrees like (YES) = 3, (NO) = 2, (NOT SURE) = 1. Please tick in the appropriate box on your opinion.

Factors			
	Yes	No	Not Sure
IRDA laid out rules to curb aggressive online selling.			
IRDA laid out new outsourcing guidelines for insurers.			
IRDA stated that claims cannot be denied on the ground of delayed filing.			
IRDA made it mandatory to link PAN and Aadhaar to the insurance policy.			
IRDA allowed the policyholders for free look period of 15 days from the receipt of policy document and 30 days in case of policies obtained through distance/electronic mode to review the T&C of the policies and cancel in case they do not agree with those T&Cs.			

17. Is there any change in your confidence in the post-2017 IRDA Act amendment (e.g. Protection of Policyholders' Interests), in the following factors? Please mention on a 3-point Likert scale with (YES) = 3, (NO) =2, (NOT SURE) =1. Please tick in the appropriate box on your opinion.

Factors			
	Yes	No	Not Sure
Is there an increase in your investment cap in life insurance policies compared to other investment options?			
Do you invest directly through online purchase of policies instead of going to life insurance agents?			

Any other comments?

Thank you for participating!

References

Bharadi, H. H. (2012). Role of insurance regulatory and development authority in Indian insurance sector. *Shiv Shakti International Journal in Multidisciplinary and Academic Research (SSIJMAR)*, *1*(4), 1–15.

Bhattacharjee, A. (2012). *Social science research: Principles, methods, and practices*. Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.

Carrie, H., Lakishyk, K., Radas, S., & Nakamoto, K. (2004). *The effect of in-store free samples on short-and long-term purchasing behavior*. Charlottesville.

CUTS. (2016). *Regulatory impact assessment in insurance sector in India*. https://cuts-ccier.org/pdf/Regulatory_Impact_Assessment_in_Insurance_Sector_in_India_Facilitating_Investments_and_Enabling_Access.pdf

Eikenhout, L. (2015). *Risk management and performance in insurance companies*. University of Twente.

Eling, M., & Pankoke, D. A. (2016). Systemic risk in the insurance sector: A review and directions for future research. *Risk Management and Insurance Review*, *19*(2), 249–284.

Insurance Regulatory and Development Authority of India. (2007). *Insurance Regulatory and Development Authority of India Act, 1999*. https://www.policyholder.gov.in/uploads/CEDocuments/IRDA%20Brochure.pdf

IRDA. (2015). *Guidelines on appointment of insurance agents, 2015*. IRDA.

IRDAI. (2017). *Protection of policyholders' interests regulations, 2017*. The Gazette of India.

IRDAI. (2018). *IRDAI Annual report 2016–17*. https://doi.org/http://www.indianemployees.com/uploads/documents/042015/1428239209-16-92.pdf

Iyer, A. (2014, February 21). *IRDA's new guidelines ensure that 'Customer is king'*. MoneyControl.

RDA. (2019). *Updated List of Life Insurers*.

Kaboub, F. (2008). Positivist paradigm. *Encyclopedia of Counselling*, *2*(Pat 2), 786–787.

Kathirvel, N., & Radhamani, S. (2013). Satisfaction level of policyholders in LIS service with reference to Tirupur district, Tamil Nadu. *IOSR Journal of Business and Management*, *9*(6), 2319–7668.

Mittal, A. (2013). E-commerce: It's impact on consumer behavior. *Global Journal of Management and Business Studies*, *3*(2), 131–138.

Morgan, D. L. (2014). Pragmatism as a paradigm for social research. *Qualitative Inquiry*, *20*(8), 1045–1053.

Nebo, G. N., & Victor O. (2016). Effects of the strategies for customer satisfaction on the performance of insurance firms in enugu metropolis. *IOSR Journal of Business and Management*, Ver. I, *18*(5), 78–90.

Pandey, A. (2017, May). Recent insurance law reforms in India. *iPleaders*. https://blog.ipleaders.in/insurance-laws-reform/

Pandey, K., & Rao, T. J. (2013). Risk management in general insurance business in India. *The IUP Journal of Financial Risk Management*, *10*, 62–82.

Rewadikar, B. S. S. (2013). The impact of insurance regulatory and development authority [IRDA]. *International Journal of Advance Research in Computer Science and Management Studies*, *1*(3), 26–30.

Siddiqui, M. H., & Sharma, T. G. (2010). Analyzing customer satisfaction with service quality in life insurance services. *Journal of Targeting, Measurement and Analysis for Marketing*, *18*(3–4), 221–238.

Singh, B. K. (2010). An empirical study on perception of consumer in insurance sector. *Indian Journal of Economics & Business*, *9*(1), 61–73.

Yadav, R. K., & Mohania, S. (2014). Role of insurance ombudsman and grievance management in life insurance services in Indian perspective. *International Letters of Social and Humanistic Sciences*, *20*, 9–13.

KeyGames: A Game Theoretic Approach to Automatic Keyphrase Extraction

Arnav Saxena[*], **Mudit Mangal**[*] and **Goonjan Jain**
Department of Applied Mathematics, Delhi Technological University
New Delhi, India
{arnavsaxenadtu,muditmangaldtu}@gmail.com, goonjan_jain@hotmail.com

Abstract

In this paper, we introduce two advancements in the automatic keyphrase extraction (AKE) space - KeyGames and pke+. KeyGames is an unsupervised AKE framework that employs the concept of evolutionary game theory and consistent labelling problem to ensure consistent classification of candidates into keyphrase and non-keyphrase. Pke+ is a python based pipeline built on top of the existing pke library to standardize various AKE steps, namely candidate extraction and evaluation, to ensure truly systematic and comparable performance analysis of AKE models. In the experiments section, we compare the performance of KeyGames across three publicly available datasets (Inspec 2001, SemEval 2010, DUC 2001) against the results quoted by the existing state-of-the-art models as well as their performance when reproduced using pke+. The results show that KeyGames outperforms most of the state-of-the-art systems while generalizing better on input documents with different domains and length. Further, pke+'s pre-processing brings out improvement in several other system's quoted performance as well.

1 Introduction

Automatic Keyphrase Extraction (AKE) is a task of identifying important words and phrases that best describe a given text document. AKE makes searching and indexing large digital collections of text feasible and finds use in a variety of Natural Language Processing (NLP) and Information Retrieval (IR) tasks ((Zhang et al., 2004), (Hulth and Megyesi, 2006), (Berend, 2011)). Owing to its widespread use, Keyphrase Extraction has emerged as a fundamental NLP task, improvements in which could lead to improvements in higher-level applications that build upon it (Danesh et al., 2015). Due to this importance, many approaches to keyphrase extraction have been proposed in the literature, majorly along two research lines: supervised and unsupervised (Hasan and Ng, 2014). However, the field faces two specific problems. Firstly, with state-of-the-art performance on keyphrase extraction being much lower than on the many core NLP tasks, there is still a fair amount of improvement possible in this space (Papagiannopoulou and Tsoumakas, 2020). Secondly, there is no effective way to compare and analyze various past AKE systems since most of them use different experimental setups. Different works have used different pre-processing and evaluation to demonstrate their performance. These discrepancies in the past studies make assessing the progress of the AKE space very challenging.

To deal with the first issue, we recognised the need to approach this challenging problem from a fresh perspective. We thus propose a game-theoretic framework for AKE - KeyGames. Players, strategies, and payoffs form an integral part of any game. In KeyGames, the candidate keyphrases act as players having two strategies each: being a keyphrase and being a non-keyphrase. Each phrase plays a game with all the other phrases before deciding upon its strategy. Next, the definition of keyphrases is paraphrased, such that it paves the way for us to develop heuristics which are then formulated into mathematical payoffs. As per our definition, a candidate keyphrase becomes a keyphrase if it satisfies the following two conditions: (a) it should be related to the main themes of the document (b) it should be related to other important phrases

[*] Equal Contribution

This work is licensed under a Creative Commons Attribution 4.0 International License. License details: http://creativecommons.org/licenses/by/4.0/.

in the document. To capture (a) in our payoffs, we use semantic relatedness between the phrases and the theme of the document (we call this confidence - higher this metric, more confident the candidate is of becoming a keyphrase). Later, we also bias a candidate's confidence based on its positional information and frequency of occurrence. To capture (b), we use semantic relatedness between the different phrases. The construction of these payoffs is discussed in detail in section 5.4. Once we have the payoffs in place, we apply evolutionary game algorithm to score and rank keyphrases.

Inspired by Tripodi and Pelillo (2017), our reasons for exploring Game theory were two fold; firstly, game theory and NLP have a intuitionistic parallel wherein both have entities interacting with each other thereby influencing each other's behaviour. Second and more important was the ability of game-theoretic frameworks to perform consistent labelling of entities under contextual constraints. We discuss this property in detail in section 3 and further demonstrate consistent labelling in action when we incorporate a preliminary constraint in KeyGames's payoffs to tackle overgeneration[1] in extracted keyphrases (KeyGames+, section 5.4). Through this effort, we reduce the overgeneration error by ~10% points across datasets.

To deal with the second issue, we built pke+[2] as an extension to pke library [3] (Boudin, 2016), Pke is a python based tool kit containing implementation of various existing state-of-the-art AKE systems. Under pke+, we have provided a pipeline with standardized modules for candidate extraction and evaluation. All of pke's implemented systems can be plugged-in to this pipeline to extract keyphrases under similar experimental setup, allowing for a fair comparision (we have implemented KeyGames simply as another system under pke's setting).

To the best of our knowledge, using a game-theoretic approach for AKE has not been proposed before and our work is first such attempt. In fact, the application of game theory to NLP is still in its infancy. We believe that our work would spur additional research in this regard.

2 Game Theory

Game Theory is a branch of applied mathematics that provides predictive power in interactive decision-making situations (Tripodi and Pelillo, 2017). A typical game consists of a finite set of players, $P = (P_1, P_2, .., P_n)$, a set of pure strategies for each player i, $\beta_i = (\beta_1, ..., \beta_n)$ and a utility or payoff function $U_i : \beta_1 \times ..\beta_n \rightarrow R$ which associates a combination of strategies played by various players to the payoffs they receive. Each player aims to settle on a strategy which maximizes its payoff. The concept of payoff, in general, refers to the satisfaction that a player derives from the outcome of a game while interacting with other players. Furthermore, in a two player game, payoffs of a player i when it plays the game with player j can be represented in the payoff matrix A_{ij}, wherein values in each cell of the matrix indicates the payoff player i gets when opting the strategy corresponding to the row represented by that cell, in the event when the j^{th} player chooses the strategy corresponding to the column represented by that cell. Nash equilibria are those strategy profiles which are best response to the strategy of the co-player and no player has the incentive to unilaterally deviate from his strategy, because there is no way to do better.

In Evolutionary Game Theory (EGT), instead of choosing a strategy with certainty as in classical game theory, players choose strategies with certain probability. We thus use the concept of mixed strategy space for each player in EGT. Mixed strategy space is defined as a probability distribution over the pure strategies a player is allowed to play. It is represented as a vector $x_i = (x_i^1, ..., x_i^m)$ where m is the number of pure strategies and each component x_i^h denotes the probability that player i chooses its h^{th} pure strategy. Note that:

$$\sum_{h=1}^{m} x_i^h = 1 \tag{1}$$

[1]Overgeneration errors occur when a system correctly predicts a candidate as a keyphrase because it contains a word that appears frequently in the associated document, but at the same time erroneously outputs other candidates as keyphrases because they contain the same word

[2]https://github.com/mangalm96/KeyGames-pke

[3]https://github.com/boudinfl/pke

An important aspect we observe within the framework of EGT is the inductive learning process wherein the players play the game repeatedly till equilibrium is achieved and keep on updating their beliefs on the state of the game and choose their strategy accordingly (Tripodi and Pelillo, 2017). The steps below are followed to model this learning process:

Step 1: The expected payoff of every pure strategy each player i can obtain (u_i^1) in a single two player game is obtained using the equation:

$$u_i^1 = A_{ij}x_j \tag{2}$$

Step 2: The expected payoff obtained by each player i in a single game (u_i^2) is calculated using:

$$u_i^2 = x_i^T A_{ij} x_j = x_i^T u_i^1 \tag{3}$$

Step 3: Discrete time version of the replicator dynamic equation (Erdem and Pelillo, 2012) is used to update the mixed strategy space of a player i after the end of each iteration as follows:

$$x_i^h(t+1) = x_i^h(t) * u_i^1/u_i^2 \tag{4}$$

where t denotes each t^{th} iteration of a set of two player games.

3 Motivation

An important factor that motivated us to pursue game theoretic models was their ability to solve Consistent Labelling Problem wherein labelling of objects is done while satisfying certain constraints as per the context of the problem (Haralick and Shapiro, 1979). We observed that AKE can be modeled as a Consistent Labelling Problem as well wherein a candidate keyphrase is to be labelled either a keyphrase or a non-keyphrase abiding to certain constraints majorly based on the definition of a keyphrase. This distinguishes our approach from the existing AKE algorithms. A game-theoretic algorithm allows us not only to exploit the contextual information from an input document like various previous works do but also lets us design constraints and heuristics and ensures that classification of candidates into keys and non-keys remains consistent with these heuristics. In fact, it has been shown that, in some cases, using only contextual information without the imposition of constraints can lead to inconsistencies in the assignment of labels to related linguistic entities (Tripodi and Pelillo, 2017). A clear example of such a case is seen while extracting keyphrases from short texts where we have fewer candidates than the number of keyphrases to be extracted. Any other scoring algorithm would classify all the candidates as keys. KeyGames, however, will classify only those candidates into keys that remain consistent with the definition of a keyphrase and will classify the rest as non-keys.

A Consistent Labelling Problem can be defined as tuple $\langle X, L, C\rangle$ where $X = \{X_1, X_2...X_n\}$ is a set of n objects, $L = \{l_1, l_2 \ldots l_m\}$ is a set of m labels and C is a real-valued, $n^2 * m^2$ matrix of compatibility coefficients also called compatibility model, $C = \{C_{ij}(\lambda, \mu) \mid \lambda, \mu \in L\}$. $C_{ij}(\lambda, \mu)$ is the measure of strength of compatibility of two hypothesis "X_i is labeled λ" and "X_j is labeled μ". C is used to impose constraints on labelling so that each label assignment is consistent. The problem is to find a label for each object such that the resulting set of object-label pairs is consistent with the constraints of the compatibility model (Haralick and Shapiro, 1979). A. Miller and Zucker (1991) observed that a consistent labelling problem is equivalent to a polymatrix game. They demonstrated that the objects are equivalent to players in game theory, the labels are strategies each player can opt for, the compatibility model is the payoff matrix designed for the game and a consistent labelling solution is the Nash Equilibrium. In our model, the candidate keyphrases serve as the objects (players), "keyphrase" and "non-keyphrase" serve as the only two labels (strategies) and finally, the mathematical signals formulated using definition of keyphrases serve as the compatibility model (payoff matrix). These signals simply imply that two candidates, one of which adheres to the definition of a keyphrase and another one which doesn't, cannot be assigned the same labels. Clearly, the first one is to be labelled a keyphrase while the second one is to be labelled a non-keyphrase.

4 Related work

Existing studies present keyphrase extraction in two steps. Firstly, a list of candidate keyphrases is generated using heuristic rules such as extracting important noun phrases or extracting phrases that match specific POS (part-of-speech) tags (Bougouin et al., 2013; Mahata et al., 2018). In the next step, ranking mechanisms are often used to select the keyphrases from among the candidates. Both supervised and unsupervised methods are used here. Detailed surveys on these are available (Hasan and Ng, 2014; Papagiannopoulou and Tsoumakas, 2020). Here, we focus on unsupervised approaches due to their many benefits (no training data, domain independence). In unsupervised approaches, many techniques have been explored including, graph-based ranking, clustering and statistical methods. Many statistical based approaches like TFIDF, KPMiner(El-Beltagy and Rafea, 2009) have been proposed. YAKE (Campos et al., 2020) recently used a combination of statistical metrics to capture context information and the spread of the terms in the document.

In graph-based approaches, a document is represented as a graph and word-word interactions are scored using graph centrality measures such as PageRank (Brin and Page, 1998) and its variants. First introduced as TextRank (Mihalcea and Tarau, 2004), words are here represented by nodes and two words are connected by an edge if they co-occur in given window size. Nodes are then ranked using PageRank. Wan and Xiao (2008) extended TextRank to SingleRank by incorporating weights to edges, using number of co-occurences between words as weights. Another graph based model, PositionRank (Florescu and Caragea, 2017) demonstrated the importance of capturing the positions and frequency of a word in a biased PageRank algorithm, preferring words that appeared earlier and more frequently in the text. There have also been attempts at extracting keyphrases that cover all the main topics of the document. This has been done either by using a clustering based approach (Liu et al., 2009) to cluster words into specific topics or by using Latent Dirichlet Allocation(LDA) model (Blei et al., 2003) to obtain the topic distribution (Liu et al., 2010; Sterckx et al., 2015). The LDA based methods require training data, making the systems corpus dependent. Clustering based approaches were first used in a graph by Bougouin et al. (2013). In their proposed system, TopicRank, candidates are clustered depending on the percentage of shared words using hierarchical clustering. A graph is constructed where each node represents a cluster and edges are weighted based on the phrases' offset position in the text. This work was extended to MultiPartiteRank (Boudin, 2018) by encoding topical and positional information within a multipartite graph structure.

With recent advances in deep learning, word embeddings are increasingly being used to study word associations in NLP. Wang et al. (2015) first proposed usage of word embedding vectors for keyphrase extraction when they used the semantic information supplied by pretrained word embeddings to formulate a weighted scheme which is then used on a weighted Page Rank algorithm. Later using a similar approach, Mahata et al. (2018) proposed a corpus dependent method making use of their domain specific trained fasttext embeddings(Bojanowski et al., 2017) for extracting keyphrases from scientific articles. Recently in a departure from graph-based approaches, Bennani-Smires et al. (2018) proposed EmbedRank where they use sentence embeddings (Sent2vec) to represent both the candidate phrases and the document in the same high-dimensional vector space. Their system then ranks the candidate phrases using cosine similarity (semantic relatedness) between the embedding of the candidate phrase and the document embedding.

KeyGames combines many of the approaches discussed above. Here we develop payoffs by capturing word-word and topic-word interactions using fasttext embeddings and later bias these interactions using frequency and positional information of a candidate.

5 Proposed Methodology

5.1 Candidate Extraction

In this work, we use Spacy's[4] noun phrases chunking parser to extract the noun phrases from the input document. Various pre-processing steps (comprising of noun chunk cleaning) are then employed on these chunks to get a list of n-grams (including unigrams). Finally, Spacy's POS tagger is used to remove those singular tokens, which are not Proper Noun or unknown entities (X). Figure 1 illustrates this entire

[4]https://spacy.io

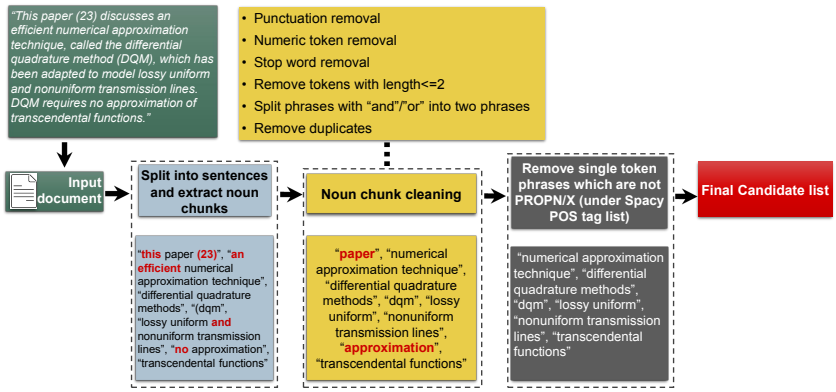


Figure 1: Candidate Extraction Methodology

candidate extraction methodology in detail with an example. This methodology forms a part of the candidate extraction module of the pke+ pipeline.

5.2 Semantic Relatedness using Embeddings

As mentioned in Section 1, we use semantic relatedness in our payoffs. To capture this relatedness, we use the concept of word embeddings. In particular, we use the pre-trained fastText embeddings to get a vector representation of our candidate keyphrases. Semantic relatedness between the candidates is then calculated using the cosine similarity between the phrase vectors. The chosen pre-trained embeddings[5] contain 1 million word vectors trained with subword information on Wikipedia 2017, UMBC web corpus, and statmt.org news dataset. The semantic relatedness between two phrases P_i and P_j is represented as S_{ij}. We choose fastText for its ability to capture both semantic and morphological similarities between words and also the similarity for Out of Vocabulary (OOV) words (Bojanowski et al., 2017).

5.3 Confidence Score

As per our definition, a keyphrase that is more related to the theme of the document is more confident to become a keyphrase. For this preliminary work, we assume that major themes of a document can be inferred from the document title. Thus we simply extract noun phrases from the input document's title (T = $\{t_1, t_2 \ldots t_m\}$), each of which acts as a thematic phrase for our algorithm. The thematic confidence score of the candidates is then calculated using the following equation:

$$Ct_i = S_{i1} + S_{i2} \ldots + S_{im} = \sum_{k=1}^{m} S_{ik} \tag{5}$$

where each ($S_{i1}, S_{i2}, ...S_{im}$) is the semantic relatedness between each keyphrase P_i and all the thematic phrases $\{t_1, t_2 \ldots t_m\}$. These individual scores (Ct_i) are then normalised after division with the highest confidence value (Ct_{max}).

Taking inspiration from previous works (Florescu and Caragea, 2017), we bias the confidence score by assigning more weight to candidate phrases that are found early in a document and are frequent. We weigh each phrase with its inverse position in the candidate list. The frequency of a phrase is taken into account by summing over all its position weights. For eg. for a phrase in the 2nd, 6th and 10th position, the positional confidence is $1/2 + 1/5 + 1/10 = 0.8$. The normalised positional confidence Cp_i, then added to our thematic confidence Ct_i gives us the final confidence score C_i for a candidate phrase P_i.

5.4 Evolutionary Game Algorithm

The candidate keyphrases participate as players ($P_i \in$ P) in our games, having two strategies each - to be a keyphrase (K) or to be a non-keyphrase (NK). Initially, all the players would have equal probabilities (0.5 each) of being a keyphrase or non-keyphrase and hence the initial mixed strategy space(x_i) for each

[5]https://fasttext.cc/docs/en/english-vectors.html

player P_i, would be

$$x_i = [0.5, 0.5]^T \tag{6}$$

As the games proceed, the probabilities in the mixed strategy space adjust appropriately following the inductive learning process after the replicator equation (Equation 4) is applied at the end of each iteration. Also, the two probabilities in the mixed strategy space of every player will always add to 1 (Equation 1). Whenever for any candidate keyphrase P_i, the probability of being a keyphrase is greater than 0.5, we say that the Phrase P_i is tending to be a keyphrase (K). Conversely, whenever the probability is lower than 0.5 (and hence the probability of being a non-keyphrase is higher than 0.5), we say that the Phrase P_i is tending to be a non-keyphrase (NK). Next, we describe the formulation of payoffs that players will receive while deciding upon their strategy (key vs. non key). Note that they will ultimately settle for a strategy for which their average payoff is maximum.

Payoffs 1.0: To construct the payoffs, we first need to understand how the candidates would select their strategies as they interact with one another. Phrase P_i while playing the game with Phrase P_j takes a decision based on the following logics:

1. If P_j is tending to be a keyphrase and also possesses a very high confidence value (C_j), P_i presumes that P_j has a very strong chance to become a keyphrase. Thus, P_i selects its strategy based on its semantic relatedness with P_j. If P_i and P_j have high semantic relatedness (S), then P_i chooses to be a keyphrase (K) as it believes it will have high contextual coherence with the input document just like the semantically similar and confident Phrase P_j. Otherwise, it tends to be a non-keyphrase.

2. If P_j is tending to be a non-keyphrase and has a very low confidence value ($\overline{C_j}$), P_i presumes that P_j will surely become a non-keyphrase and again selects its strategy based on its semantic relatedness with P_j. In this case, if P_i and P_j have high semantic relatedness (S) then P_i chooses to be a non-keyphrase as it gets reasons to believe that it will not be contextually coherent with the document just like the semantically similar and non-confident Phrase P_j. Otherwise, it tends to be a keyphrase.

3. If P_j is tending to be a keyphrase and has a very low confidence value ($\overline{C_j}$) or if P_j is tending to be a non-keyphrase with a very high confidence value (C_j), P_i realizes that P_j is not sure about its decision and hence bases its strategy on its own confidence value. If P_i has high confidence (C_i), then it chooses to be a keyphrase as high confidence indicates that the Phrase P_i is contextually coherent with the document and hence is capable of explaining the main topics of the document.

Payoffs 2.0: In a bid to bias our AKE system against producing over-generated keyphrases, we introduce a factor of fuzzy string match, "F" (based on the Levenshtein distance) on top of Payoff 1.0 built in the last discussion. The core idea is that if two candidates have similar spelling or similar set of tokens, then the system should deter from labelling both of them as keyphrases together. (only the candidate which performs well on other heuristics now stands a chance to become a keyphrase). We use fuzz.token set ratio() function from fuzzywuzzy[6], a python based library to calculate this factor F. A few examples of this function in action-
F(lossy transmission, transmission) = 100
F(pade approx, rational approx) = 84
F(numerical methods,transmission lines) = 34
To bring out this impact of fuzzy match, we tweak the first condition discussed under payoff 1.0 as follows while keeping the other two conditions unchanged:

1. If P_j is tending to be a keyphrase and also possesses a very high confidence value (C_j), P_i presumes that P_j has a very strong chance to become a keyphrase. Thus, P_i selects its strategy based on its semantic relatedness *and its fuzzy similarity* with P_j. If P_i and P_j have high semantic relatedness (S) *and low fuzzy similarity* ($\overline{F}$), then P_i chooses to be a keyphrase as it believes it will have high

[6] 1. https://github.com/seatgeek/fuzzywuzzy

Phrase P_i \ Phrase P_j	Strategies	KEY	NON KEY
	KEY	Payoff α_1 • P_j has high confidence and P_i is contextually similar to P_j(C_j * S) • P_j has low confidence and P_i has high confidence ($\overline{C_j}$* C_i)	Payoff α_2 • P_j has high confidence and P_i has high confidence (C_j * C_i) • P_j has low confidence and P_i is dissimilar to P_j ($\overline{C_j}$* $\overline{S}$)
	NON KEY	Payoff α_3 • P_j has high confidence and P_i is dissimilar to P_j (C_j*$\overline{S}$) • P_j has low confidence and P_i has low confidence ($\overline{C_j}$* $\overline{C_i}$)	Payoff α_4 • P_j has high confidence and P_i has low confidence (C_j * $\overline{C_i}$) • P_j has low confidence and P_i is similar to P_j ($\overline{C_j}$* S)

Table 1: Strategies of Phrase P_i summarised depending upon strategies of Phrase P_j.

Phrase P_i \ Phrase P_j	Strategies	Key	Non Key
	Key	C_jS +$\overline{C_j}C_i$	C_jC_i+$\overline{C_j}$ $\overline{S}$
	Non Key	$C_j\overline{S}$+ $\overline{C_j}$ $\overline{C_i}$	$C_j\overline{C_i}$ +$\overline{C_j}S$

Table 2: Payoff 1.0 Matrix. C and S denote the Confidence and Similarity indicators respectively.

Phrase P_i \ Phrase P_j	Strategies	Key	Non Key
	Key	$C_jS\overline{F}$ +$\overline{C_j}C_i$	C_jC_i+$\overline{C_j}$ $\overline{S}$
	Non Key	$C_j\overline{S}$+ $\overline{C_j}$ $\overline{C_i}$ + C_jSF	$C_j\overline{C_i}$ +$\overline{C_j}S$

Table 3: Payoff 2.0 Matrix. C,S and F denote the Confidence, Similarity and Fuzzy Match indicators respectively.

contextual coherence with the input document just like the semantically similar and confident Phrase P_j while the low F_{ij} ensures that it will not be causing over-generation. Otherwise, it tends to be a non-keyphrase.

As in any two-by-two game with players having two strategies each, four scenarios are possible when two phrases are involved in our games: (I) both phrases opt for becoming keyphrase (II) both phrases opt for becoming non-keyphrase (III) phrase one opts to become a keyphrase while phrase two doesn't (IV) phrase two opts to become a keyphrase while phrase one does not. For each of these four scenarios, we develop four mathematical signals (α_1, α_2, α_3, α_4) which ultimately become our payoffs. These four signals employ the use of indicators S, $\overline{S}$, C_i, C_j, $\overline{C_i}$, $\overline{C_j}$, F, $\overline{F}$ to model the three decision logics discussed above. We further demonstrate the formulation of Payoffs 1.0 in Table 1.The Payoff 1.0 matrix and Payoff 2.0 matrix are shown in Table 2 and Table 3 respectively.

Lastly, we implement evolutionary game algorithm (Algorithm 1) to score and rank keyphrases. KeyGames (KG) use Payoffs 1.0 and KeyGames+ (KG+) use Payoffs 2.0. Note that with regards to Algorithm 1, equilibrium is the state when the probability of becoming a keyphrase for all candidates remains constant across two consecutive iterations. We simulate this by letting the algorithm run for a large enough number of iterations(i.e. 50).

Algorithm 1 KeyGames Algorithm

Input: Candidates list ck; Confidence score C; Semantic Relatednes S
Init Mixed Strategy Space, x = [ ]
Init Weight of a candidate i, $W(i)$ = 0
for each phrase i in ck **do**
    $x(i)$ = [P(K)=0.5, P(NK)=0.5]^T
while (equilibrium not achieved) **do**
    for each phrase i in ck **do**
        u_1 = $[0, 0]^T$
        u_2 = 0
        for each phrase j in ck **do**
            payoffMatrix$[0, 0]$ = C_jS +$\overline{C_j}C_i$
            payoffMatrix$[0, 1]$ = C_jC_i + $\overline{C_j}$ $\overline{S}$
            payoffMatrix$[1, 0]$ = $C_j\overline{S}$+ $\overline{C_j}$ $\overline{C_i}$
            payoffMatrix$[1, 1]$ = $C_j\overline{C_i}$ + $\overline{C_j}S$
            $temp$ = payoffMatrix $*x(j)$
            $u_2 += x(i)^T * temp$
            $u_1 += temp$
        $x_{(t+1)}(i) = x_t(i) * u_1/u_2$
        $ck[i].W(i) += x_{(t+1)}(i)[0]$
sort(ck) by descending W and extract topn keys
Output: List of Topn keyphrases

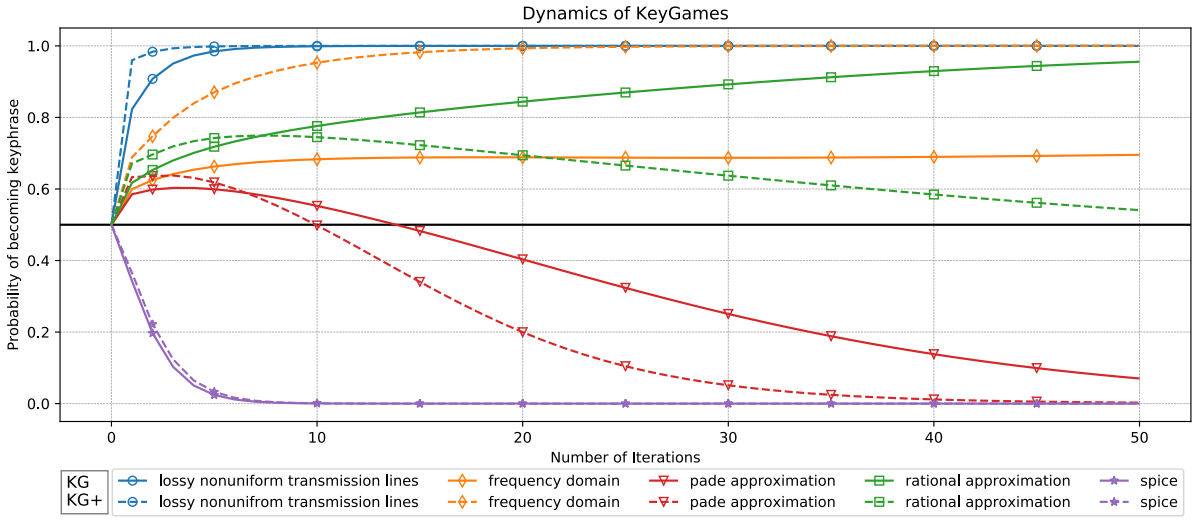


Figure 2: Using KG, 'lossy nonuniform transmission lines', 'rational approximation' and 'frequency domain', all end up in the extracted keyphrase list at the equilibrium position, but 'rational approximation' and 'frequency domain' don't make it to the top 10 keyphrase list as they achieve equilibrium at a very slow rate. Under KG+, however, overgenerated phrases like 'pade approximation' and 'rational approximation' both move towards becoming non-key at a faster pace then their KG counterparts. Also note that KG+, while punishing over-generated candidates, rewards non-overgenerated phrases like 'frequency domain' and it ends up as top 10 keyphrase (partially matching with a ground truth keyphrase) thereby increasing diversity in the KG+ extracted keyphrase list.

Title: Accurate modeling of **lossy nonuniform transmission lines** by using **differential quadrature methods**.
Input Text: This paper discusses an efficient **numerical approximation technique**, called the **differential quadrature method** (DQM), which has been adapted to model lossy uniform and nonuniform transmission lines... Using the DQM, the frequency-domain Telegrapher's partial differential equations for transmission lines... DQM reduces interconnects into **multiport models** whose port voltages and currents are related by rational formulas in the frequency domain. The rationalization process in DQM is comparable with the Pade approximation of asymptotic waveform evaluation (AWE) applied to transmission lines. AWE employs a complex moment-matching process to obtain rational approximation... Due to global sampling of points in the DQM approximation, it requires far fewer grid points in order to build accurate discrete models than other numerical methods do. The DQM-based time-domain model can be readily integrated in a circuit simulator like SPICE.
Assigned Keyphrases: lossy nonuniform transmission lines, differential quadrature method, numerical approximation technique, frequency-domain Telegrapher PDE, partial differential equations, multiport models, multiconductor transmission lines, rationalization process
KeyGames Keyphrases: lossy nonuniform transmission lines, differential quadrature methods, numerical approximation technique, transmission lines, multiport models, accurate discrete models, numerical methods, rational formulas
KeyGames+ Keyphrases: lossy nonuniform transmission lines, differential quadrature methods, numerical, approximation technique, multiport models, global sampling, small set, frequency domain, accurate discrete models

Table 4: An example of keyphrases extracted using KG and KG+ from a sample Inspec article (397.txt)

5.5 Keyphrase Ranking:

Keyphrases are ranked in the order in which they achieve equilibrium. We measure this rate for a candidate by summing over its probability of becoming a keyphrase at the end of each iteration (Algorithm 1: Line 18). The earlier a candidate becomes a keyphrase, the faster its probability of becoming a keyphrase (P(K)) reaches one. Hence, its summation of P(K) across all iterations would be higher. Table 4 shows ranked keyphrases extracted from a sample Inspec article. Figure 2 illustrates the dynamics of a typical KeyGame for selected phrases of the sample.

6 Experimentation and Results

6.1 Datasets and Evaluation Metrics

To evaluate the performance of KeyGames, we carry out experiments on three publicly available English datasets. The first dataset, Inspec (Hulth, 2003), consists of short documents from scientific journal

Method	Inspec				SemEval				DUC			
	P	R	F_1	$F_1{}^{o}$	P	R	F_1	$F_1{}^{o}$	P	R	F_1	$F_1{}^{o}$
SingleRank	40.31	36.41	38.26	29.69*	6.8	4.71	5.57	1.99*	22.48	27.47	24.73	27.2
TopicRank	41.02	34.69	37.59	27.9	16.1	11.16	13.18	12.1	19.31	23.6	21.24	21.39*
PositionRank	40.56	36.61	38.48	27.49*	10.6	7.35	8.68	7.85*	23.33	28.52	25.67	27.01*
MultipartiteRank	41.15	37.2	39.07	30.6	18.2	12.61	**14.9**	14.5	22.21	27.19	24.45	23.41*
EmbedRank	41.44	37.43	39.33	35.8*	6.2	4.29	5.08	3.58*	22.71	27.71	24.96	29.4*
YAKE	36.7	32.8	34.63	31.6	0.6	0.4	0.5	12.25*	3.95	4.84	4.35	14.18*
KeyGames(KG)	43.38	38.79	**40.48**†	-	17.57	12.13	14.35	-	25.71	31.43	**28.28**†	-
KeyGames+(KG+)	43.15	38.73	40.87	-	17.17	11.85	14.02	-	23.89	29.21	26.28	-

Table 5: Comparison of KG and KG+ with some unsupervised systems. Precision (P), Recall (R), and F-score (F_1) for Top 10 Keyphrases (using pke+) are reported. † indicates significance at the 0.01 level using Student's t-test over the next best system. For a fair comparison, the original reported F-scores ($F_1{}^{o}$) of the existing systems are provided. Missing results are accounted for by reimplementation using pke (indicated by *)

abstracts. Like previous works (Mihalcea and Tarau, 2004; Bougouin et al., 2013), we use the test set of 500 documents and the uncontrolled set of annotated keyphrases as ground truth for our analysis. The second dataset, SemEval 2010 (Kim et al., 2010) contains ACM full length papers. In our experiments, we use the 100 documents from the test set and the combined set of annotated keyphrases. The third dataset, DUC 2001 (Wan and Xiao, 2008), is a collection of 308 medium length news articles. Similar to candidate extraction, we use common evaluation to transparently compare our approach with other systems. We compute exact matched true positives between the assigned and extracted keys (both stemmed) and evaluate performances in terms of the micro-averaged Precision, Recall, and F_1 measure for top 10 keyphrases.

6.2 Performance Comparision

In Table 5, we compare KG (with Payoffs 1.0) and KG+ (with Payoffs 2.0) to existing unsupervised state-of-the-art systems - SingleRank (Wan and Xiao, 2008), TopicRank (Bougouin et al., 2013), PositionRank (Florescu and Caragea, 2017), MultipartiteRank (Boudin, 2018), EmbedRank[7] (Bennani-Smires et al., 2018) and YAKE[8] (Campos et al., 2020). Direct comparisons are done on the scores (P,R,F_1) obtained using the pke+ implementation. We also mention the original scores($F_1{}^{o}$) as reported in the respective papers to compare it with our pke+ implementations.

We observe that KeyGames consistently outperforms most of the existing systems across the three datasets, each with different document length, covering two different domains. Moreover, the pke+ pipeline (more specifically, the pre-processing) helps most systems improve/retain performances as originally quoted. This further indicates the effectiveness of the new candidate extraction. Among existing systems, EmbedRank performs admirably on short texts (Inspec/DUC) but fails to replicate it on long texts (as also reported by the authors themselves). Only MPRank gives consistent results across the datasets but does not match the effectiveness of KeyGames, especially on Inspec and DUC, where the improvements are statistically significant while the difference on SemEval is statistically insignificant. The performance on longer texts like SemEval has historically been lower, often attributed to the hypothesis that it is more difficult to extract keyphrases correctly from longer documents because of a much bigger search space (Hasan and Ng, 2014).

Compared to KG, the KG+ system, which incorporates the over-generation constraint, shows a dip in performance even though it results in keys with less overgeneration (Table 6) (and hence more diversity). We calculate this over-generation percentage as the the sum of the number of phrases having more than 75% fuzzy match with any other keyphrase in the extracted/assigned keyphrase list as

Dataset	Inspec	SemEval	DUC
Assigned Keys	29.59%	46.99%	22.71%
KG	35.49%	63.4%	30.68%
KG+	26.12%	53.5%	20.92%

Table 6: Over-generation in Assigned keys & Extracted Keys of KG and KG+

[7]EmbedRank's reported scores are macro-averaged. Hence, we mention their reimplemented scores

[8]YAKE's original reported score are based on the entire SemEval dataset (and not just the test set)

a share of the total number of keyphrases extracted/assigned across the entire dataset -

$$og\% = 100\,\frac{\sum_{i \in N} \text{count(overgenerated keys)}}{\sum_{i \in N} \text{count(total keys)}} \tag{7}$$

where *overgenerated keys* are keyphrases with more than *75% fuzzy match(F)* score with any other key in the assigned/extracted keyphrase list of doc i, *total keys* are the total number of assigned/extracted keyphrases of doc i and N is the number of documents in a dataset.

Bennani-Smires et al. (2018) also observed such a dip in results on these diverse and less overlapping keys in their experiments. It can perhaps be explained by high over-generation present in the assigned keys themselves (Table 6). In Table 5 we highlight the results of the KG system and not KG+ because of its more consistent performance across datasets.

7 Conclusion and Future Work

In this paper, we discussed our work on a novel, unsupervised game-theoretic approach to automatic keyphrase extraction after modeling the task as a consistent labelling problem. We paraphrased the definition of a keyphrase, conceptualized logics to model the same, and ultimately designed keyphrase games by quantifying these logics into mathematical payoffs. For carrying out experiments, we used pke to construct a new pipeline pke+, standardizing candidate extraction, and evaluation steps across different systems. Our proposed method consistently outperforms most of the existing systems on three datasets with different domains and lengths, indicating the potential of game theory as an alternative. In the future, we plan on working with Infection and Immunization Dynamics (InImDyn) (Rota Bulò et al., 2011), which approximately simulates replicator dynamics in linear time, thereby reducing time-complexity of KeyGames. We plan to experiment with different heuristics in our payoffs as well. We would also like to explore other methods like document embedding to extract the theme of the document.

References

Douglas A. Miller and Steven Zucker. 1991. Copositive-plus lemke algorithm solves polymatrix games. *Operations Research Letters*, 10:285–290, 07.

Kamil Bennani-Smires, Claudiu Musat, Andreea Hossmann, Michael Baeriswyl, and Martin Jaggi. 2018. Simple unsupervised keyphrase extraction using sentence embeddings. In *Proceedings of the 22nd Conference on Computational Natural Language Learning*, pages 221–229. Association for Computational Linguistics.

Gábor Berend. 2011. Opinion expression mining by exploiting keyphrase extraction. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 1162–1170. Asian Federation of Natural Language Processing.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Florian Boudin. 2016. pke: an open source python-based keyphrase extraction toolkit. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, pages 69–73, Osaka, Japan, December.

Florian Boudin. 2018. Unsupervised keyphrase extraction with multipartite graphs. 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 667–672. Association for Computational Linguistics.

Adrien Bougouin, Florian Boudin, and Béatrice Daille. 2013. Topicrank: Graph-based topic ranking for keyphrase extraction. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 543–551. Asian Federation of Natural Language Processing.

S. Brin and L. Page. 1998. The anatomy of a large-scale hypertextual web search engine. In *Seventh International World-Wide Web Conference (WWW 1998)*.

Ricardo Campos, Vítor Mangaravite, Arian Pasquali, Alípio Mário Jorge, Celia Nunes, and Adam Jatowt. 2020. Yake! keyword extraction from single documents using multiple local features. *Inf. Sci.*, 509:257–289.

Soheil Danesh, Tamara Sumner, and James H. Martin. 2015. Sgrank: Combining statistical and graphical methods to improve the state of the art in unsupervised keyphrase extraction. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, pages 117–126. Association for Computational Linguistics.

Samhaa R. El-Beltagy and Ahmed A. Rafea. 2009. Kp-miner: A keyphrase extraction system for english and arabic documents. *Inf. Syst.*

Aykut Erdem and Marcello Pelillo. 2012. Graph transduction as a noncooperative game. *Neural Computation*, 24(3):700–723.

Corina Florescu and Cornelia Caragea. 2017. PositionRank: An unsupervised approach to keyphrase extraction from scholarly documents. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vancouver, Canada. Association for Computational Linguistics.

R. M. Haralick and L. G. Shapiro. 1979. The consistent labeling problem: Part i. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):173–184.

Kazi Saidul Hasan and Vincent Ng. 2014. Automatic keyphrase extraction: A survey of the state of the art. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1262–1273. Association for Computational Linguistics.

Anette Hulth and Beáta B. Megyesi. 2006. A study on automatically extracted keywords in text categorization. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 537–544. Association for Computational Linguistics.

Anette Hulth. 2003. Improved automatic keyword extraction given more linguistic knowledge. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, EMNLP '03, pages 216–223, Stroudsburg, PA, USA. Association for Computational Linguistics.

Su Nam Kim, Olena Medelyan, Min-Yen Kan, and Timothy Baldwin. 2010. Semeval-2010 task 5 : Automatic keyphrase extraction from scientific articles. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 21–26. Association for Computational Linguistics.

Zhiyuan Liu, Peng Li, Yabin Zheng, and Maosong Sun. 2009. Clustering to find exemplar terms for keyphrase extraction. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 257–266. Association for Computational Linguistics.

Zhiyuan Liu, Wenyi Huang, Yabin Zheng, and Maosong Sun. 2010. Automatic keyphrase extraction via topic decomposition. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 366–376. Association for Computational Linguistics.

Debanjan Mahata, John Kuriakose, Rajiv Ratn Shah, and Roger Zimmermann. 2018. Key2vec: Automatic ranked keyphrase extraction from scientific articles using phrase embeddings. 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 634–639. Association for Computational Linguistics.

Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*.

Eirini Papagiannopoulou and Grigorios Tsoumakas. 2020. A review of keyphrase extraction. *WIREs Data Mining and Knowledge Discovery*, 10(2):e1339.

Samuel Rota Bulò, Marcello Pelillo, and Immanuel Bomze. 2011. Graph-based quadratic optimization: A fast evolutionary approach. *Computer Vision and Image Understanding*.

Lucas Sterckx, Thomas Demeester, Johannes Deleu, and Chris Develder. 2015. Topical word importance for fast keyphrase extraction. In *Proceedings of the 24th International Conference on World Wide Web*, page 121–122. Association for Computing Machinery.

Rocco Tripodi and Marcello Pelillo. 2017. A game-theoretic approach to word sense disambiguation. *Computational Linguistics*, 43(1):31–70.

Xiaojun Wan and Jianguo Xiao. 2008. Single document keyphrase extraction using neighborhood knowledge. In *Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 2*, AAAI'08, pages 855–860. AAAI Press.

Rui Wang, Wei Liu, and Chris McDonald. 2015. Using word embeddings to enhance keyword identification for scientific publications. In *Australasian Database Conference*, pages 257–268. Springer.

Yongzheng Zhang, Nur Zincir-Heywood, and Evangelos Milios. 2004. World wide web site summarization. *Web Intelli. and Agent Sys.*, 2(1):39–53, January.

Journal of Tropical Pediatrics, 2020, 00, 1–7
doi: 10.1093/tropej/fmaa110
Original Paper

OXFORD

Leave against Medical Advice in Children: Rural Indian Perspective

Debadyuti Datta MD,[1] Tanushree Mondal MD,[2] Mihir Sarkar MD,[3] Soumita Chel BTec, MSc,[4] Sankar Kumar Das MD,[5] Arnab Nandy MD,[1] Divyoshanu M Ivan BTec Trainee,[6] *and* Rakesh Mondal MD[5]

[1]Department of Pediatrics, North Bengal Medical College, Darjeeling, West Bengal, India
[2]Community Medicine, Medical College Kolkata, Kolkata, West Bengal, India
[3]Pediatric Medicine, Medical College Kolkata, Kolkata, West Bengal, India
[4]Alumni University of Glasgow, Expert Data Science, Kolkata, West Bengal, India
[5]Department of Pediatric Medicine, North Bengal Medical College, Darjeeling, West Bengal, India
[6]Department of Biotechnology, Delhi Technological University, Delhi, India
Correspondence: Dr Rakesh Mondal, Department of Pediatrics, Medical College Kolkata, Kolkata, West Bengal, India.
E-mail: <ivanrakesh2001@gmail.com>
Work should be attributed to: Rheumatology Unit; Department of Pediatric Medicine, North Bengal Medical College, West Bengal, India.

ABSTRACT

Objectives: To determine the burden and etiological factors of leave against medical advice (LAMA) in Indian children.
Methodology: In this prospective study, legal guardians of 528 patients who took the decision of LAMA were interviewed (using structured question answers based multi-option) and data were captured over a period of 16 months. The resultant database was analyzed using standard statistical methods.
Results: About 6.12% of childhood LAMA cases were dealt out of total pediatric admission including newborns. Neonatal preponderance noted in cases of LAMA from intensive care unit (ICU; 57.14%, $p < 0.05$). The overall male ($n = 293$) to female ($n = 235$) ratio in this study was 1.25:1. Number of LAMA patients was higher from rural area (83.33%), mostly admitted in ICU (93.65%, Pearson's chi-squared test, $p < 0.05$). Higher proportion (29.47%) of patients with infection availed LAMA from neonatal age group but overall LAMA patients fall under category of respiratory disorders (22.35%). Interest of the domestics issues other than suffering child was considered primary during LAMA for those admitted in ward as compared with ICU patients [odds ratio (OR): 1.73, CI: 1.02–2.94, $p < 0.05$]. ICU patients were reportedly to be taken to private health care facility (OR: 1.93, CI: 1.06–3.49, $p < 0.05$). Duration of stay before taking LAMA from ward was <7 days (85.59%, OR: 0.19, CI: 0.11–0.35, $p < 0.05$). Upper-lower socio-economic class attributed financial constraint as the main reason for LAMA (Pearson's chi-squared test, Chi-square value: 152.23, $p < 0.05$).
Conclusions: This study tried to elucidate the determinants of childhood LAMA in rural Indian setting.

KEYWORDS: children, LAMA, risk factors, treatment refusal

© The Author(s) [2020]. Published by Oxford University Press. All rights reserved.
For permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/tropej/advance-article/doi/10.1093/tropej/fmaa110/6043117 by Delhi Technological University user on 05 January 2021

INTRODUCTION

Sometimes it is seen that patients leave the hospital without completion of treatment before physician in charge decides to discharge them; this act of voluntarily leaving the hospital is widely known as leave against medical advice (LAMA) or discharge on risk bond (DORB) [1]. Especially, discharge of hospitalized children against medical advice is a serious health concern. At the time of LAMA or DORB, children often remain sick because of interruption of ongoing treatment. This kind of health-seeking behavior exposes partially treated children to high risks of mortality and long term sequel. Parental disagreement with health care delivery system for their children also reflects negligence of one's child health. Besides, treating physician faces legal and ethical challenges from such discharges [2, 3]. The act of LAMA hinders the attainment of Sustainable Developmental Goal number 3: Good health and well-being [4]. There are few studies related to this situation, which have been conducted in pediatric age group (mainly African countries), but such studies we believe are unable to depict the whole extent and true nature of the problem [5]. These studies showed the reasons for LAMA vary depending on socio-economical, cultural and environmental factors. In India, by far only one attempt had been made to document this problem in pediatric age group. Considering the social and cultural aspects in India, where government is investing huge amount of money to reduce neonatal mortality rate, infant mortality rate, under 5 mortality rate through various health schemes and programs [6], the magnitude and epidemiological factors related to LAMA need to be explored in detail.

MATERIALS AND METHODS

We conducted a prospective observational study from December 2018 to March 2020, in the pediatric department of a district-level tertiary care teaching hospital in West Bengal, India. Patients aged from day 0 to 12 years, who took LAMA, were included after taking written informed consent from either parent or legal guardian. The study protocol was approved beforehand by the Institutional Ethics Committee for clinical research. Although we sought to include consecutive patients when consent was available, our sampling strategy was purposive since we excluded patients not accompanied by close family caregivers and those whose caregivers were not consenting for interview.

The process of LAMA initiates as soon as caregivers communicate the decision of going LAMA. Primary consultant had a discussion with caregivers regarding various queries and counseled adequately of medical consequences before their decision. The process of LAMA is similar to discharge on advice. However, a written declaration from caregiver is recorded followed by issuance of 'discharge/LAMA' certificate. Such discharge certificates never prejudice public health and safety [7].

Data were collected in structured format age, gender, residence(rural/urban), primary diagnosis and condition of patient at the time of LAMA, hospital area where the child was admitted (emergency/ward/intensive care unit [ICU]), duration of hospitalization before LAMA, person who initiated discussion on LAMA, socio-demographic information as defined by modified Kuppuswamy scale [8]. These parameters were compared in patients admitted in ward and patients admitted in ICU.

To understand the rationale behind the LAMA opted by family members, a structured closed ended (Likert-type scales) questionnaire was formulated by an expert committee considering the dimensionality of the issue, which included items on economic constraints, social support, type of illness (terminal, critical), satisfaction level to treatment of child's illness, behavior of health care staff, dissatisfaction to hospital environment (cleanliness, food, hygiene and beds), iatrogenic complication, seeking alternative therapy, personal or family reason.

The questionnaire made simple and in understandable language to the respondents. After initial framing it was reviewed and revised by expert committee. Pilot testing on 30 respondents was carried out to validate the questionnaire. If there were multiple reasons, we accepted all the responses. In case of incomplete or conflict issues, data were excluded. The number of admissions and discharges in the

Downloaded from https://academic.oup.com/tropej/advance-article/doi/10.1093/tropej/fmaa110/6043117 by Delhi Technological University user on 05 January 2021

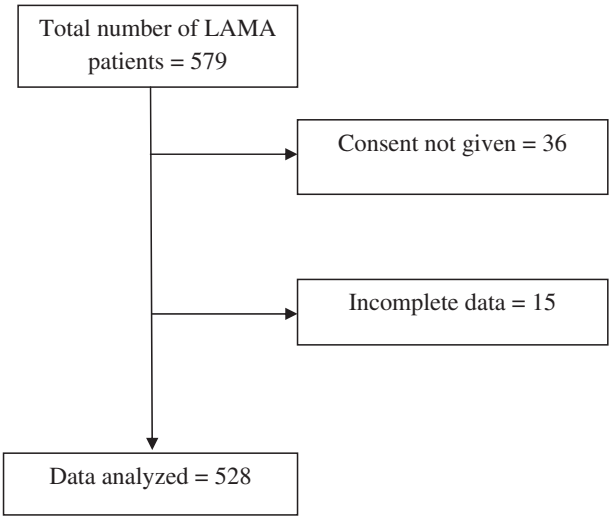


Fig. 1. Flow diagram showing the study participants.

pediatric department during the study period were collected from hospital record-book.

Postgraduate students involved in the study were trained to assist parents to complete the interview in understandable language. Five simulation based mock scenario was performed after explaining the questionnaire to the trainees. These sessions were conducted by the senior faculties. Only principal investigator collected the data in the indexed study. Regular follow-up and monitoring were conducted every week by senior faculties after the initiation of study.

Statistical Packages for the Social Sciences (SPSS) version 23 was used for statistical analysis of data of this study. Categorical data were presented as count and percentage. Qualitative data had been summarized as mean, median, standard deviation and range. Key proportions have been depicted with 95% confidence interval (CI). Categorical variables have been compared between subgroups by Pearson's chi-squared test. In all comparisons, statistical significance has been inferred if p value was < 0.05.

RESULTS

During study period total admission was 9456 (males 66.1%), out of which 579 patients (6.12%) took LAMA. Fifty one cases were not completely interviewed. Thus, data from 528 patients have been analyzed (Fig. 1).

Demographic characteristics of patients in different LAMA subgroups are depicted in Table 1. It is evident that of the 528 cases analyzed, 63 (11.93%) received treatment in ICU. 190 (36%) were newborns and 338 (64%) were pediatric patients. Mean age of LAMA patients from ICU vs. ward was 1.07 years ($\pm$ 2.24) vs. 2.46 years ($\pm$ 3.27). LAMA from ICU showed significant neonatal preponderance (57.14%, Pearson's chi-squared test, $p < 0.05$).The overall male ($n = 293$) to female ($n = 235$) ratio in this study was 1.25:1. No gender predilection has been noted. Number of LAMA patients was higher from rural area (83.33%), mostly admitted in ICU (93.65%, Pearson's chi-squared test, $p < 0.05$).

Age group wise, broad categories of diseases of all LAMA patients are shown in Fig. 2. Higher proportion (29.47%) of patients with infection availed LAMA from neonatal age group but overall LAMA patients fall under category of respiratory disorders (22.35%).

Table 2 presents a comparison of factors associated with LAMA between these two subgroups. In majority of cases primary decision maker for LAMA was father (72.73%). The decision of LAMA was taken in favor of the family for those admitted in ward as compared with ICU patients [odds ratio (OR): 1.73, CI: 1.02–2.94, $p < 0.05$). Caregivers said that most of the patients (81.44%) would be taken to home after availing LAMA; however, ICU patients were reportedly to be taken to private health care facility (OR: 1.93, CI: 1.06–3.49, $p < 0.05$). Mean duration of hospitalization before availing LAMA was 3.51 days ($\pm$ 4.90). In most of cases duration of stay before taking LAMA from ward was <7 days (85.59%, OR: 0.19, CI: 0.11–0.35, $p < 0.05$).

Reasons for LAMA among patients of different socio-economic status are depicted in Table 3. Approximately 56% of LAMA cases occurred in those of upper-lower socio-economic status.

Downloaded from https://academic.oup.com/tropej/advance-article/doi/10.1093/tropej/fmaa110/6043117 by Delhi Technological University user on 05 January 2021

Table 1: Basic demographic characteristics of LAMA patients

Parameter	Whole LAMA cohort ($n = 528$)	ICU LAMA ($n = 63$)	Ward LAMA ($n = 465$)	p value
Age group (%)				<0.05
Neonate	190 (36)	36 (57.14)	154 (33.11)	
Pediatric	338 (64)	27 (42.86)	311 (66.89)	
Gender(%)				0.795
Male	293 (55.49)	34 (53.96)	259 (55.69)	
Female	235 (44.51)	29 (46.04)	206 (44.31)	
Socio-economic status (%)				0.771
Upper middle	12 (2.27)	1 (1.58)	11 (2.37)	
Lower middle	212 (40.16)	23 (36.52)	189 (40.65)	
Upper lower	294 (55.68)	37 (58.73)	257 (55.27)	
Lower	10 (1.89)	2 (3.17)	8 (1.72)	
Place of living (%)				0.019
Rural	440 (83.33)	59 (93.65)	381 (81.94)	
Urban	88 (16.67)	4 (6.35)	84 (18.04)	

Notes: LAMA = leave against medical advice; ICU = intensive care unit. A p value in the last column is from Pearson's chi-squared test.

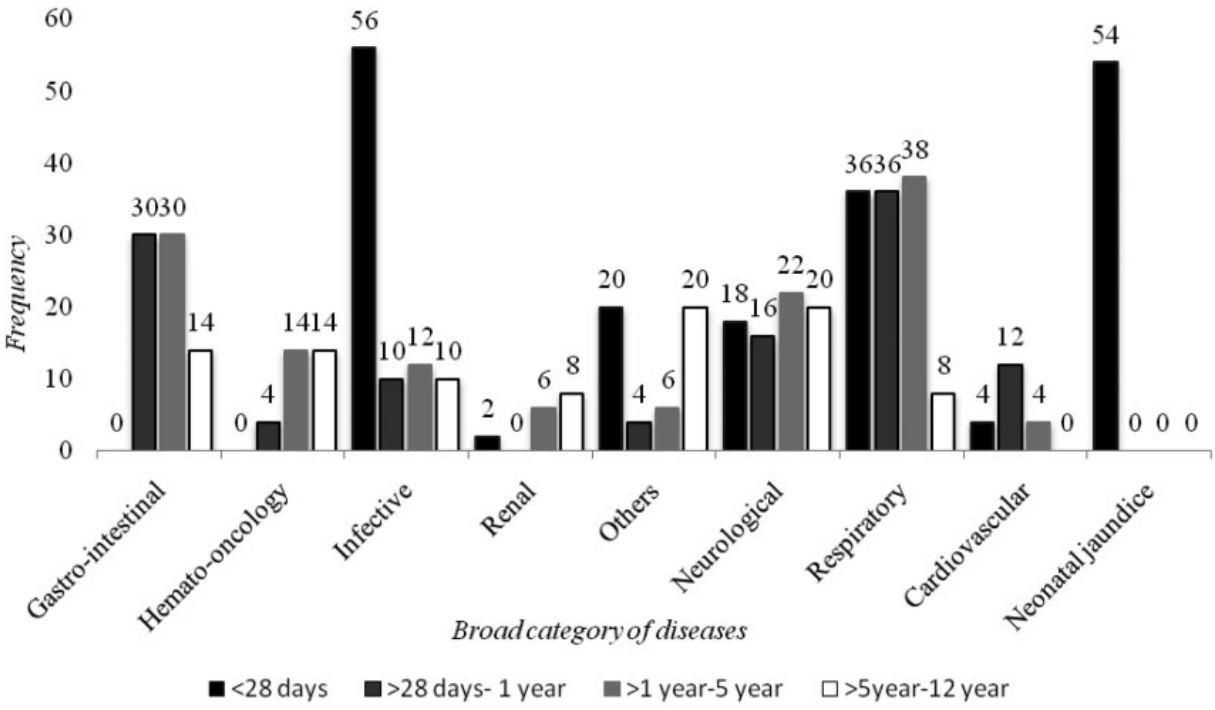


Fig. 2. Distribution of broad categories of diseases of all LAMA patients according to age.

Majority of upper-lower socio-economic class attributed financial constraint as the main reason for LAMA (Pearson's chi-squared test, Chi square value: 153.667, $p < 0.05$).

DISCUSSION

This study observed a prevalence (6.12%) of LAMA similar to studies conducted in Northern India [5], Southern Nigeria, Kuwait, Iran [9–11]; much less

Downloaded from https://academic.oup.com/tropej/advance-article/doi/10.1093/tropej/fmaa110/6043117 by Delhi Technological University user on 05 January 2021

Table 2: Factors associated with leave against medical advice

Parameter	Whole LAMA cohort ($n = 528$)	ICU LAMA ($n = 63$)	Ward LAMA ($n = 465$)	OR (CI 95%)	p value
Primary decision maker (%)					0.33
Father	384 (72.73)	49 (77.78)	335 (72.04)	1.3582 (0.7251–2.5439)	
Other family members	144 (27.27)	14 (22.22)	130 (27.95)		
Decision in whose interest (%)					
Family	230(43.56)	28 (44.44)	270 (58.06)	1.7308 (1.0188 to 2.9404)	
Child	298(56.44)	35 (55.56)	195 (41.94)		0.04
Destination taken to after LAMA (%)					
Home	430(81.44)	45 (71.43)	385 (82.80)	1.9250 (1.0593 to 3.4982)	
Private care	98(18.56)	18 (28.57)	80 (17.20)		0.03
Duration of stay before availing LAMA (%)					
<7 days	432 (81.82)	34 (53.97)	398 (85.59)	0.1974 (0.1129–0.3452)	
>7 days	96 (18.18)	29 (46.03)	67 (14.41)		0.0001
Previous admission (%)					
Yes	48 (9.09)	3(4.76)	45 (9.6)	(0.1406 to 1.5488)	0.46
No	480 (90.91)	60 (95.24)	420 (90.32)		

Notes: LAMA = leave against medical advice; ICU = intensive care unit; OR = odds ratio; CI 95% = confidence interval 95%.

Table 3: Comparing the socio-economic status of the patients with the reasons for leave against medical advice

Socio-economic status	Dissatisfaction ($n = 119$)	Family obligations ($n = 88$)	Financial problem ($n = 140$)	Poor prognosis ($n = 91$)	Symptomatic relief ($n = 90$)	p value
Lower (%)	1 (0.85)	1 (1.14)	3 (2.14)	4 (4.34)	1 (1.11)	
Lower middle (%)	77 (64.7)	43 (48.86)	2 (1.43)	38 (41.76)	52 (57.77)	
Upper lower (%)	34 (28.57)	42 (47.73)	134 (95.71)	48 (52.75)	36 (40)	<0.05
Upper middle (%)	7 (5.88)	2 (2.27)	1 (0.72)	1 (1.15)	1 (1.11)	

Notes: Pearson's chi-squared = 153.667. A p value in the last column is from Pearson's chi-squared test.

than that of other African countries [12]. LAMA rates vary between countries and even in a same country. Low prevalence (<1%) of LAMA was reported in Central Nigeria and Oman [13, 14]. This variation appears to be a function of different study settings, geographical and socio-economical factors [15]. Our study shows that in neonates there is a tendency towards higher rate of LAMA in ICU setting which is statistically significant, similar to elsewhere [9, 16].Overall there is no significant gender bias among LAMA cases. This is consistent with the literature [11]. Like other countries [16] majority patients who were taken LAMA belonged to rural background.

The commonest admission diagnosis among newborns was sepsis. However, overall LAMA

Downloaded from https://academic.oup.com/tropej/advance-article/doi/10.1093/tropej/fmaa110/6043117 by Delhi Technological University user on 05 January 2021

population shows respiratory system involvement. El Malek *et al.* [10] and Ndu *et al.* [17] suggested the same.

In majority of cases, primary decision maker for LAMA was father but it did not influence the incidence. LAMA was taken in favor of the family for those admitted in ward as compared to ICU patients. Critically ill patients were shifted to another hospital. Due to lack of studies these facts are not comparable. Most families availed LAMA within 7 days of admission with a mean of 3.51 days (± 4.90). This type of early discharge leads to wastage of major resources [15].

In accordance with other studies [5, 16], financial constrain is cited as principal cause of LAMA. It is significantly prevalent in upper-lower and lower-middle socio-economic classes. This is a global finding with exception of Iran, where temporary improvement was cited as most common cause [11].

In India, adult data on LAMA are limited and mostly retrospective. There is significant knowledge gap regarding predictors of LAMA, affecting not only the patients but also the health care system [18]. Studies on LAMA in Indian children are nearly non-existing. There is need [19] to explore the causes of LAMA in Indian setting, but current data largely reflects African scenario. We failed to find a single published study in last 5 years highlighting factors deciding childhood LAMA both in ICU and ward set-up.

This is a single-center study which may affect generalization of the study findings. Many reasons for LAMA like satisfaction level to mode of treatment, duration of treatment, behavior of health care staffs and hospital environment were clubbed and broadly categorized as dissatisfaction for ease of statistical analysis. LAMA patients were lost to follow up so we were not able to comment on the outcome.

In conclusion we can say that, despite limitations our works are an attempt to elucidate the outlook of childhood LAMA. We need a strategic guideline that helps dealing with LAMA cases and devise preventive strategies at the same time.

What is Already Known?

Burden and determinants of childhood LAMA are nearly non-existing in Indian literature

What This Study Adds

Burden and related factors of LAMA in children in Indian setting identified

ETHICAL CLEARANCE OF THE STUDY AND PERMISSION NUMBER

Ethical clearance obtained from Institutional Ethics committee with Ref No. IEC/NBMC/2018-2019/80 Dated 7 January 2019.

REFERENCES

1. Taqueti VR. Leaving against medical advice. N Engl J Med 2007;357:213–5.
2. Hwang SW, Li J, Gupta R, *et al.* What happens to patients who leave hospital against medical advice? CMAJ 2003; 168:417–20.
3. Beauchamp TL, Childress JF. Principles of Biomedical Ethics, 5th edn. New York: Oxford University Press; 2001.p, 454.
4. Park K. Park's Textbook of Preventive and Social Medicine, 25th edn. Jabalpur: M/s Banarsidas Bhanot Publishers, 2019, 524.
5. Awasthi S, Pandey N. Rural background and low parental literacy associated with discharge against medical advice from a tertiary care government hospital in India. ClinEpidemiol Global Health 2015;3:24–8.
6. Park K. Park's Textbook of Preventive and Social Medicine, 25th edn. Jabalpur: M/s Banarsidas Bhanot Publishers, 2019, 610–28.
7. Wbhealth.gov.in.2020. [online] Available at: https://www.wbhealth.gov.in/uploaded_files/go/SPSRC-2603.pdf (25 August 2020, date last accessed).
8. Park K. Park's Textbook of Preventive and Social Medicine, 25th edn. Jabalpur: M/s Banarsidas Bhanot Publishers, 2019, 745–6.
9. Onyiriuka AN. Pediatric discharge against medical advice: experience from a Nigerian secondary healthcare institution. Med J Islam Repub Iran 2011;25: 194–9.
10. El Malek VA, Alexander S, Al Anezi F. Discharge against medical advice among children admitted into pediatric wards at Al-Jahra Hospital. Kuwait Kuwait Med J 2014;46: 28–31.
11. Roodpeyma S, Hoseyni SAE. Discharge of children from hospital against medical advice. World J Pediatr 2010;6: 353–6.
12. Meremikwu MM, Ehiri JE, Nkanga DG, *et al.* Socioeconomic constraints to effective management of Burkitt's lymphoma in South-eastern Nigeria. Trop Med Int Health 2005;10:92–8.
13. Adebowale Adefemi S. The rate, reasons and predictors of hospital discharge against medical advice among inpatients of a tertiary health facility in North-Central Nigeria. Am J Health Res 2015;3:11.

Downloaded from https://academic.oup.com/tropej/advance-article/doi/10.1093/tropej/fmaa110/6043117 by Delhi Technological University user on 05 January 2021

14. Al-Sadoon M, Al-Shamousi K. Discharge against medical advice among children in Oman: a university hospital experience. Sultan Quboos Univ Med J 2013;13:534– 8.
15. Devpura B, Bhadesia P, Nimbalkar S, *et al.* Discharge against Medical Advice at Neonatal Intensive Care Unit in Gujarat, India. Int J Pediatr 2016;2016:1–8. http://dx.doi.org/10.1155/2016/1897039.
16. Ibekwe RC, Muoneke VU, Nnebe-Agumadu UH, *et al.* Factors influencing discharge against medical advice among paediatric patients in Abakaliki, Southeastern Nigeria. J Trop Pediatr 2007;55:39–41.
17. Ndu IK, Asinobi IN, Ekwochi U, *et al.* Discharge against medical advice (DAMA) among the paediatric age group in Enugu State University Teaching Hospital Parklane, Enugu. J Experimental Res 2016;4:55–62.
18. Gunchan P, Gautam PL, Rubina K, *et al.* Prospective evaluation of patients leaving against medical advice of a tertiary care hospital: comparison of emergency and intensive care units. Int J Med Public Health 2018;8:18–23.
19. Roy MP, Gupta R. Leaving against medical advice: underestimated challenge in children. Indian Pediatr 2016;53:264.

Downloaded from https://academic.oup.com/tropej/advance-article/doi/10.1093/tropej/fmaa110/6043117 by Delhi Technological University user on 05 January 2021

Multi-functional circuit using two Current Feedback Operational Amplifiers

Yashwant Singh[a], Suyash Jain[b], Shubham Snehill[c], Ram Bhagat[d], DC Meena[e]
Department of Electrical Engineering
Delhi Technological University
Delhi, India

[a]yashsh.97@gmail.com, [b]jain.suyash28@gmail.com, [c]shubhamsnehill@gmail.com, [d]rambhagat@dtu.ac.in, [e]dcmeena@dce.ac.in

***Abstract*— This research proposes a newly discovered Current Feedback Operational Amplifier (CFOA) based multi-functional circuit. This paper presents a new Single Resistance Controlled Oscillator (SRCO) and from the same configuration a High Pass Filter (HPF) and Band Pass Filter (BPF) have been implemented. The oscillating frequency of the proposed oscillator can be controlled using a single resistance only. The circuit uses two CFOAs,** three **resistors and two capacitors (one grounded and one virtually grounded). Commercial AD844 IC has been used to verify the workability of the circuit. Also, sensitivity analysis has been carried out for the circuit which directly indicates the proper functionality of the proposed configuration in practical circuits.**

Keywords— Active Filters, Analog Signal Processing, Current Feedback Operational Amplifier, Grounded Capacitor, Single Resistance Controlled Oscillator.

I. INTRODUCTION

The CFOA is a popular active building block and better than the traditional building blocks such as operational amplifiers and current conveyors as seen in recent publications in terms of slew rate and bandwidth product. Active filters and oscillators have always received a special attention in analog circuit design and signal processing since it has various applications in instrumentation, control systems and analog signal processing [1-2]. Many linear and non-linear oscillators have already been introduced out of which Sinusoidal oscillators (linear oscillator) has maximum practical applications in the electrical domain. It turns out to be more advantageous when the oscillation frequency can be controlled through a single resistance only without disturbing the condition of oscillation and hence the nomenclature single resistance-controlled oscillator (SRCO). Many authors in previous research works have proposed numerous structures of oscillators using different building blocks. Few among these include the traditional ICs like operational amplifier and current conveyor whereas other oscillators employ new building blocks like Voltage Differencing Inverting Buffered Amplifier (VDIBA), Voltage Differencing-Differential Input Buffered Amplifiers (VD-DIBA), Four-terminal Floating Nullor (FTFN), Operational Trans-resistance Amplifier (OTRA), Differential Voltage Complementary Current Conveyor (DVCCC), Current Differencing Transconductance Amplifier (CDTA) and Current Feedback Operational Amplifier (CFOA). It is evident from the recent publications that sinusoidal oscillators using these modern building blocks are more precise, fast and also provides wider frequency ranges relative to the oscillators procured with the use of traditional counterparts. There are numerous SRCO circuits which have been realized with help of CFOAs as cited in [13-23]. Oscillators can be designed following various approaches like state variable approach or by transformation of bandpass filter. In [17],[18] and [24], multiple SRCO designs are proposed all of them following the state variable approach. The main emphasis of this paper remains with designing oscillators which uses CFOA and has exceptional performance under permissible error limits. In this paper a new SRCO has been proposed using transformational approach shown in Fig. 2. Remarkably, the proposed circuit also realizes second order high pass filter (HPF) and second order band pass filter (BPF) by breaking the feedback path at point 'P' and applying sinusoidal input signal to z-terminal of CFOA through resistor R_1. The filter circuits have independent tunability of cut of frequency, bandwidth/quality factor and voltage gain. The proposed oscillator employs two CFOAs, three resistors and two capacitors. Designed circuit has been verified by analysing the results of its hardware implementation and also sensitivity analysis has been conducted.

This paper has been divided into six sections. The introduction of the discussed topic is mentioned in first section. Section II contains a short literature review of relevant work performed in past. Section III includes the proposed configuration of SRCO. Non-ideal and sensitivity analysis of the proposed circuit has been done in Section IV. Section V and Section VI includes the Experimental results and Concluding remarks respectively.

II. LITERATURE REVIEW

Oscillators can be obtained using different active blocks as discussed in the previous section. Every oscillator that has been proposed till now differ not only on the basis of their performance but also on quantity of the passive elements used in their respective designs. Evaluating the performance of circuits is considered a traditional criterion to classify them in analog design, but classification can also be performed by evaluating complexity of the designs. The complexity of design can be evaluated by observing the total number of building blocks or passive elements incorporated. Another important factor that distinguishes the performance is the number of grounded/floating capacitors in circuitry design. In [3-24] use of modern building blocks states that these provides better results when compared to traditional Op-Amp

978-1-7281-9287-1/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on January 05,2021 at 07:02:06 UTC from IEEE Xplore. Restrictions apply.

based oscillators asserting superiority over traditional Op-Amps in many ways. In [3], the proposed SRCO is using two VDIBA's but only three passive elements which is one of the simplest designs for an oscillator. In [4-11] only one building block is used employing different number of passive elements with various types of capacitors. Out of the two capacitors, either both can be grounded or only one grounded and one be a floating capacitor. One can witness implementations given in [4], [5], [10] and [11] have used four passive elements each but in [4] only it has both capacitors grounded while [5], [10] and [11] makes use of grounded and floating capacitor both.

CFOA being a commercial and easily available IC relative to other modern building blocks, has always been one of the prime choices for many researchers. The circuits based on CFOA are highly functional, show satisfactory performance and have easy hardware implementations. References [14-24] employs only CFOA as building blocks but differ in quantity of peripherals used. [14-19] uses two building blocks where [14],[16],[17],[18] and [19] employs only five passive elements whereas [15] employs six passive elements, but all of them has two grounded capacitors. References [20-22] proposes oscillators based on single CFOA and uses five/seven passive elements. They all have one floating and one grounded capacitor. Also there have been oscillators that use three building blocks, six passive elements and both the capacitors as grounded [23-24].

The SRCO design proposed in this paper uses two CFOAs, two capacitors (one grounded and one virtually grounded) and three resistors. Considering the circuit performance wise, it shows results very close to the expected theoretical results. The proposed oscillator in its filter form gives bandpass and high pass filters making the circuit highly versatile.

III. PROPOSED CONFIGURATION

A symbolic notation of the current feedback operational amplifier is shown in Fig. 1 where x is low impedance current inverting terminal, y is high resistance voltage non-inverting input terminal, z is high impedance current output terminal and w is low impedance output voltage terminal.

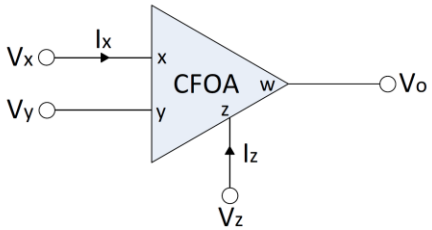


Fig. 1 Symbolic representaion of CFOA

The CFOAs discussed in this paper are characterized by following equations depicted in hybrid matrix [1] given below:

$$\begin{bmatrix} V_x \\ I_y \\ I_z \\ V_w \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} I_x \\ V_y \\ V_z \\ I_w \end{bmatrix} \quad (1)$$

The single resistance controlled sinusoidal oscillator presented in this paper is shown in Fig. 2, which employs a minimum of 5 passive components including 2 capacitors (one grounded and one virtually grounded), and three resistors and two CFOAs. Given below are the Characteristic Equation (CE) along with CO and FO derived for the proposed SRCO:

$$CE: S^2 + \left(\frac{1}{R_2C_1} - \frac{1}{R_4C_5}\right)S + \frac{1}{R_3R_4C_1C_5} = 0 \quad (2)$$

$$CO: R_2C_1 = R_4C_5 \quad (3)$$

$$FO: \omega_0 = \sqrt{1/R_3R_4C_1C_5} \quad (4)$$

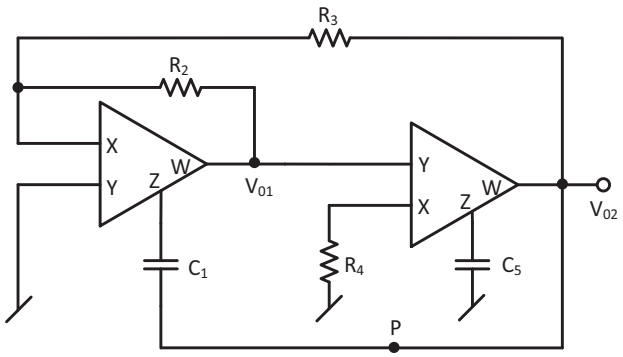


Fig. 2 Proposed Sinusoidal Oscillator circuit

Also, from (3) and (4) it can be deduced that the proposed circuit has independent control of FO and CO. Remarkably, the proposed circuit mentioned in Fig. 2 also realizes HPF and BPF at nodes V_{01} and V_{02} respectively by breaking the feedback path at point 'P' in Fig. 2 and applying input signal to z-terminal of the CFOA through capacitor C_1.

The transfer function of HPF is:

$$\frac{V_{01}(s)}{V_{in}(s)} = \frac{S^2}{S^2 + \frac{S}{R_2C_1} + \frac{1}{R_3R_4C_1C_5}} \quad (5)$$

From the above transfer function, the cut-off frequency (ω_0), quality factor (Q) and voltage gain (H_0) are obtained as follows:

$$\omega_0 = \sqrt{1/R_3R_4C_1C_5} \quad (6)$$

$$Q = R_2\sqrt{C_1/R_3R_4C_5} \quad (7)$$

$$H_0 = 1 \quad (8)$$

Similarly, the transfer function of BPF is:

$$\frac{V_{02}(s)}{V_{in}(s)} = \frac{\frac{S}{R_4C_5}}{S^2 + \frac{S}{R_2C_1} + \frac{1}{R_3R_4C_1C_5}} \quad (9)$$

The above transfer function corresponds to BPF with cut-off frequency (ω_0), Bandwidth (BW) and H_0 are as follows:

$$\omega_0 = \sqrt{1/R_3R_4C_1C_5} \quad (10)$$

$$BW = 1/R_2C_1 \quad (11)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on January 05,2021 at 07:02:06 UTC from IEEE Xplore. Restrictions apply.

$$H_0 = R_2C_1/R_4C_5 \quad (12)$$

From above equations it has been observed that the realized BPF circuit has independent tunability of cutoff frequency, BW and voltage gain. While for the active HPF, the voltage gain remains constant at unity.

IV. Non-Ideal and Sensitivity Analysis

The non-ideal performance characteristics can be realized by considering a small resistance R_x at x-terminal of CFOA, parasitic resistance (R_z) in parallel combination to parasitic capacitance (C_z) at z-terminal of CFOA. The port equation of the non-ideal CFOA is given by:

$$\begin{bmatrix} V_x \\ I_y \\ I_z \\ V_w \end{bmatrix} = \begin{bmatrix} 0 & \alpha & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \gamma & 0 & 0 & 0 \\ 0 & 0 & \beta & 0 \end{bmatrix} \begin{bmatrix} I_x \\ V_y \\ V_z \\ I_w \end{bmatrix} \quad (13)$$

Here, the voltage transfer ratio from x-terminal to y-terminal is α, voltage transfer ratio from w-terminal to z-terminal is β and γ is the current transfer ratio from x-terminal to z-terminal of CFOA. A comparison between ideal and non-ideal parameters of the proposed HPF, BPF and SRCO (assuming R_x=0, R_z=∞ and C_z=0) are tabulated in Table 1 and Table 2 where parameters with '¯' indicates the non-ideal quantities with respect to ideal parameters. Also, from Table 1 and Table 2 it can be deduced that the parasitic effects of the CFOA does have very small effect on CO, FO, ω_0, Q and H_o.

The sensitivity of a function F with respect to parameter of interest T is defined as:

$$S_T^F = \frac{T}{F}\frac{\partial F}{\partial T}$$

The sensitivity analysis of various parameters of the realized SRCO, LPF and BPF are tabulated in Table 3. Also, from Table 3, it has been observed that the proposed SRCO, HPF and BPF promises very low passive sensitivity.

Table 1. COMPARISON BETWEEN IDEAL AND NON-IDEAL PARAMETERS OF SRCO

SRCO		**Frequency of oscillation**	**Condition of oscillation**
Oscillator	**Ideal**	$\omega_0 = \sqrt{\frac{1}{R_3R_4C_1C_5}}$	$R_2C_1 = R_4C_5$
	Non-Ideal	$\overline{\omega_0} = \sqrt{\frac{\alpha_1\gamma_1\alpha_2\beta_2\gamma_2}{R_3R_4C_1C_5}}$	$R_4C_5 \leq \frac{\alpha_2\beta_2\gamma_2R_2C_1}{\gamma_1}$

Table 2. COMPARISON BETWEEN IDEAL AND NON-IDEAL PARAMETERS OF HPF AND BPF

Filters		**Frequency**	**Voltage Gain**	**BW / Q**
Highpass Filter	**Ideal**	$\omega_0 = \sqrt{\frac{1}{R_3R_4C_1C_5}}$	$H_0 = 1$	$Q = R_2\sqrt{\frac{C_1}{R_3R_4C_5}}$
	Non-Ideal	$\overline{\omega_0} = \omega_0\sqrt{\alpha_1\gamma_1\alpha_2\beta_2\gamma_2}$	$\overline{H_0} = \alpha_1H_0$	$\bar{Q} = Q\sqrt{\frac{\alpha_2\beta_2\gamma_2}{\alpha_1\gamma_1}}$
Bandpass Filter	**Ideal**	$\omega_0 = \sqrt{\frac{1}{R_3R_4C_1C_5}}$	$H_0 = \frac{R_2C_1}{R_4C_5}$	$BW = \frac{1}{R_2C_1}$
	Non-Ideal	$\overline{\omega_0} = \omega_0\sqrt{\alpha_1\gamma_1\alpha_2\beta_2\gamma_2}$	$\overline{H_0} = H_0\frac{\alpha_2\beta_2\gamma_2}{\gamma_1}$	$\overline{BW} = \alpha_1\gamma_1BW$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on January 05,2021 at 07:02:06 UTC from IEEE Xplore. Restrictions apply.

Table 3. SENSITIVITY ANALYSIS OF SRCO, HPF AND BPF

Filter	Sensitivity
SRCO	$S_{R_2}^{\omega_0} = 0\,,\, S_{R_3}^{\omega_0} = S_{R_4}^{\omega_0} = S_{C_1}^{\omega_0} = S_{C_5}^{\omega_0} = \frac{-1}{2}$
HPF	$S_{R_2}^{\omega_0} = 0\,,\, S_{R_3}^{\omega_0} = S_{R_4}^{\omega_0} = S_{C_1}^{\omega_0} = S_{C_5}^{\omega_0} = \frac{-1}{2}$ $S_{C_1}^{Q} = \frac{1}{2}\,,\, S_{R_2}^{Q} = 1\,,\, S_{R_3}^{Q} = S_{R_4}^{Q} = S_{C_5}^{Q} = \frac{-1}{2}$ $S_{R_2}^{H_0} = S_{R_3}^{H_0} = S_{R_4}^{H_0} = S_{C_1}^{H_0} = S_{C_5}^{H_0} = 0$
BPF	$S_{R_2}^{\omega_0} = 0\,,\, S_{R_3}^{\omega_0} = S_{R_4}^{\omega_0} = S_{C_1}^{\omega_0} = S_{C_5}^{\omega_0} = \frac{-1}{2}$ $S_{C_1}^{BW} = \frac{1}{2}\,,\, S_{R_2}^{BW} = 1\,,\, S_{R_3}^{BW} = S_{R_4}^{BW} = S_{C_5}^{BW} = \frac{-1}{2}$ $S_{C_1}^{H_0} = S_{R_2}^{H_0} = 1\,,\; S_{R_3}^{H_0} = 0\,,\; S_{R_4}^{H_0} = S_{C_5}^{H_0} = -1$

V. EXPERIMENTAL RESULTS

The Feasibility of the proposed circuit depicted in Fig. 2 has been verified by hardware implementation using two commercial AD-844 type CFOAs biased with DC potential of ±12V and about 5% tolerance passive RC components. The values of the various passive components selected for realizing the proposed oscillator are R_2 =1.5 kΩ, R_3 = R_4= 1 kΩ, and C_1 =C_5=1 nF (where R_2 is implemented using 10 kΩ pot.). Similarly, the values of the various passive components selected for realizing active filters are R_2 = 1.05 kΩ, R_3 = 1 kΩ, R_4 = 920 Ω and C_1 =C_5=1 nF where R_2 is implemented using 10 kΩ pot. and R_4 using 1 kΩ pot.

The experimental output of SRCO transient wave is shown in Fig. 3. The Frequency recorded on Keysight DSO-X 3034T oscilloscope is 152.05 kHz against the designed frequency of 159 kHz which shows the marginal error is less than 4.37% with a voltage of 3.942V (pk-pk). The Fig. 4 shows the Fourier transform of the obtained sinusoidal oscillator which also indicates the percentage THD value of 1.56% only.

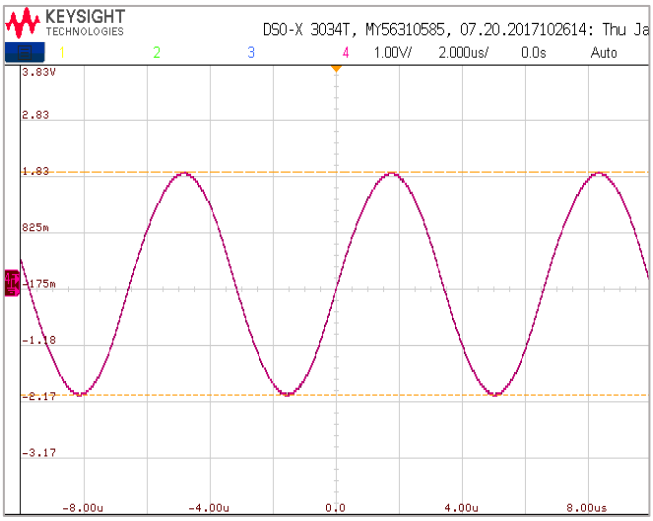


Fig. 3 Transient output waveform of SRCO recorded on Oscilloscope

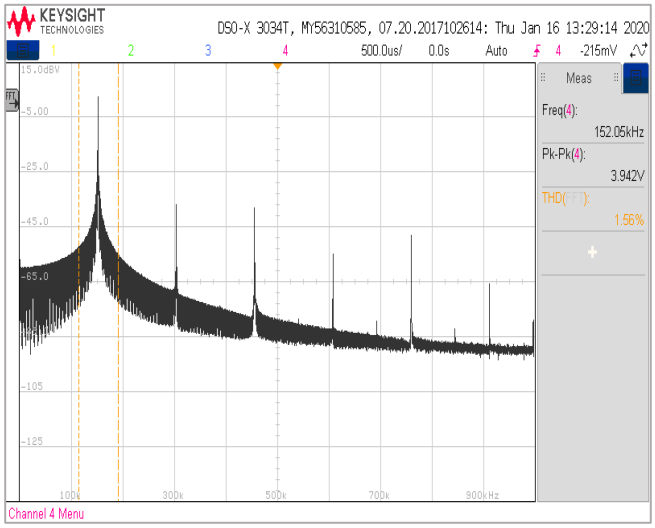


Fig. 4 Fourier transform of the Sinusoidal output wave

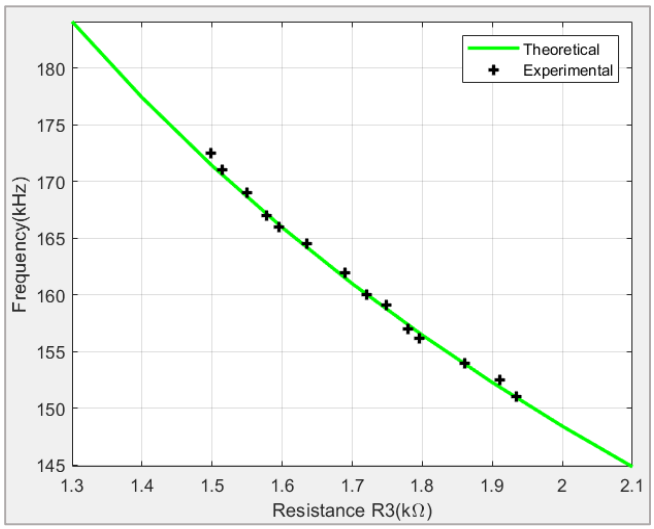


Fig. 5 Variation of FO of SRCO with Resistance

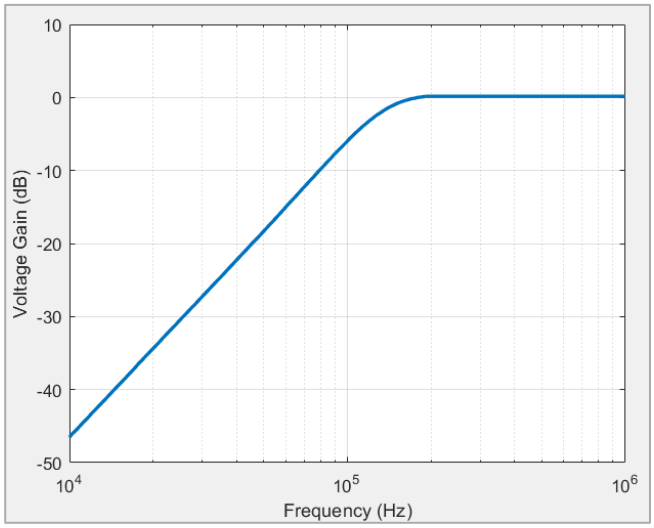


Fig. 6 Frequency response of HPF plotted on MATLAB using .csv file from oscilloscope

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on January 05,2021 at 07:02:06 UTC from IEEE Xplore. Restrictions apply.

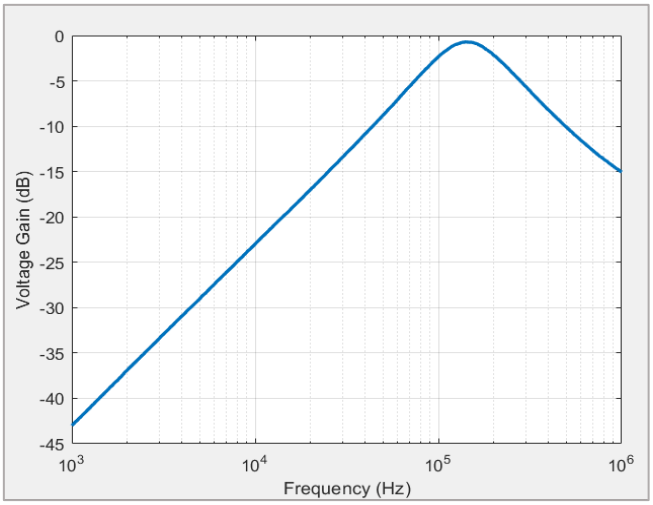


Fig. 7 Frequency response of BPF plotted on MATLAB using .csv file from oscilloscope

Fig. 5 shows the variation of frequency of oscillation (FO) with resistance obtained experimentally. It can be observed that the theoretical values of FO against R_3 are in close proximity of Experimentally obtained values. Fig. 6 and Fig. 7 shows the frequency response of the proposed HPF and BPF plotted on MATLAB using .csv data file from oscilloscope and hence proved the workability of the active filter circuits.

VI. Concluding Remarks

A new single resistance-controlled sinusoidal oscillator with independent control of CO and FO has been derived Fig. 2 which employs two CFOAs, three resistors and two capacitors (one grounded capacitor and one virtually grounded capacitor). The proposed multi-functional circuit also realizes tunable HPF and BPF by breaking the feedback at point 'P' and applying input signal to z-terminal of the CFOA through capacitor C_1. The workability of the circuit has been established through hardware implementation using commercial AD-844 type CFOAs. The circuits presented in this paper enjoys low passive sensitivity. The experimental results are in close agreement with the theoretical results.

References

[1] R. Senani, "Realization of a class of analog signal processing/signal generation circuits: Novel configurations using current feedback op-amps," *Frequenz: J. Telecommun.*, vol. 52, no. 9/10, pp. 196–206, 1998.

[2] A. M. Soliman, "Applications of the current feedback amplifier," *Analog Integr. Circuits Signal Process.*, vol. 11, no. 3, pp. 265–302, 1996.

[3] Pushkar, Kanhaiya & Singh, Ghanshyam & Goel, Rajendra., "CMOS VDIBAs-Based Single-Resistance-Controlled Voltage-Mode Sinusoidal Oscillator", *Circuits and Systems*, vol. 08, no. 1, pp. 14-22, 2017.

[4] Pushkar, K. & Bhaskar, D. & Prasad, Dinesh., "Single-Resistance-Controlled Sinusoidal Oscillator Using Single VD-DIBA", *Active and Passive Electronic Components*, vol. 2013, 2013.

[5] Pushkar, K.L., Goel, R.K., Gupta, K., Vivek, P. and Ashraf, J., "New VD-DIBA-Based Single-Resistance-Controlled Sinusoidal Oscillator", *Circuits and Systems*, vol. 7, no. 13, pp. 4145-4153, 2016.

[6] Lee, C.-T & Wang, H.Y., "Minimum realisation for FTFN-based SRCO", *Electronics Letters*, vol. 37, no. 20, pp. 1207 – 1208, 2001.

[7] Bhaskar, D., "Grounded-capacitor SRCO using only one PFTFN", *Electronics Letters*, vol. 38, pp. 1156 – 1157, 2002.

[8] Çam, Uğur., "A Novel Single-Resistance-Controlled Sinusoidal Oscillator Employing Single Operational Transresistance Amplifier", *Analog Integrated Circuits and Signal Processing,* vol. 32, pp. 183-186, 2002.

[9] Aggarwal, Varun & Kılınç, Selçuk & Çam, Uğur, "Minimum component SRCO and VFO using a single DVCCC", *Analog Integrated Circuits and Signal Processing*, vol. 49, pp. 181-185, 2006.

[10] Chien, H.-C., "CDTA and CCCDTA-based mixed-mode single-resistance-controlled sinusoidal oscillators", *Journal of Applied Science and Engineering,* vol. 17, no. 4, pp. 437-446, 2014.

[11] Prasad, Dinesh & Bhaskar, D. & Singh, Abdhesh, "Realisation of Single-Resistance-Controlled Sinusoidal Oscillator: A new application of the CDTA", *WSEAS Transactions on Electronics*, vol. 5, Issue 6, pp. 257-259, 2008.

[12] Herencsar, Norbert & Vrba, Kamil & Koton, Jaroslav & Lahiri, Abhirup, "Realisations of single-resistance-controlled quadrature oscillators using generalised current follower transconductance amplifier and unity-gain voltage-follower", *International Journal of Electronics*, vol. 97, Issue 8, pp. 897-906, 2010.

[13] Mongkolwai, Pratya & Pukkalanun, Tattaya & Dumawipata, Teerasilapa & Tangsrirat, Worapong., "CFOA-based single resistance-controlled quadrature oscillator", *Proceedings of the SICE Annual Conference*, pp. 1147 – 1150, 2008.

[14] R. Senani and Vinod Kumar Singh, "Novel single-resistance-controlled oscillator configuration using current feedback amplifiers", IEEE transactions on Circuits and Systems, 1996.

[15] S. I. Liu and J.-H. Tsay, "Single-resistance-controlled sinusoidal oscillator using current-feedback amplifiers," *Int. J. Electron.*, vol. 80, no. 5, pp. 661–664, 1996.

[16] Bhaskar, D.R., and Senani, R.: "New CFOA based single element-controlled sinusoidal oscillators", IEEE transactions on instrumentation and measurement, vol. 55, no. 6, pp. 2014-2021, 2006.

[17] S.S. Gupta, R. Senani, "State variable synthesis of single resistance controlled grounded capacitor oscillators using only two CFOAs" in *IEE Proc.Circuits Syst.,* vol.145*,* No.2, April 1998.

[18] S. S. Gupta and R. Senani, "State variable synthesis of single-resistance-controlled grounded capacitor oscillators using only two CFOAs: additional new realisations," in *IEE Proceedings - Circuits, Devices and Systems*, vol. 145, no. 6, pp. 415-418, Dec. 1998.

[19] S. Celma, P. A. Martinez, and A. Carlosena, "Current feedback amplifiers based sinusoidal oscillators," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 41, no. 12, pp. 906–908, 1994.

[20] Liu, S.I., Shih, C.S., and Wu, D.S.: "Sinusoidal oscillators with single element control using a current feedback amplifier", Int. J. Electron., vol. 77, no. 6, pp. 1007-1013, 1994.

[21] P. A. Martinez, S. Celma, and J. Sabadell, "Designing sinusoidal oscillators with current-feedback amplifiers," *Int. J. Electron.*, vol. 80, no. 5, pp. 637–646, 1996.

[22] M. T. Abuelma'atti and A. M. Al-shahrani, "New CFOA-based sinusoidal oscillators," *Int. J. Electron.*, vol. 82, no. 1, pp. 27–32, 1997.

[23] D. R. Bhaskar, S. S. Gupta, R. Senani, A. K. Singh, "New CFOA-based sinusoidal oscillators retaining independent control of oscillation frequency even under the influence of parasitic impedances", Analog Integr Circ Sig Process, vol. 73, pp. 427-437, 2012.

[24] R. Senani and S. S. Gupta, "Synthesis of single-resistance-controlled oscillators using CFOAs: simple state-variable approach," in *IEE Proceedings - Circuits, Devices and Systems*, vol. 144, no. 2, pp. 104-106, April 1997.

[25] Soliman, A.M., "Transformation of a Floating Capacitor Oscillator to a Family of Grounded Capacitor Oscillators", *International Journal of Electronics*, vol. 98, Issue 3, pp. 289-300, 2011.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on January 05,2021 at 07:02:06 UTC from IEEE Xplore. Restrictions apply.

AIP Conference Proceedings

Numerical simulation of n-TiO2/p-CIGS solar cell

Cite as: AIP Conference Proceedings **2265**, 030675 (2020); https://doi.org/10.1063/5.0017203
Published Online: 05 November 2020

Gagandeep, Mukhtiyar Singh, Ramesh Kumar, and Sonia Kumari

View Online Export Citation

Meet the Next Generation of Quantum Analyzers
And Join the Launch Event on November 17th

Register now
Zurich Instruments

AIP Publishing

© 2020 Author(s).

Numerical Simulation of n-TiO2/p-CIGS Solar Cell

Gagandeep[1, a)], Mukhtiyar Singh[2], Ramesh Kumar[1, b)] and Sonia Kumari[1]

[1]*Department of Physics, Guru Jambheshwar University of Science & Technology, Hisar, India*
[2]*Department of Applied Physics, Delhi Technological University, Delhi, India*

[a)]gagankundu22@gmail.com
Corresponding author: [b)] rameshkumar.bibiyan@gmail.com

Abstract. The mostly used buffer layer in CIGS solar cell is CdS. This buffer layer is toxic. Therefore, the p-type Cu (In, Ga)Se_2 (CIGS) absorber layer is simulated with non toxic and wide band gap TiO_2 material as n-type buffer layer in AFORS-HET software under the illumination condition of AM 1.5 G. To see the effect of absorber layer thickness on the performance parameters of the cell, it is varied from 200-1900 nm. The highest efficiency of the cell is obtained 16.03% for optimized thickness of absorber layer i.e., at 1900 nm. The temperature influences the performance of the cell and its performance is maximum at temperature 300 K. The external quantum efficiency behavior with wavelength is observed. The reported results in this study show that TiO_2 material is better replacement of CdS material.

Key Words: Solar Cell, Efficiency, TiO2, AFORS-HET, CIGS

INTRODUCTION

The chalcopyrite Cu(In, Ga)Se_2 (CIGS) material based solar cells is very promising thin film solar cell and play very important role in the enhancement of photovoltaic technology [1-3]. The CIGS solar cell has excellent outdoor stability. The CIGS based solar cells are attracting considerable space application interest because of its stability against particle irradiation. Its stability is proved by electron and proton irradiation test of CdTe and CIGS solar cells which is superior to Ⅲ-Ⅴ or silicon solar cells. The laboratory scale power conversion efficiency (PCE) of this cell is attaining 20.4% on a flexible polymer substrate [4] and 22.6% on a soda lime glass (SLG) substrate [5]. The CdS plays the role of buffer layer in CIGS solar cell and resulting the highest above mentioned device efficiency. The CdS layer on CIGS is formed by chemical bath deposition (CBD) method. The CdS layer have relatively low band gap energy (2.4-2.5 eV). Hence, the finest performance of the cell is limited by the parasitic absorption in short wavelength region [6]. The CdS buffer layer is also toxic in nature.

Therefore, different buffer layer with a wider band gap is proposed [7- 9]. W. Hsu et. al [10] experimentally studied that amorphous TiO_2 act as buffer layer on CIGS cell and improved the performance of the device

Keeping the observation in mind, we simulated the TiO_2 as buffer layer in CIGS solar cell and obtained the PCE is 16.03%. Our theoretical results reported herein are improved the performance of the device.

SIMULATION DETAILS

Here, we simulate a polycrystalline CIGS based solar cell by n-TiO_2 /p-CIGS heterojunction structure with ITO and Mo are front and back contact respectively. In this paper, we use AFORS-HET (Automat for simulation of heterostructure) [11-13] to analyze the working of CIGS based thin film solar cell. The simulated structure is shown in figure 1.

DAE Solid State Physics Symposium 2019
AIP Conf. Proc. 2265, 030675-1–030675-4; https://doi.org/10.1063/5.0017203
Published by AIP Publishing. 978-0-7354-2025-0/$30.00

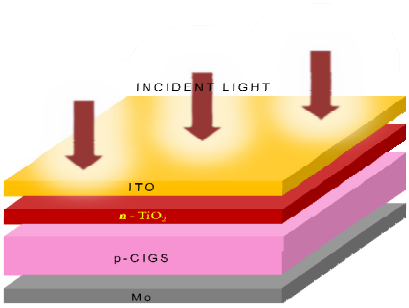


FIGURE 1. Schematic structure of simulated CIGS solar cell

The thickness of the CIGS layer is varied from 200-1900 nm. The variations in the values of the performance parameters of the cell such as short current density (J_{sc}), open circuit voltage (V_{oc}), fill factor (FF) and PCE (η) are noticed with absorber layer thickness. The impact of temperature on the performance parameters of the device is observed. The J-V (current density-voltage) characteristics are investigated for this structure with optimized absorber layer thickness. The quantum efficiency spectra with wavelength are also plotted. The layer parameters used in this work are taken from different theoretical and experimental studies [14-15] and mentioned in Table 1.

TABLE1. Layer parameters used in simulation

Parameter	**n-TiO_2**	**p-CIGS**
Layer thickness (nm)	90	200-1900
Dielectric constant	100	10
Electron affinity (eV)	4	4.2
Band gap (eV)	3.2	1.4
Optical band gap (eV)	3.2	1.4
Valence band density (cm^{-3})	1×10^{21}	2.2×10^{18}
Conduction band density (cm^{-3})	2×10^{20}	1.8×10^{19}
Hole mobility ($cm^2\ V^{-1}\ s^{-1}$)	0.006	100
Electron mobility ($cm^2\ V^{-1}\ s^{-1}$)	0.006	25
Thermal velocity of electron(cm/s)	1×10^{6}	1×10^{7}
Thermal velocity of hole(cm/s)	4×10^{6}	1×10^{7}
Doping concentration(cm^{-3})	5×10^{19}	2×10^{13}
Layer density(g cm^{-3})	2.328	2.328

RESULTS AND DISCUSSIONS

The thickness of CIGS layer is varied from 200-1900 nm and performance of the cell is observed. The cell parameters variation with absorber layer thickness is displayed in fig.2. The value of cell parameters such as J_{sc}, V_{oc}, FF and η are changed from 23.06 mA/cm^2 to 28.01 mA/cm^2, 606.4 mV to 697.7 mV, 81.95% to 82% and 11.46% to 16.03% respectively. It is explained based on the fact that larger thickness of CIGS layer absorbs more radiation and produce more carrier i.e. electron-hole pair and probably increase the performance parameters i.e. V_{oc}, J_{sc}, and η of the device. The FF of the cell is increased up to 82.24% at 1200 nm absorber thickness but decreases up to 82% at 1900 nm absorber thickness. The optimum thickness of CIGS layer is around 1900 nm from which the highest

efficiency of 16.03% is obtained for the n-TiO_2/p-CIGS cell. The results reported herein are accorded with the general trends of the device i.e. increased layer thickness improves the efficiency of the cell [16].

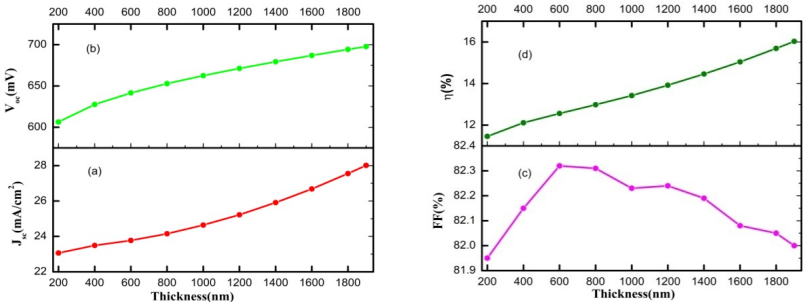


FIGURE2. Variation of solar cell parameters (a) J_{sc}, (b) V_{oc}, (c) FF and (d) Efficiency with absorber layer thickness

The variation of efficiency, FF, J_{sc} and V_{oc} with operating temperature is shown in figure 3. The performance of the cell, are also influenced by the operating temperature. The cell parameters variation is discussed in the temperature range 300-600 K under a constant illumination of 1000 W/m^2. The characteristics parameters of the cell namely V_{oc}, FF and η are strongly influenced by the operating temperature and get reduced with increase in temperature of the cell but J_{sc} remains constant. From the observed results we analyzed that the η of the cell is decreased from 16.03% to 0.15% in the range 300-600 K. It may be probably happened due to that, as we increase the temperature of the cell then there is increased in the reverse saturation current density J_o which is caused by the increase in the intrinsic carrier concentration and decrease in the band gap of the material. The conductivity of the cell is also affected by the temperature. At optimized value of absorber layer thickness J-V characteristics of the cell is also studied and obtain the η of 16.03% with FF of 82%. We also calculate the V_{oc} = 697.7 mV and J_{sc} = 28.01 mA/cm^2 respectively and results are shown in fig. 4(a).

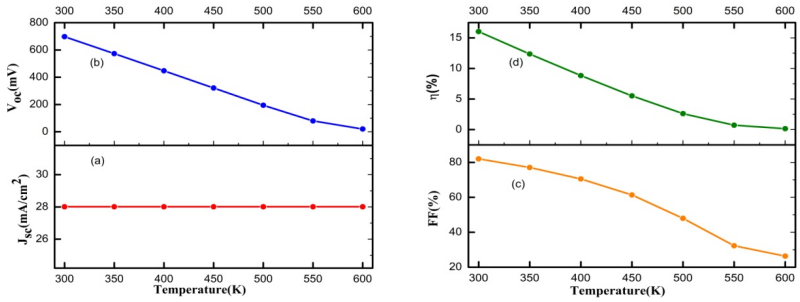


FIGURE3. Variation of solar cell parameters (a) J_{sc}, (b) V_{oc}, (c) FF and (d) Efficiency with operating temperature

The radiation absorbed by the cell is given by the spectra of EQE (external quantum efficiency). The QE spectra provides an information about the absorption of light by the CIGS device as compared to incident one. The EQE provides the detailed study about all photons whether they are absorbed, reflected or transmitted and provide us a maximum value in the wavelength range of 600-900 nm. All the absorbed photon by the cell are counted by the IQE (internal quantum efficiency) and gives maximum in 400-900 nm of wavelength but IQE1 becomes almost unity in this range of wavelength and counts only that photons which are absorbed by the electrical layers. The results of quantum efficiency are shown in fig 4 (b)

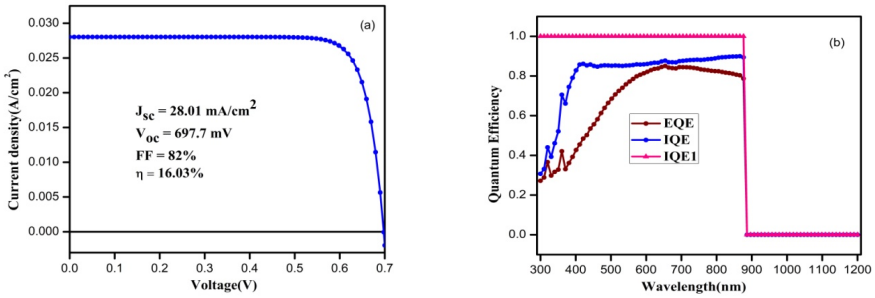


FIGURE4. (a) J-V characteristics (b) Quantum efficiency with wavelength

CONCLUSIONS

Theoretically, we simulated the TiO_2 layer as buffer layer in CIGS solar cell. The thickness of the absorber layer influences the performance of the device. The maximum value of power conversion efficiency is 16.03% at 1900 nm absorber thickness. The cell performance is influenced by the temperature and have attained maximum value at 300 K. The J-V curve and variation of external quantum efficiency of the solar cell with wavelength are also plotted. The obtained results in this study boost the performance of TiO_2 as buffer layer.

ACKNOWLEDGMENT

One of the authors (**Ramesh Kumar**) is thankful to the **University Grants Commission, INDIA** for giving the financial support through START-UP GRANT under FRPS (No. F.30-364/2017 (BSR), Dated 15/12/2017). The authors highly acknowledge the AFORS-HET software developer for providing free access of the software.

REFERENCES

1. J. Löckinger, S. Nishiwaki, T. P. Weiss, B. Bissig et al., Solar Energy Materials and Solar Cells, 174, 397 (2018).
2. C. Zhang, H. Zhu, X. Liang, D. Zhou, Y. Guo, X. Niu, Z. Li, J. Chen, Y. Mai, Vacuum, 141, 89 (2017).
3. Gagandeep, M. Singh and R. Kumar, AIP Conference Proceedings, 2093, 020018 (2019).
4. A. Chirila et.al., Nat. Mater. 12, 1107 (2013).
5. P. Jackson et.al., Phys. Status Solidi-R 10, 583(2016).
6. N. Naghavi et.al., Prog. Photovolt.18,411 (2010).
7. S. Spiering et.al.,Solar Energ.Mater. Sol.C 144, 544 (2016).
8. J. Lindahl et. al., IEEE J. Photovolt.3, 1100 (2013).
9. T. M.Friedlmeir et. al., IEEE J. Photovolt 5, 1487(2015).
10. W. Hsu et.al., Sci. Rep. 5, 16028 (2015).
11. R.Varache et. al., Sol. Energ. Mat. Sol. C.141, 14 (2015).
12. R. Stangl, J. Haschke and C. Leendertz,Solar Energy, Intech (R.D. Rugescu (Ed.), Croatia, 2009), pp. 432.
13. R.Stangl, M. Kriegel and M. Schmidt, Proceeding of the WCPEC-4, Fourth World Conference on Photovoltaic Energy Conversion (Hawaii, USA), PP. 1350 (2006).
14. M. Mostefaoui et. al., Energy Procedia, 74, 736 (2015).
15. S. Dabbabi, T. B. Nasr, N. K. Turki, Results in Physics, 7, 4020 (2017).
16. S. Tobbeche and H. Amar, JNTM, 04, 89 (2014).

The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1463-5771.htm

Outsourcing and reverse supply chain performance: a triple bottom line approach

Saurabh Agrawal
Delhi School of Management, Delhi Technological University, Delhi, India, and
Rajesh Kumar Singh
Management Development Institute, Gurgaon, India

Received 25 September 2020
Revised 7 November 2020
Accepted 8 November 2020

Abstract

Purpose – Outsourcing has become a strategic decision for most of the organizations in the present context of globalization and circular economy. The objective of proposed study is to empirically investigate the relationship between the outsourcing benefits and triple bottom line (TBL)-based performance of reverse supply chain (RSC) comprising of economic, environmental and social factors.

Design/methodology/approach – In the present study, the partial least square path approach of structural equation modeling has been applied for testing hypotheses. A questionnaire was developed for collecting the information and responses from the experts of electronics industry, India. Measurement models were tested for ensuring the fitness of data for further applying partial least square path approach of SEM.

Findings – The results provide insightful information on outsourcing and RSC performance. It reveals that outsourcing decisions are important and have great influence on the economic, environmental and social performance of RSC.

Research limitations/implications – The study will provide useful guidance to the organizations for outsourcing decisions. One of the limitations of the study is that it has considered data collected from the Indian electronics industry. Therefore, results may not be generalized.

Originality/value – The third-party service providers have core competencies, and they are the experts of providing RSC services. It is also apparent from the exploration of past research and theoretical background that there are limited number of studies focusing on the impact of outsourcing decisions in RSC performance specifically from the TBL perspective. Hence, the proposed study is an effort to explore the impact of outsourcing RSC activities on the performance of RSC from a sustainability perspective.

Keywords Reverse supply chain, Reverse logistics, Outsourcing, PLSPM, Structural equation modeling, Indian electronics industry

Paper type Research paper

1. Introduction

Supply chains have been major focus of attention for researchers, academicians and business enterprises. Time responsiveness and cost optimization have been the ultimate goal of supply chain professionals. Singh *et al.* (2007) have observed that while developing strategies for competitiveness, sustainability of operations should be a priority for management. Now many organizations are outsourcing their logistics operations to become sustainable and competitive (Singh *et al.*, 2018). Zarbakhshnia *et al.* (2020a, b) developed a model for outsourcing sustainable reverse supply chain (RSC) and observed that a RSC can make significant contributions to the sustainability aspects of an organization. Until the recent past, the focus had been more on studying the products flow from manufacturer to the consumers and ignored the reverse flows (Bernon and Cullen, 2007). With the increase in the returns of products because of online selling and competitiveness, RSC is becoming an essential part of business. RSC, which is also, referred to as reverse logistics "*is the process of planning, implementing and controlling the efficient, cost effective flow of raw materials, in*

Authors would like to express sincere thanks to Editor of the journal and reviewers for constructive suggestions in revising this manuscript.

Benchmarking: An International Journal
© Emerald Publishing Limited
1463-5771
DOI 10.1108/BIJ-09-2020-0498

process inventory, finished goods and related information from the point of consumption to the point of origin for the purpose of recapturing value or proper disposal" (Rogers and Tibben-Lembke, 1999). RSC differs significantly in comparison to a forward supply chain (Tibben-Lembke and Rogers, 2002). Zarbakhshnia *et al.* (2020a, b) demonstrated that RSC is a key strategy of any supply chain of a business because of lack of raw material and its increased costs. They also mentioned that environmental consciousness is also promoting need of RSC and organizations are adopting, implementing this for their competitive advantages. Also, RSC is an essential component of most of the circular economy models. The concept of circular economy has gained enormous attention because of growing pressure on governments and policy makers to introduce policies to become more sustainable by better managing the available resources (Dubey *et al.*, 2019a).

In the forward supply chain, the demand forecasting of the products is relatively easier in comparison to RSC. In RSC, the consumers may return the product at any point of life cycle in unpredictable condition in terms of quality of returned product. These uncertainties in product returns make RSC more stochastic in nature compared to the forward supply chain. Therefore, product return policies are very important because of uncertainties, and they must be designed carefully for the efficient collection of returned products in the RSC (Östlin *et al.*, 2008). Coordination among supply chain partners plays a very important role in effectiveness of supply chain operations (Kumar and Singh, 2017). However, it is a challenge to balance among customer services, regulatory and operational requirements and competitiveness in the market. While the forward supply chain is the core business of most of the organizations, the RSC has been considered as a cost driven activity. Still, RSC has become essential for organizations because of regulatory and environmental concerns (Ravi and Shankar, 2005). Many organizations have adopted and implemented the RSC successfully for their business advantages. RSC adoption and implementation faces various challenges because of lack of experience in the industry. Lambert *et al.* (2011) discussed these issues for the respective RSC processes.

Allocation of financial resources (Morgan *et al.*, 2018; Lambert *et al.*, 2011), risk of adopting RSC (Ferrer and Ketzenberg, 2004; Senthil *et al.*, 2018), disposition of returned products (Hazen, 2011; Agrawal *et al.*, 2019), outsourcing RSC activities (Kushwaha *et al.*, 2020), integration of forward and RSC (Pedram *et al.*, 2017), performance evaluation system (Han and Trimi, 2018), information technology (IT) infrastructure (Chileshe, 2019), product return policy (Batarfi *et al.*, 2017) and availability of technology (Lambert *et al.*, 2011) are some of major issues observed from the literature. Because of these issues and challenges faced by organizations in carrying out RSC activities, many organizations have explored the option of outsourcing RSC activities.

Organizations outsource part of operations or entire operations to other organizations depending on their business model (Gunasekaran and Spalanzani, 2012). According to Daugherty and Droge (1997), the decision to outsource functions or processes depends on a number of factors. By outsourcing, organizations may give more attention to their core competencies and may reduce investment in resources while responding well to their customers in a dynamic business environment (Kakabadse and Kakabadse, 2005). However, some organizations incurred losses because of problems faced during operations (Tadelis, 2007). Many organizations prefer outsourcing RSC activities because these activities are not part of their core businesses. Another reason for outsourcing is economic benefits to the organizations.

According to Ravi and Shankar (2015), main reason of adopting reverse logistics in Indian manufacturing industry is economic. However, outsourcing RSC activities may also improve the environmental and social performance of the RSC. In the present business environment of online retailing and e-commerce, managing RSCs has become very challenging for electronics industry. The world electronics goods market is expected to grow to approximately three trillion US$ by 2020 (Future Market Insights, 2019). The demand for electronic products in India is also very high and expected to grow at a compound annual growth rate of 41%

during 2017–2020 to reach US$ 400 bn by 2020 (The Related Chambers of Commerce and Industry of India, 2017). However, the Indian electronics market is dependent on the import of parts and components (Agrawal *et al.*, 2015). Returning of products in online retailing is another challenge. Most of the companies are going for outsourcing of reverse logistics operations for becoming sustainable (Chand *et al.*, 2018).

Gorane and Kant (2016) have found that outsourcing has been used moderately by Indian organizations. Very limited research has been done on impacts of outsourcing RSC activities on performance of supply chains (Singh *et al.*, 2018). Organizations are adopting the triple bottom line (TBL) approach comprising of social, economic and environmental dimensions of performance. Organizations can make their RSCs more efficient by outsourcing of different logistics activities. Uncertainty of the acquisition time impacts quality of returned products. Minimizing waste, creating recovering facilities and controlling cost of operations are major challenges for RSC managers (Chakrabory, 2018). Dubey *et al.* (2017) have advocated the need for integrated RSC. To reduce cost and optimizing RSC operations, outsourcing is emerging as a sustainable option. However, sufficient research has not been done in this context. Therefore, we would try to answer following research questions through this study.

RQ1. Will outsourcing RSC activities impact economic performance of supply chains?

RQ2. Will outsourcing RSC activities impact environmental performance of supply chains?

RQ3. Will outsourcing RSC activities impact social performance of supply chains?

The study involves development of hypotheses related to outsourcing RSC activities and TBL performance. Hypotheses are tested using the information and responses gathered from the experts in electronics industry. The study intends to investigate the relation between outsourcing and TBL performance of RSC in the Indian electronics industry. The main construct outlines the benefits of outsourcing in electronics industry. Its association with the other constructs outlining the economic, environmental and social performance of RSC is tested. The latent variables of all constructs have reflective type of observed variables. The intent is to understand the association of benefits of outsourcing with TBL performance of RSC. The measurements models of the constructs developed were tested for fitness of data for further modeling using the partial least square path modeling (PLSPM) approach. PLSPM was selected because it requires smaller sample size; more importantly, it does not need normally distributed data, and it is nonparametric (Hair *et al.*, 2017). Recently, PLSPM has been applied by Feng *et al.* (2018), Ardakani *et al.* (2019) and Chandak *et al.* (2019) in the area of supply chain management.

The remaining part of the study is organized as follow. Section 2 explores the literature related to outsourcing in RSC and TBL performance and explains hypothesis development. Section 3 discusses research methodology comprising of hypothesis testing along with the data collection method. Section 4 deals with results and discussion, including the measurement model and PLSPM model of the study. The study is concluded in Section 5.

2. Literature review

Outsourcing is the term used for "using outside resource" (Quinn and Hilmer, 1994). The organizations adopting RSC systems need to decide on outsourcing various functions or processes to the other outside organizations (Kotabe and Zhao, 2002). Organizations outsource part of operations or entire operations to the other organization depending on their business model (Gunasekaran and Spalanzani, 2012). In 80s, organizations used outsourcing as a cost reducing tool by outsourcing service-related activities (Lacity and Hirschheim, 1993). Later on, in 90s they realized the benefits of outsourcing and started focusing on core activities and outsourcing noncore activities (Hamel and Prahalad, 1990). Activities are also

outsourced for economic benefits while focusing on core competencies (Boyson *et al.*, 1999). If an organization is not good on a particular activity and it is not influencing the competitiveness, the activity must be outsourced (Insigna and Werle, 2000). Arnold (2000) correlated the core competencies with competitiveness and found that core activities are directly related with core activities.

In RSC, activities, which are not the core activities, may be outsourced to the other service providers (Wu *et al.*, 2005). RSC needs specialized network for the collection, segregation, inspection, disposition and processing of returns which requires specialized IT infrastructure and working professionals (Ko and Evans, 2007). Most of the organizations are not experienced and resourced enough to perform RSC functions themselves and prefer outsource them to other service providers (Bai and Sarkis, 2019). Krumwiede and Sheu (2002) insisted on outsourcing RSC activities if resources and capabilities are limited. Govindan *et al.* (2012) cited reduced cost as a major reason for outsourcing RSC activities because of large volume handled by the service providers. IT infrastructure also plays an important role in reduced cost of RSC functions (Huscroft *et al.*, 2013). Mello *et al.* (2008) explored the RSC and found that outsourcing is done for quality and cost reduction. Cost, quality, flexibility, responsiveness and availability of latest technological advancement are the reasons for outsourcing the RSC functions (Kumari *et al.*, 2015; Li *et al.*, 2018a, b; Bai and Sarkis, 2019). Outsourcing is a major decision of RSC and need a lot consideration before any decision. While there are many benefits of outsourcing, sustainability benefits are yet to be explored. The outsourcing RSC may greatly influence the sustainability aspects of an organization (Lee and Lam, 2012; Bai and Sarkis, 2019). The proposed study explores the relation between the outsourcing benefits and the sustainability performance of RSC.

In total, three components of sustainability, i.e. economic, environment and social all together are named as TBL (Elkington, 1998). The economic component addresses the issues related to money for managing product returns, the environment component helps in preserving environment while the social component focuses on developing skills, health and safety, equity and fairness to the employees (Vachon and Mao, 2008; Bansal, 2002; McKenzie, 2004). In most of studies on the RSC, economic performance is the major criteria for the organizations. However, more than 70% of organizations across in the industry are stressed to take initiative for measuring performance of environment and social performance along with economic component of TBL aspects of sustainability (Hubbard, 2009). The return process becomes complex with consideration of economic and social components (Bai and Sarkis, 2019).

There are very limited studies related to RSC focusing on TBL performance. Presley *et al.* (2007) introduced all three TBL components in relation to the RSC. Sarkis *et al.* (2010) further emphasized on the role of RSC in achieving the TBL performance goals in an organization. Agrawal *et al.* (2016) explored the sustainability into RSC and developed the model for measuring its TBL performance. Singh and Agrawal (2018) analyzed disposition strategies in reverse supply chains. Govindan *et al.* (2019) developed the third-party supplier selection model considering criteria based on TBL performance. Gopal and Thakkar (2016) investigated the sustainable supply chain practices for Indian automobile industry. Thanki and Thakkar (2018) developed a performance measurement framework for lean and green supply chain. Bai and Sarkis (2019) have discussed about RSC contribution toward sustainability by developing the model for selection of third-party reverse logistic service providers. There are number of studies focusing on third-party RSC service providers (Govindan *et al.*, 2019; Tavana *et al.*, 2016; Akbari, 2018; Li *et al.*, 2018a, b). The outsourcing in RSC has been mainly focused on third-party selection (Min and Ko, 2008; Govindan and Murugesan, 2011; Guarnieri *et al.*, 2015; Evangelista *et al.*, 2018). The proposed study explores the outsourcing in relation to TBL performance of RSC. Zarbakhshnia *et al.* (2018) proposed a hybrid approach, on the selection of third-party reverse logistics providers (3PRLPs) for a car parts manufacturing company and benchmark it with the Multi-Objective Optimization on

the basis of Ratio Analysis (MOORA) method. Cheng and Lee, 2010 investigated the relative importance of reverse logistics service requirements and illustrated selection of an appropriate third-party logistics providers (3PLPs) using the analytical network process (ANP). Ordoobadi (2009) described the multi-phased decision model for the consideration of both strategic and economic issues of outsourcing reverse logistics activities. Serrato *et al.* (2007) analyzed and formulated the Markov decision model of the outsourcing. They also justified how the threshold for outsourcing decreases while the probability of crossing any fixed threshold increases with the return fraction. Logozar (2008) discussed the characteristics of business logistics in the manufacturing firms. It could be helpful in understanding of outsourcing logistics services. Gu *et al.* (2019) constructed an evolutionary game theoretical model to investigate the cooperation among Natural Resource and Energy Intensive (NREI) companies. Prajapati *et al.* (2019) proposed a hybrid framework using modified step-wise weight assessment ratio analysis (SWARA) and weighted aggregated sum product assessment (WASPAS) to achieve the objective of this research. Yu *et al.* (2018) proposed a two-stage stochastic biobjective mixed integer programming model for the network design problem of a multi-product multi-echelon sustainable reverse logistics system under uncertainty, which aims at providing a set of Pareto solutions between profitability and environmental performance. Zarbakhshnia *et al.* (2018) proposed a multiple attribute decision-making (MADM) model to rank and select 3PRLPs, using fuzzy SWARA to weight the evaluation criteria. SWARA–WASPAS is also utilized for planning the block chain technology by Modgil and Sonwaney (2019). Govindan *et al.* (2018) presented a multi-perspective framework for reverse logistics implementation using the lens of stakeholder theory and analyzed drivers' barriers against the dimensions and categories, served as a basis for the development of the referred framework. Bai and Sarkis (2018) provided integrating and extending data and decision tools for sustainable third-party reverse logistics. Li *et al.* (2018a, b) investigated an integrated cumulative prospect theory (CPT) based on a hybrid-information multi-criteria decision making (HI-MCDM) approach to appraise and identify sustainable 3PRLPs. Zhikang (2017) presented current situation and issues of the reverse logistic of the automobile industry in China and also used SWOT analysis to develop reverse logistic strategies. Senthil *et al.* (2018) analyzed and prioritized risks in a reverse logistics network using hybrid multi-criteria decision-making methods. Ameknassi *et al.* (2016) analyzed the relationship between logistics outsourcing decisions with strategic planning issues. Espino-Rodríguez *et al.* (2017) identified the role of the outsourcing strategy in the hotel sector. They also determined what strategic benefits and risks determine an outsourcing strategy by hotel managers. Giri *et al.* (2017) proposed a model through a series of data sets and derived analytically the pricing and return product collection decisions for the supply chain under five different scenarios, namely, centralized, decentralized (Nash game) and manufacturer-led, retailer-led and third party-led decentralized scenarios. Ji *et al.* (2018) used the mean-squared deviation weight method to obtain the weights of criteria for selecting an outsourcing provider based on the combined Multi-Attributive Border Approximation Area Comparison - Elimination and Choice Translating Reality (MABAC-ELECTRE) method using single-valued neuromorphic linguistic sets. Tian *et al.* (2019) adopted the take-back pattern of vehicle reverse logistics in China using grey Decision Making Trial and Evaluation Laboratory (DEMATEL) and fuzzy-VIKOR combined method. Chakraborty *et al.* (2020) explored the crucial barriers to IT implementation in the largely unorganized logistics sector of an emerging economy, India. Golroudbary *et al.* (2019) proposed a hybrid simulation framework with the combination of agent based modeling and system dynamics to address the dynamics ricks effects in the logistic delivery.

Outsourcing is a very common phenomenon in global business. The outsourcing in RSC is also very popular because of nature of business and several other advantages. Based on the literature review discussed above, the research identified some of the key outsourcing

benefits. These are OB1: Focus on core activity of the organization, OB2: Economical to outsource RSC activities, OB3: Third party may provide sophisticated technology, system and better services, OB4: Reduces investment in RSC facilities, OB5: More flexibility, OB6: Better customer satisfaction responsiveness, OB7: Third party takes responsibility of entire product life cycle and OB8: Third party may have better environment management. There is a general perception that all these benefits directly or indirectly led to the economic performance of RSC. Recently, Govindan *et al.* (2013) developed a model for evaluating the economic, environmental and social performance of a supplier. In the study, they reported that an organization could assure sustainability even if they are outsourcing RSC activities. The research work examined whether outsourcing contributed to the economic, environmental and social performance of RSC. The hypotheses are developed to test whether the benefits are positively related with economic, environmental and social performance of RSC. The responses were gathered through the questionnaire developed, based on the theoretical background of hypothesis. The hypotheses developed are as follows.

H1a. Benefits of outsourcing RSC are positively related with economic performance of RSC.

H1b. Benefits of outsourcing RSC are positively related with environmental performance of RSC.

H1c. Benefits of outsourcing RSC are positively related with social performance of RSC.

3. Methodology

The intent of the research is to analyze the relation between outsourcing and TBL performance of the RSC. The constructs were developed and their observed indicators were identified with the help of past literature review. The questionnaire is attached in the Appendix. The data were collected through survey of Indian electronics industry by using Google Form and also by visiting personally to the various firms. According to Ryan and Garland (1999), all scales such as five-point, seven-point or ten-point are comparable for the analytical tools such as structural equation modeling (SEM). However, the five-point scale was utilized for the convenience of respondents, and it was used by many researchers in the past. The five-point Likert scale was utilized, and responders were asked to opt for the appropriate option based on their opinion. The data were transferred into the MS excel sheet for all the indicators. XLSTAT software was used for further statistical analysis. The data entered was thoroughly checked for any discrepancy, and appropriate modifications were made to complete the MS excel data sheet for transferring into the software. In order to test the hypotheses, s SEM has been utilized because of its ability to analyze measurement model, factor analysis and multi-linear regression modeling all together (Gefen *et al.*, 2000). The SEM is a second-generation multivariate data analysis technique employed to assess the factors in the analysis phase. Structural model shows the path relationship between constructs using latent variables along with observed variables (Kaplan, 2008). SEM offers two approaches; one is covariance-based AMOS and other one is PLSPM approach. In AMOS, numbers of observed indicators cannot be less than minimum required number, while in PLSPM, one observed indicator can be also analyzed efficiently (Peng and Lai, 2012). The proposed study used PLSPM to analyze the relation among constructs because of exploratory nature of the study (Leimeister *et al.*, 2009). It is utilized because it does not require the data to be normally distributed, and consequently a small data sample size may be used for the study (Reinartz *et al.*, 2009). PLSPM was developed by Herman Wold in early 1980s. It is a multi-linear regression analysis method consisting of factor analysis feature for analyzing relationship between constructs and their observed indicators. SEM requires analysis of measurement

model first, and after that SEM models are analyzed. XLSTAT software analyzes the measurement model and SEM together and provides the result both in graphical and tabular forms. Measurement model ensures the fitness of data for the further analysis by testing the validity and reliability of the constructs using observed indicators (Chin, 1998). The study evaluated internal consistency reliability, indicator reliability, individual reliability of the item, the internal consistency and the convergent and discriminant validity. After analyzing the measurement model, PLSPM is utilized for SEM to test the hypotheses developed for examining the relation between constructs. The data collection utilized for the study is explained in following subsection.

3.1 Data collection

To analyze the relation of outsourcing and TBL performance, a questionnaire was framed with the help of theoretical background and hypotheses. In order to collect the data, responses were recorded on a five-point Likert scale. The four experts from the academics and five from Indian electronics industry revised the initial draft of questionnaire for improving the questionnaire. Google Form was used for collecting the data. A five-point Likert scale with level of agreement of respondent as 5 for Very High, 4 for High, 3 for Medium, 2 for Low and 1 for Very Low; was utilized for the research. The questionnaire was self-administered and was managed online using Google Form. The questionnaire link was emailed to the 700 industry experts, and 253 responses were received. Some of them were incomplete so discarded, and effectively 208 responses were utilized for the research. The study ensures the participation of respondents from various level of management performing different functions including logistics department, supply chain department, marketing and sales, corporate strategy and human resource departments. No response bias was ensured before using the data for the analysis.

3.2 Nonresponse bias

As discussed in the previous section, effectively 208 responses were received. A total of 120 responses were received in first month of study. The remainder responses were received after sending the reminder emails to the responders. The nonresponse bias was tested by comparing these two groups and nonrespondents of sample size 120 respondents (Dubey *et al.*, 2019b). A *t*-distribution test was performed at 95% confidence level. The *p*-value obtained was greater than 0.5, which concludes that there is no response bias.

3.3 Common method bias

Common method biases were examined with the help of statistical analysis and correlation marker technique, and it was found that common method variance is nonsubstantial.

4. Results and discussion

In RSC, many organizations are dependent on the outsourcing of activities to other service providers. The outsourcing decisions have been focused mostly on the economic performance. The study focuses on the all three aspects of performance based on the TBL concept. As discussed above, hypotheses were developed for testing and investigation. The steps being followed are discussed in subsequent sections.

4.1 Measurement model

The measurement models were analyzed before the structural model analysis by using PLSPM. The measurement model was developed for different constructs based on hypotheses and was tested for the fitness of data. The results obtained using software are

shown in the tabular form in Table 1. Individual reliability, composite reliability, convergent validity and discriminate validity met the criteria as per standards described in the literature frequently. The values of different parameters are shown in Table 1 and Table 2.

These findings fall within the acceptable limit prescribed by the researchers. The evidence of construct validity represent the adequate fit of measurement model (Iacobucci, 2010) and indicate for evaluating second part of the analysis that is SEM analysis.

4.2 Structural equation model

All the proposed hypotheses are tested by using PLSPM technique of SEM. The XLSTAT software was utilized, and the parameters of measurements are summarized in Table 3 and Figure 1. The values of R^2, Q^2, f^2 and q^2 values are within limit as per requirements of measurement model. The *t*-value for hypothesis H1a is 9.087, from Table 3, which is higher than the required *t*-value at 95% of confidence level. Therefore, the hypothesis "Benefits of outsourcing RSC are positively related with economic performance of RSC" is accepted. This implies that outsourcing helps in economic benefits to the organizations. Similarly, H1b and H1c are also accepted, as shown in Table 3. These results indicate that outsourcing benefits are positively related with all three respective components of TBL performance of RSC. Bai and Sarkis (2019) observed that economic and cost benefits are not sufficient for the organizations seeking sustainable practices. According to Tavana *et al.* (2016), third-party

Table 1. Statistics results of measurement model

Latent variable (Cronbach's alpha)	Loadings	AVE	Cr
OBS: outsourcing benefits (Cronbach's alpha = 0.79)		0.576	0.901
OB1: Focus on core activity of the organization	0.776		
OB2: Economical to outsource RSC activities	0.620		
OB3: Third party may provide sophisticated technology, system and better services	0.830		
OB4: Reduces investment in RSC facilities	0.865		
OB5: More flexibility	0.783		
OB6: Better customer satisfaction responsiveness	0.681		
OB7: Third party takes responsibility of entire product life cycle	0.837		
OB8: Third party may have better environment management	0.761		
ECP: economic performance (Cronbach's alpha = 0.85)		0.743	0.945
EC1: Return on investment	0.923		
EC2: Recapturing value	0.940		
EC3: Logistics cost optimization	0.767		
EC4: Recycle efficiency	0.838		
EC5: Annual sales of remanufactured products	0.713		
EC6: Disposal costs	0.904		
EVP: environmental performance (Cronbach's alpha = 0.91)		0.662	0.919
EV1: Minimum energy consumption	0.813		
EV2: Optimum use of raw material	0.852		
EV3: Transport optimization	0.812		
EV4: Reduced packaging	0.829		
EV5: Use of recycled material	0.732		
EV6: Waste reduction	0.816		
SCP: social performance (Cronbach's alpha = 0.83)		0.613	0.889
SC1: Community complaints	0.854		
SC2: Customer health and safety	0.847		
SC3: Stake holders participation	0.734		
SC4: Employment stability	0.629		
SC5: Donations to community	0.842		
SC6: Employee benefits	0.789		

service provider's selection is critical for the economic and environmental performance of an organization. Govindan *et al.* (2013) also found that outsourcing activities may also help in improving TBL performance and developed a model for supplier selection based on TBL for Indian industry. Recently, Zarbakhshnia *et al.* (2020a, b) also supports the findings of the research that reverse logistics contribute to TBL, i.e. three major components of the sustainability.

5. Conclusions, implications and future scope

The objective of proposed study is to empirically investigate the relationship between the outsourcing benefits and TBL based performance of RSCs. The literature review was carried out based on theoretical background, and three hypotheses were developed. The

Table 2. Constructs correlation and AVE for Model 3

Constructs	AVE	OBS	ECP	EVP	SCP
OBS	0.576	*0.759*	0	0	0
ECP	0.743	0.579	*0.862*	0	0
EVP	0.662	0.441	0.487	*0.814*	0
SCP	0.613	0.683	0.439	0.697	*0.783*

Table 3. Structural equation model results

Hypotheses	Relationship	Path coefficient	*t*-value	f^2 effect size	q^2 effect size	Results
H1a	OBS–ECP	0.535	9.087	0.439	0.265	Accepted
H1b	OBS–EVP	0.667	12.84	0.464	0.292	Accepted
H1c	OBS–SCP	0.531	9.004	0.432	0.294	Accepted

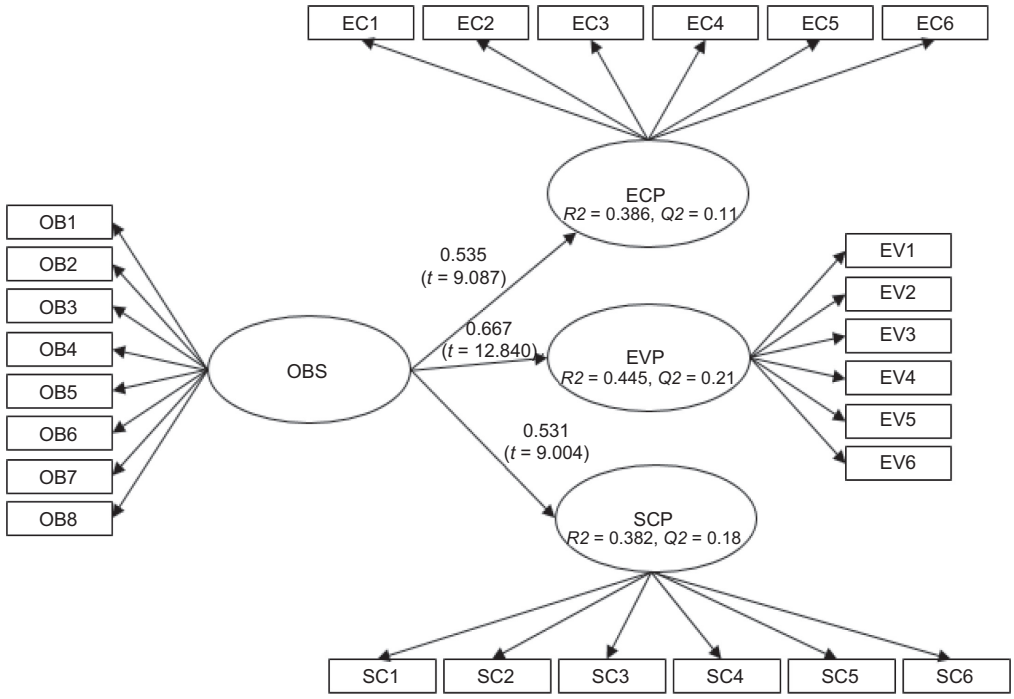


Figure 1. Structural equation model for outsourcing benefits

questionnaire was framed to carry out survey of the industry. Responses were collected, and measurement model was analyzed to ensure fitness of data for further analysis. PLSPM was utilized for structural modeling of the various constructs formed for testing the hypotheses. Results of study indicate that a lot of importance is being given to the RSC, which is evident from the responses of industry. The study inspects the effect of outsourcing decisions on the TBL performance of RSC.

5.1 Implications

The study tested various outsourcing benefits and found that these benefits are related with TBL performance of RSC. The findings of the study will help the managers, researchers and academicians in decision-making in context to RSC outsourcing issues. Managers may utilize this analysis for justifying the use of corporate social responsibility funds in outsourcing activities.

5.2 Limitations and future scope

The study provides a useful insight into the Indian electronics industry. Based on survey of Indian electronics industry, association and contribution of RSC with TBL performance is examined and ascertained. However, the present study is restricted to only TBL performance of RSC. In future, more performance aspects may be included for the study. In addition to this, the study was limited to small sample size of the electronics industry. In future, a larger sample size may be taken in other sectors also for comparison of findings and generalization of results. Apart from it, performance parameters related to circular economy may also be considered for analyzing the impact of outsourcing benefits on other sustainability parameters.

References

Agrawal, S. and Singh, R.K. (2019), "Analyzing disposition decisions for sustainable reverse logistics: Triple Bottom Line approach", *Resources, Conservation and Recycling*, Vol. 150, p. 104448.

Agrawal, S., Singh, R.K. and Murtaza, Q. (2015), "A literature review and perspectives in reverse logistics", *Resources, Conservation and Recycling*, Vol. 97, pp. 76-92.

Agrawal, S., Singh, R.K. and Murtaza, Q. (2016), "Triple bottom line performance evaluation of reverse logistics", *Competitiveness Review*, Vol. 26 No. 3, pp. 289-310.

Akbari, M. (2018), "Logistics outsourcing: a structured literature review", *Benchmarking: An International Journal*, Vol. 25 No. 5, pp. 1548-1580.

Ameknassi, L., Aït-Kadi, D. and Rezg, N. (2016), "Integration of logistics outsourcing decisions in a green supply chain design: a stochastic multi-objective multi-period multi-product programming model", *International Journal of Production Economics*, Vol. 182, pp. 165-184.

Ardakani, A., Davood and Soltanmohammadi, A. (2019), "Investigating and analysing the factors affecting the development of sustainable supply chain model in the industrial sectors", *Corporate Social Responsibility and Environmental Management*, Vol. 26 No. 1, pp. 199-212.

Arnold, U. (2000), "New dimensions of outsourcing: a combination of transaction cost economics and the core competencies concept", *European Journal of Purchasing and Supply Management*, Vol. 6 No. 1, pp. 23-29.

Bai, C. and Sarkis, J. (2018), "Integrating sustainability into supplier selection: a grey-based TOPSIS analysis", *Technological and Economic Development of Economy*, Vol. 24 No. 6, pp. 2202-2224.

Bai, C. and Sarkis, J. (2019), "Integrating and extending data and decision tools for Sustainable third party reverse logistics provider selection", *Computers and Operations Research*, Vol. 110, pp. 188-207.

Bansal, P. (2002), "The corporate challenges of sustainable development", *The Academy of Management Executive*, Vol. 16 No. 2, pp. 122-131.

Batarfi, R., Jaber, M.Y. and Aljazzar, S.M. (2017), "A profit maximization for a reverse logistics dual-channel supply chain with a return policy", *Computers and Industrial Engineering*, Vol. 106, pp. 58-82.

Bernon, M. and Cullen, J. (2007), "An integrated approach to managing reverse logistics", *International Journal of Logistics: Research and Applications*, Vol. 10 No. 1, pp. 41-56.

Boyson, S., Corsi, T., Dresner, M. and Rabinovich, E. (1999), "Managing effective third party logistics relationships: what does it take?", *Journal of Business Logistics*, Vol. 20 No. 1, p. 73.

Chakraborty, K., Mondal, S. and Mukherjee, K. (2018), "Developing a causal model to evaluate the critical issues in reverse supply chain implementation", *Benchmarking: AnInternational Journal*, Vol. 25 No. 7, pp. 1992-2017.

Chakraborty, S., Sharma, A. and Vaidya, O.S. (2020), "Achieving sustainable operational excellence through IT implementation in Indian logistics sector: an analysis of barriers", *Resources, Conservation and Recycling*, Vol. 152, p. 104506.

Chand, M., Bhatia, N. and Singh, R.K. (2018), "ANP-MOORA-based approach for the analysis of selected issues of green supply chain management", *Benchmarking: An International Journal*, Vol. 25 No. 2, pp. 642-659.

Chandak, A., Chandak, S. and Dalpati, A. (2019), "The analysis of impact of critical dimensions on performance in context of supply chain: evidence from Indian automobile industry", Proceedings of Recent Advances in Interdisciplinary Trends in Engineering & Applications (RAITEA) 2019, doi: 10.2139/ssrn.3350929, (accessed 12 March 2019).

Cheng, Y. and Lee, F. (2010), "Industrial marketing management outsourcing reverse logistics of high-tech manufacturing firms by using a systematic decision-making approach: TFT-LCD sector in Taiwan", *Industrial Marketing Management*, Vol. 39 No. 7, pp. 1111-1119.

Chileshe, N., Jayasinghe, R.S. and Rameezdeen, R. (2019), "Information flow-centric approach for reverse logistics supply chains", *Automation in Construction*, Vol. 106, p. 102858.

Chin, W.W. (1998), "The partial least squares approach to structural equation modeling", *Modern methods for business research*, Vol. 295 No. 2, pp. 295-336.

Daugherty, P.J. and Dröge, C. (1997), "Organizational structure in divisionalized manufacturers: the potential for outsourcing logistical services", *International Journal of Physical Distribution and Logistics Management*, Vol. 27 No. 5, pp. 337-349.

Dubey, R., Gunasekaran, A. and Papadopoulos, T. (2017), "Green supply chain management: theoretical framework and further research directions", *Benchmarking: An International Journal*, Vol. 24 No. 1, pp. 184-218.

Dubey, R., Gunasekaran, A., Childe, S.J., Papadopoulos, T. and Helo, P. (2019a), "Supplier relationship management for circular economy: influence of external pressures and top management commitment", *Management Decision*, Vol. 57 No. 4, pp. 767-790.

Dubey, R., Gunasekaran, A., Childe, S.J., Blome, C. and Papadopoulos, T. (2019b), "Big data and predictive analytics and manufacturing performance: integrating institutional theory, resource-based view and big data culture", *British Journal of Management*, Vol. 30 No. 2, pp. 341-361.

Elkington, J. (1998), "Partnerships from cannibals with forks: The triple bottom line of 21st-century business", *Environmental Quality Management*, Vol. 8 No. 1, pp. 37-51.

Espino-Rodríguez, T.F., Chun-Lai, P. and Gil-Padilla, A.M. (2017), "Does outsourcing moderate the effects of asset specificity on performance? An application in Taiwanese hotels", *Journal of Hospitality and Tourism Management*, Vol. 31, pp. 13-27.

Evangelista, P., Santoro, L. and Thomas, A. (2018), "Environmental sustainability in third-party logistics service providers: a systematic literature review from 2000–2016", *Sustainability*, Vol. 10 No. 5, p. 1627.

Feng, M., Ford, M.W. and Evans, J.R. (2018), "An empirical investigation of the Baldrige framework using applicant scoring data", *International Journal of Quality and Reliability Management*, Vol. 35 No. 8, pp. 1599-1616.

Ferrer, G. and Ketzenberg, M.E. (2004), "Value of information in remanufacturing complex products", *IIE Transactions*, Vol. 36 No. 3, pp. 265-277.

Future Market Insights (2019), "Global Consumer Electronics Market: key players focusing on developing technologically advanced products", Mar 05, Press Release, available at: https://www.futuremarketinsights.com/press-release/consumer-electronics-market.

Gefen, D., Straub, D. and Boudreau, M.C. (2000), "Structural equation modeling and regression: guidelines for research practice", *Communications of the relation for information systems*, Vol. 4 No. 1, p. 7.

Giri, B.C., Chakraborty, A. and Maiti, T. (2017), "Pricing and return product collection decisions in a closed-loop supply chain with dual-channel in both forward and reverse logistics", *Journal of Manufacturing Systems*, Vol. 42, pp. 104-123.

Golroudbary, S., Zahraee, S., Awan, U. and Kraslawski, A. (2019), "Sustainable operations management in logistics using simulations and modelling: a framework for decision making in delivery management", *Procedia Manufacturing*, Vol. 30, pp. 627-634.

Gopal, P.R.C. and Thakkar, J. (2016), "Sustainable supply chain practices: an empirical investigation on Indian automobile industry", *Production Planning and Control*, Vol. 27 No. 1, pp. 49-64.

Gorane, S.J. and Kant, R. (2016), "Supply chain practices: an implementation status in Indian manufacturing organisations", *Benchmarking-An International Journal*, Vol. 23 No. 5, pp. 1076-1110.

Govindan, K. and Bouzon, M. (2018), "From a literature review to a multi-perspective framework for reverse logistics barriers and drivers", *Journal of Clean Production*, Vol. 187, pp. 318-337.

Govindan, K. and Murugesan, P. (2011), "Selection of third-party reverse logistics provider using fuzzy extent analysis", *Benchmarking: An International Journal*, Vol. 18 No. 1, pp. 149-167.

Govindan, K., Kadzinski, M., Ehling, R. and Miebs, G. (2019), "Selection of a sustainable third-party reverse logistics provider based on the robustness analysis of an outranking graph kernel conducted with ELECTRE I and SMAA", *Omega*, Vol. 85, pp. 1-15.

Govindan, K., Khodaverdi, R. and Jafarian, A. (2013), "A fuzzy multi criteria approach for measuring sustainability performance of a supplier based on triple bottom line approach", *Journal of Cleaner Production*, Vol. 47, pp. 345-354.

Govindan, K., Palaniappan, M., Zhu, Q. and Kannan, D. (2012), "Analysis of third-party reverse logistics provider using interpretive structural modeling", *International Journal of Production Economics*, Vol. 140 No. 1, pp. 204-211.

Gu, W., Wei, L., Zhang, W. and Yan, X. (2019), "Evolutionary game analysis of cooperation between natural resource-and energy-intensive companies in reverse logistics operations", *International Journal of Production Economics*, Vol. 218, pp. 159-169.

Guarnieri, P., Sobreiro, V.A., Nagano, M.S. and Serrano, A.L.M. (2015), "The challenge of selecting and evaluating third-party reverse logistics providers in a multicriteria perspective: a Brazilian case", *Journal of Cleaner Production*, Vol. 96, pp. 209-219.

Gunasekaran, A. and Spalanzani, A. (2012), "Sustainability of manufacturing and services: investigations for research and applications", *International Journal of Production Economics*, Vol. 144 No. 1, pp. 35-47.

Hair, J.F. Jr, Matthews, L.M., Matthews, R.L. and Sarstedt, M. (2017), "PLS-SEM or CB-SEM: updated guidelines on which method to use", *International Journal of Multivariate Data Analysis*, Vol. 1 No. 2, pp. 107-123.

Hamel, G. and Prahalad, C.K. (1990), "The core competence of the corporation", *Harvard Business Review*, Vol. 68 No. 1, pp. 79-91.

Han, H. and Trimi, S. (2018), "A fuzzy TOPSIS method for performance evaluation of reverse logistics in social commerce platforms", *Expert Systems with Applications*, Vol. 103, pp. 133-145.

Hazen, B.T. (2011), "Strategic reverse logistics disposition decisions: from theory to practice", *International Journal of Logistics Systems and Management*, Vol. 10 No. 3, pp. 275-292.

Hubbard, G. (2009), "Measuring organizational performance: beyond the triple bottom line", *Business Strategy and the Environment*, Vol. 18 No. 3, pp. 177-191.

Huscroft, J.R., Hazen, B.T., Hall, D.J., Skipper, J.B. and Hanna, J.B. (2013), "Reverse logistics: past research, current management issues, and future directions", *International Journal of Logistics Management*, Vol. 24 No. 3, pp. 304-327.

Iacobucci, D. (2010), "Structural equations modeling: fit indices, sample size, and advanced topics", *Journal of Consumer Psychology*, Vol. 20 No. 1, pp. 90-98.

Insigna, R.C. and Werle, M.J. (2000), "Linking outsourcing to business strategy", *The Academy of Management Executive*, Vol. 14 No. 4, pp. 58-70.

Ji, P., Zhang, H.Y. and Wang, J.Q. (2018), "Selecting an outsourcing provider based on the combined MABAC–ELECTRE method using single-valued neutrosophic linguistic sets", *Computers and Industrial Engineering*, Vol. 120, pp. 429-441.

Kakabadse, A. and Kakabadse, N. (2005), "Outsourcing: current and future trends", *Thunderbird International Business Review*, Vol. 47 No. 2, pp. 183-204.

Kaplan, D. (2008), *Structural Equation Modeling: Foundations and Extensions: Foundations and Extensions*, Sage Publications, Vol. 10.

Ko, H.J. and Evans, G.W. (2007), "A genetic algorithm-based heuristic for the dynamic integrated forward/reverse logistics network for 3PLs", *Computers and Operations Research*, Vol. 34 No. 2, pp. 346-366.

Kotabe, M. and Zhao, H. (2002), "A taxonomy of sourcing strategic types for MNCs operating in China", *Asia Pacific Journal of Management*, Vol. 19 No. 1, pp. 11-27.

Krumwiede, D.W. and Sheu, C. (2002), "A model for reverse logistics entry by third-party providers", *Omega*, Vol. 30 No. 5, pp. 325-333.

Kumar, R. and Singh, R.K. (2017), "Coordination and responsiveness issues in SMEs supply chains: a review", *Benchmarking: An International Journal*, Vol. 24 No. 3, pp. 635-650.

Kumari, S., Singh, A., Mishra, N. and Garza-Reyes, J.A. (2015), "A multi-agent architecture for outsourcing SMEs manufacturing supply chain", *Robotics and Computer-Integrated Manufacturing*, Vol. 35 No. 1, pp. 36-44.

Kushwaha, S., Ghosh, A. and Rao, A.K. (2020), "Collection activity channels selection in a reverse supply chain under a carbon cap-and-trade regulation", *Journal of Cleaner Production*, Vol. 260, p. 121034.

Lacity, M. and Hirschheim, R. (1993), *Information Systems Outsourcing; Myths, Metaphors, and Realities*, John Wiley & Sons, New York.

Lambert, S., Riopel, D. and Abdul-Kader, W. (2011), "A reverse logistics decisions conceptual framework", *Computers and Industrial Engineering*, Vol. 61 No. 3, pp. 561-581.

Lee, C.K.M. and Lam, J.S.L. (2012), "Managing reverse logistics to enhance sustainability of industrial marketing", *Industrial Marketing Management*, Vol. 41 No. 4, pp. 589-598.

Leimeister, S., Leimeister, J.M., Knebel, U. and Krcmar, H. (2009), "A cross-national comparison of perceived strategic importance of RFID for CIOs in Germany and Italy", *International Journal of Information Management*, Vol. 29 No. 1, pp. 37-47.

Li, Y.L., Ying, C.S., Chin, K.S., Yang, H.T. and Xu, J. (2018a), "Third-party reverse logistics provider selection approach based on hybrid-information MCDM and cumulative prospect theory", *Journal of Cleaner Production*, Vol. 195, pp. 573-584.

Li, Y., Kannan, D., Garg, K., Gupta, S., Gandhi, K. and Jha, P.C. (2018b), "Business orientation policy and process analysis evaluation for establishing third party providers of reverse logistics services", *Journal of Cleaner Production*, Vol. 182, pp. 1033-1047.

Logozar, K. (2008), "Outsourcing reverse logistics", *Zagreb International Review of Economics and Business*, Vol. 11 No. 2, pp. 35-45.

McKenzie, S. (2004), *Social Sustainability: Towards Some Definitions*, Hawke Research Institute, University of South Australia Magill.

Mello, J.E., Stank, T.P. and Esper, T.L. (2008), "A model of logistics outsourcing strategy", *Transportation Journal*, Vol. 47 No. 4, p. 5.

Min, H. and Ko, H.J. (2008), "The dynamic design of a reverse logistics network from the perspective of third-party logistics service providers", *International Journal of Production Economics*, Vol. 113 No. 1, pp. 176-192.

Modgil, S. and Sonwaney, V. (2019), "Planning the application of blockchain technology in identification of counterfeit products: sectorial prioritization", *IFAC-PapersOnLine*, Vol. 52 No. 13, pp. 1-5.

Morgan, T.R., Tokman, M., Richey, R.G. and Defee, C. (2018), "Resource commitment and sustainability: a reverse logistics performance process model", *International Journal of Physical Distribution and Logistics Management*, Vol. 48 No. 2, pp. 164-182.

Ordoobadi, S.M. (2009), "Outsourcing reverse logistics and remanufacturing functions: a conceptual strategic model", *Management Research News*.

Östlin, J., Sundin, E. and Björkman, M. (2008), "Importance of closed-loop supply chain relationships for product remanufacturing", *International Journal of Production Economics*, Vol. 115 No. 2, pp. 336-348.

Pedram, A., Yusoff, N.B., Udoncy, O.E., Mahat, A.B., Pedram, P. and Babalola, A. (2017), "Integrated forward and reverse supply chain: a tire case study", *Waste Management*, Vol. 60, pp. 460-470.

Peng, D.X. and Lai, F. (2012), "Using partial least squares in operations management research: a practical guideline and summary of past research", *Journal of Operations Management*, Vol. 30 No. 6, pp. 467-480.

Prajapati, H., Kant, R. and Shankar, R. (2019), "Prioritizing the solutions of reverse logistics implementation to mitigate its barriers: a hybrid modified SWARA and WASPAS approach", *Journal of Cleaner Production*, Vol. 240, p. 118219.

Presley, A., Meade, L. and Sarkis, J. (2007), "A strategic sustainability justification methodology for organizational decisions: a reverse logistics illustration", *International Journal of Production Research*, Vol. 45 Nos 18-19, pp. 4595-4620.

Quinn, J.B. and Hilmer, F.G. (1994), "Strategic outsourcing", *MIT Sloan Management Review*, Vol. 35 No. 4, p. 43.

Ravi, V. and Shankar, R. (2005), "Analysis of interactions among the barriers of reverse logistics", *Technological Forecasting and Social Change*, Vol. 72 No. 8, pp. 1011-1029.

Ravi, V. and Shankar, R. (2015), "Survey of reverse logistics practices in manufacturing industries: an Indian context", *Benchmarking: An International Journal*, Vol. 22 No. 5, pp. 874-899.

Reinartz, W., Haenlein, M. and Henseler, J. (2009), "An empirical comparison of the efficacy of covariance-based and variance-based SEM", *International Journal of Research in Marketing*, Vol. 216 No. 4, pp. 332-344.

Rogers, D.S. and Tibben-Lembke, R.S. (1999), *Going Backwards: Reverse Logistics Trends and Practices*, Vol. 2, Reverse Logistics Executive Council, Pittsburgh, PA.

Ryan, C. and Garland, R. (1999), "The use of a specific non-response option on Likert-type scales", *Tourism Management*, Vol. 20 No. 1, pp. 107-113.

Sarkis, J., Helms, M.M. and Hervani, A.A. (2010), "Reverse logistics and social sustainability", *Corporate Social Responsibility and Environmental Management*, Vol. 17 No. 6, pp. 337-354.

Senthil, S., Murugananthan, K. and Ramesh, A. (2018), "Analysis and prioritization of risks in a reverse logistics network using hybrid multi criteria decision making methods", *Journal of Cleaner Production*, Vol. 179, pp. 716-730.

Serrato, M.A. and Ryan, S.M. (2007), "A Markov decision model to evaluate outsourcing in reverse logistics", *International Journal of Production Research*, Vol. 45 Nos 18-19, pp. 4289-4315.

Singh, R.K. and Agrawal, S. (2018), "Analyzing disposition strategies in reverse supply chains: fuzzy TOPSIS approach", *Management of Environmental Quality: An International Journal*, Vol. 29 No. 3, pp. 427-443.

Singh, R.K., Garg, S.K. and Deshmukh, S.G. (2007), "Strategy development for competitiveness: a study on Indian auto component sector", *International Journal of Productivity and Performance Management*, Vol. 56 No. 4, pp. 285-304.

Singh, R.K., Gunasekaran, A. and Kumar, P. (2018), "Third party logistics (3PL) selection for cold chain management: a fuzzy AHP and fuzzy TOPSIS approach", *Annals of Operations Research*, Vol. 267 Nos 1-2, pp. 531-553.

Tadelis, S. (2007), "The innovative organization: creating value through outsourcing", *California Management Review*, Vol. 50 No. 1, pp. 261-277.

Tavana, M., Zareinejad, M., Di Caprio, D. and Kaviani, M.A. (2016), "An integrated intuitionistic fuzzy AHP and SWOT method for outsourcing reverse logistics", *Applied Soft Computing*, Vol. 40 No. 1, pp. 544-557.

Thanki, S. and Thakkar, J. (2018), "A quantitative framework for lean and green assessment of supply chain performance", *International Journal of Productivity and Performance Management*, Vol. 67 No. 2, pp. 366-400.

The Related Chambers of Commerce and Industry of India (2019), "Second-hand market to cross Rs 1,15,000 crore by 2015: ASSOCHAM Economic slowdown creates higher demand for second-hand goods and machinery", available at: http://assocham.org/newsdetail.php?id=4351.

Tian, G., Liu, X., Zhang, M., Yang, Y., Zhang, H., Lin, Y., Ma, F., Wang, X., Qu, T. and Li, Z. (2019), "Selection of take-back pattern of vehicle reverse logistics in China via Grey-DEMATEL and Fuzzy-VIKOR combined method", *Journal of Cleaner Production*, Vol. 220, pp. 1088-110.

Tibben-Lembke, R.S. and Rogers, D.S. (2002), "Differences between forward and reverse logistics in a retail environment", *Supply Chain Management: International Journal*, Vol. 7 No. 5, pp. 271-282.

Vachon, S. and Mao, Z. (2008), "Linking supply chain strength to sustainable development: a country-level analysis", *Journal of Cleaner Production*, Vol. 16 No. 15, pp. 1552-1560.

Wu, F., Li, H.Z., Chu, L.K. and Sculli, D. (2005), "An outsourcing decision model for sustaining long-term performance", *International Journal of Production Research*, Vol. 43 No. 12, pp. 2513-2535.

Yu, H. and Solvang, W.D. (2018), "Incorporating flexible capacity in the planning of a multi-product multi-echelon sustainable reverse logistics network under uncertainty", *Journal of Cleaner Production*, Vol. 198, pp. 285-303.

Zarbakhshnia, N., Soleimani, H. and Ghaderi, H. (2018), "Sustainable third-party reverse logistics provider evaluation and selection using fuzzy SWARA and developed fuzzy COPRAS in the presence of risk criteria", *Applied Soft Computing*, Vol. 65, pp. 307-319.

Zarbakhshnia, N., Wu, Y., Govindan, K. and Soleimani, H. (2020a), "A novel hybrid multiple attribute decision-making approach for outsourcing sustainable reverse logistics viable bearable economic equitable social", *Journal of Cleaner Production*, Vol. 242, p. 118461.

Zarbakhshnia, N., Wu, Y., Govindan, K. and Soleimani, H. (2020b), "A novel hybrid multiple attribute decision-making approach for outsourcing sustainable reverse logistics", *Journal of Cleaner Production*, Vol. 242, p. 118461.

Zhikang, L. (2017), "Research on development strategy of automobile reverse logistics based on SWOT analysis", *Procedia engineering*, Vol. 174, pp. 324-330.

Further reading

Chin, W.W., Marcolin, B.L. and Newsted, P.R. (2003), "A partial least squares latent variable modeling approach for measuring interaction effects: results from a Monte Carlo simulation study and an electronic-mail emotion/adoption study", *Information Systems Research*, Vol. 14 No. 2, pp. 189-217.

Gunasekaran, A., Irani, Z., Choy, K.L., Filippi, L. and Papadopoulos, T. (2015), "Performance measures and metrics in outsourcing decisions: a review for research and applications", *International Journal of Production Economics*, Vol. 161, pp. 153-166.

Prakash, C. and Barua, M.K. (2016), "A combined M CDM approach for evaluation and selection of third-party reverse logistics partner for Indian electronics industry", *Sustainable Production and Consumption*, Vol. 7, pp. 66-78.

Skinner, L.R., Bryant, P.T. and Glenn Richey, R. (2008), "Examining the impact of reverse logistics disposition strategies", *International Journal of Physical Distribution and Logistics Management*, Vol. 38 No. 7, pp. 518-539.

Yang, H.L. and Wang, C.S. (2007), "Integrated framework for reverse logistics", *New Trends in Applied Artificial Intelligence*, Springer Berlin Heidelberg, Vol. 45, pp. 501-510.

Appendix

Name of the Organization …….Year of Establishment……..

Annual Turnover (Rupees) …… Number of Employees _____

Please rate the following factors for your organization on the scale of 1 to 5 (1 for Very Low, 2 for Low, 3 for Medium, 4 for High, and 5 for Very High)

Questionnaire	**Please mark any one of the following for each question**				
Outsourcing Benefits	**1**	**2**	**3**	**4**	**5**
Focus on core activity of the organization					
Economical to outsource RSC activities					
Third party may provide sophisticated technology, system, and better services					
Reduces investment in RSC facilities	**1**	**2**	**3**	**4**	**5**
More flexibility					
Better customer satisfaction responsiveness					
Third Party takes responsibility of entire product life cycle					
Third party may have better environment management					
Economic Performance	**1**	**2**	**3**	**4**	**5**
Return On Investment					
Recapturing Value					
Logistics Cost Optimization					
Recycle Efficiency					
Annual Sales of re-manufactured products					
Disposal Costs					
Environmental Performance	**1**	**2**	**3**	**4**	**5**
Minimum Energy Consumption					
Optimum use of raw material					
Transport Optimization					
Reduced Packaging					
Use of recycled material					
Waste Reduction					
Social Performance	**1**	**2**	**3**	**4**	**5**
Response to Community complaints					
Customer Health and Safety					
Stake Holders Participation					
Employment Stability					
Donations to Community					
Employee Benefits					

Respondent Profile

Please provide the following information. All the personal information provided by you would be confidential and will not be shared with anyone. I appreciate your time. Thanks

(i) Name (Optional)

(ii) Designation

(iii) Total Industry Experience

(iv) Your Expertize

About the authors

Dr. Saurabh Agrawal is Assistant Professor in Mechanical, Production and Industrial Engineering Department at Delhi Technological University, Delhi, India. He has vast experience of academics, research and the industry both in India and in the USA. His research focus is in the areas of supply chain management, reverse logistics, sustainability and e-waste management. He has completed his undergraduate degree from Indian Institute of Technology, Roorkee, India. He has received his Master of Business Administration from Oregon State University, USA and Master in Technology from Indian Institute of Technology, Delhi, India. He has published research papers in Resources, Conservation and Recycling, Journal of Cleaner Production, Competitiveness Review, etc.

Dr. Rajesh Kumar Singh is Professor in Operations Management Area at Management Development Institute (MDI), Gurgaon, India. Currently, he is serving as an Associate Editor to International Journal of Consumers Studies (Wiley publication), International Journal of Global Business and Competitiveness (Springer) and in Editorial Board of many reputed journals. His areas of interest include supply chain management, Industry 4.0 and circular economy. He has about 180 research papers published in reputed international/national journals and conferences. He has published papers in journals such as Production Planning and Control, Annals of Operations Research, International Journal of Production Research, Resources, Conservation and Recycling, Journal of Cleaner Production, Industrial Management and Data Systems, Singapore Management Review, International Journal of Productivity and Performance Management, Benchmarking: An International Journal, Journal of Modelling in Management, Competitiveness Review: An International Business Journal, Global Journal of Flexible Systems and Management, International Journals of Productivity and Quality Management, International Journal of Services and Operations Management, International Journal of Automotive Industry and Management, International Journal of Logistic Systems Management and IIMB Management Review and Productivity Promotion. Rajesh Kumar Singh is the corresponding author and can be contacted at: rajesh.singh@mdi.ac.in

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:12/December -2020 Impact Factor- 5.354 www.irjmets.com

RECENT ADVANCEMENTS IN SOLAR FURNACE TECHNOLOGY

Sameer Kapur*1, Satyam Shandilya*2, Vishal Tiwari*3, J.P. Kesari*4

*1,2,3B. Tech Student (4th Year), Department of Mechanical Engineering, Delhi Technological University, Delhi – 110042, India.

*4Associate Professor, Department of Mechanical Engineering, Delhi Technological University, Delhi – 110042, India.

ABSTRACT

As the pollution and climate change is a real concern, the whole world is looking for alternatives to the fossil fuels. This has led to an increase in dependance on non-conventional sources of energy like wind, hydro and solar. There has been increasing use of non-conventional sources of energy for achieving high temperatures. One such method is the use of Solar Furnace for reaching temperatures up to $3,500^0$ C. In this review paper, a brief overview of different types of solar furnace and their main components have been described. The various parameters affecting the performance, efficiency, and achievable temperature have also been discussed. Various solar furnaces operational around the globe are discussed in this paper such as Odeillo Furnace in France, Solar Tower in Jülich, Germany, solar furnace in Mexico and USA. The principle, working, and dimensional specification have been discussed along with the infrastructural placement. Various recent achievements and innovations in the field of solar furnace have also been presented.

Keywords: Concentrated Solar Power (CSP), Concentrator, Heliostat, Odeillo Furnace, Solar Furnace

I. INTRODUCTION

The rapid climate change has raised concerns regarding the rational use of conventional energy sources and development and use of non-conventional sources of energy. There has been extensive research in the field of energy, exploring sources of clean energy having minimum environmental impacts. Solar Energy, is by far, the greatest single source of clean energy which can be converted for use in various applications. Some of the ways developed for harnessing solar energy include passive and active solar heating, heliostat fields, solar furnace, et cetera.

Solar furnace is based on the principle of concentrating solar radiation or sunlight on a small area with the use of an optical system consisting of collector and heliostat, achieving high temperatures. The other name used for, Solar furnace, is solar energy concentrators due to its capability to produce high temperatures on a small area with concentrated radiations. Solar furnaces, fitted with large parabolic mirrors, the temperature attainable is close to 3,5000 Celsius. This is generally because of the use of a large number of strategically placed parabolic reflectors or heliostat onto a concentrator, leading to a large coverage of solar radiation field. The use of this method for harnessing solar power is gaining immense popularity with applications in wide industries.

The basic advantage of using a Solar furnace is the production of no pollutants, making it a clean source of energy. The application area includes solutions for heating water, space heating air conditioning, foundry applications, hydrogen fuel production, high-temperature material testing, and extensive research work. Some of the other methods capable of producing such high temperatures are exploding wires, detonations, induction heating, and many more. But all these methods, have possible drawbacks such as shortness of time, by-products produced and many more.

II. LITERATURE REVIEW

In the research paper, the authors discuss the temperature control architecture design and development methodology with analysis results and help of two experimental tests. Further, it has discussed the various parts used for the solar furnace model as well.

In the paper, a detailed study of heliostat calibration and tracking control methods is discussed with all the errors flawing the result and successful operation of the solar furnace. The various calibration systems have been discussed as well as classified.

The research paper provides a detailed project report of the construction, assembly and testing of a 5 KW solar furnace developed by students over time. The complete description of steps involved in the completion of the project have been discussed along with detailed results and analysis.

In this paper, author has described the optical structure design for a solar furnace. Monte Carlo simulations were done to gauge the influence of the optical factors on the concentrated solar heat flux distribution. Heliostat optimum tilt was studied and the errors associated with the measurement of these optical parameters effect were discussed. Results were drawn comparing the experimentally measured irradiance with the simulated results.

In this paper, the author has discussed the working of Odeillo Furnace and its parts along with their dimensions. The dynamic heliostat mounting mechanism has been discussed. An experiment to study the heat flux distribution along with graphical outcomes is also discussed.

The author has discussed the Mexican high radiative flux solar furnace in this paper. The parts along with their dimensions have also been discussed. The two tests that were initiated with the operation of HRFSF, the acquisition of images on a Lambertian target and the melting of metals is covered and the experimental setup is briefly discussed. The results of the two tests have been discussed along with the relevant graphs.

In this paper, Mexican solar furnace along with its infrastructural properties is discussed. The hexagonal mirrors for the spherical frame have been discussed along with spherical aberration and reorientation facets. The heat flux distribution has been studied along with all the irradiance, receiver and collector graphs.

The author has discussed the SERI's high flux solar furnace. Its unique design has been discussed along with dimensional parameters. Single stage and two stage performance of the furnace has been discussed. Accentuator, computer hardware, computer software and data acquisition software parameters have been discussed. The heat flux distribution has been discussed along with relevant graph plots.

In this research paper, various case studies of solar furnace around the world is discussed, their parts and specifications. Concentrated Solar Power generated cooking system for the midday meal system in Delhi is considered and a theoretical analysis is done for the load calculation and economic feasibility.

The author discusses about the transient simulation solar tower concept with open volumetric technology. The open volumetric technology receives air as heat transfer fluid and it is used to convert electricity by using a gas turbine. This model is being developed for a location Barstow-Daggett, USA. Simulation is performed in MATLAB and its performance has been studied.

The author has described the modeling of a high temperature storage system with open volumetric receiver technology, in which air is used as heat transfer medium. Thermal energy can be stored and release as per the requirement via HTF air. The storage model was developed in MATLAB.

In this paper, the author has discussed the brief history of solar furnace and the various experiments that have been conducted and innovative ideas which are being pursued, the advantages and difficulties of concentrated solar radiation is discussed. The present opportunities and future trends are also presented.

III. BRIEF HISTORY

The application of solar furnaces is not new and many researchers have studied it over time. It was Archimedes in 212 B.C. who used this principle to destroy the Roman fleet by concentrating solar radiations using a large number of mirrors. Over the 17th and 18th centuries, the use of both lenses and mirrors were explored. Lavoisier was successful in building a solar furnace with a diameter of 5 feet and was successful in attaining the melting point of platinum. It was in 1921, Straubel constructed the first reflecting type modern furnace attaining a temperature close to 3,0000 C. Another example is of Aluminum Alloy sheet Solar furnace operational at Convair for the study of "High-Temperature Materials". However, the largest solar furnace is located at Montlouis, France. It has a 6.5 feet diameter parabolic searchlight mirrors and the number of such furnaces is six. It also has a large 35 feet diameter solar furnace with attached 3,500 small mirrors attached in a parabolic shape steel frame. Similarly, a 27 feet diameter solar furnace was built in Algiers consisting of 144 electropolished aluminum panels shaped in the form of a parabola.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:12/December -2020 Impact Factor- 5.354 www.irjmets.com

IV. TYPES OF SOLAR FURNACE

Solar furnaces can be classified mainly into Direct Solar furnace and Heliostat Solar furnace. Direct Solar furnaces have a simpler principle, using a lens to converge the direct light to the focal point. The light concentrated at the focal point acts as a furnace.

4.1 Direct Solar Furnace

Direct Solar Furnace can be further divided into various types.

4.1.1 Single Lens Direct type

This type of Solar Furnace does not have practical applications. This acts as an example for explaining the principle of the solar furnace. It includes a single lens for converging or concentrated the light at the focal point increasing its temperature.

4.1.2 Multiple Lens type

In this type, multiple lenses are used to focus light through the first series of lens. This then converges to the focal point through the second row of lens using plane mirrors as the means to direct the path towards the focal point i.e., the furnace.

4.1.3 Single Parabolic Direct type

This is also one of the simplest types where instead of a lens, a parabolic mirror is used to reflect the light towards the focal point.

4.1.4 Lens-type furnace

In Lens type Solar Furnace, an arrangement of multiple lens and mirror is used for focusing sunlight. The basic process can be defined as, the incident rays get refracted from the first lens, traveling towards the second lens. Any incident rays which might be moving outside get reflected by the mirrors located. After passing through the first lens, the rays get refracted from the second lens, converging at a single point or focal point of the whole apparatus.

These types of solar furnaces were used in the very early days of discovery. This type of solar furnace cannot change its position or keep track of the sun as the day progresses, which is a big disadvantage. Thus, they became obsolete leading to the development and introduction of heliostats. The use of heliostat in solar furnaces allows its application in heavy industrial applications where high temperatures are involved. This type of solar furnaces has a high investment cost which gets compensated by its low operating cost.

4.2 Heliostat Solar Furnace

Heliostat Solar furnaces are further divided into multiple types depending upon the orientation of the optical axis.

4.2.1 Heliostat Optical Axis Horizontal

In a Horizontal axis heliostat solar furnace, a plane mirror acting as a heliostat reflects the solar rays coming from the sun to a parabolic mirror. The reflection from the mirror makes the rays parallel and horizontal to each other. The rays reflect from the parabolic mirror and converge at the focal point which acts as the receiver. Odeillo solar furnace uses this principle for solar concentration.

4.2.2 Heliostat Optical Axis Vertical

In a vertical axis heliostat solar furnace, the heliostat reflects the sun rays directly falling on it to the parabolic mirror in such a way that the rays become vertical after reflection and converges to the focal point, acting as the receiver.

4.2.3 Heliostat Optical Axis Horizontal and Vertical

This is a hybrid version of the horizontal and vertical optical axis heliostat and can be used for customized design and applications.

V. MAIN PARTS

The main parts of a solar furnace are discussed below.

5.1 Heliostat

Heliostat is an instrument whose function is to focus sunlight continuously in a fixed direction. In a solar furnace, the heliostat tracks the sun and orients itself directing the sunlight to the concentrator. This leads to the stabilization of the sun's rays, allowing it to focus at a particular spot for optimum performance. In general, heliostats include plane and parabolic mirrors whose orientation is maintained with the help of controllers and encoders.[1] The categorization of a heliostat based on the mounting of the actuators, serial and parallel. In a serial type, the heliostat has a simple design and control but with reduced stability. Whereas, in the parallel type heliostat, provides increased stability. The mode of operation for a heliostat system is classified as closed-loop and open-loop control mode. The open-loop control mode is used by heliostats which are rigid and whose properties like geometry, rigidity, transmission ratio, etc. are maintained between successive calibrations. The closed-loop control mode is used for non-rigid heliostat and it tracks the sun's position accurately. Its fast control provides low errors in operation. The desirable properties in a heliostat used in concentrated solar power plant are high reflectivity, high optical precision, high tracking accuracy, and resistant structure.[2]

5.2 Parabolic Concentrator

The function of the parabolic concentrator is to concentrate or direct the solar energy at a particular location known as the focal point. The concentrator has several curved mirrors that help in focusing sunlight at a particular spot. The construction of a parabolic mirror as a whole is not feasible and economical.[3] Hence, many plane or parabolic mirrors are attached to the concentrator for operation. The use of several mirrors allows the incoming reflection to be focused on the collector, usually located in a tower. Although the concentrator may not be able to concentrate sunlight on a single spot, the focus area can be considered to be a circle or an ellipse. The sun rays reflected from the heliostat are focused on using the parabolic concentrator. The compound parabolic concentrator can be categorized as symmetric and asymmetric. The absorbers employed in concentrators are generally tubular and fin-type with pipe absorbers.

5.3 Attenuator/ Shutter

The attenuator is used to reduce or change the power quickly being transmitted. It regulates the amount of sunlight transmitted, thus, changing the power according to requirements. The response of the shutter should be fast and it depends on several factors such as the size of the solar furnace, location, and material. The panel used for the attenuator is generally made from aluminum. The control of the attenuator is digital, allowing it to send commands for changing the position and energy transmission.[1]

Some of the other parts of the solar furnace include the test table or reactor table. It can be defined as the location of placing the specimen in testing applications. It also consists of measurement sensors, high-temperature receivers, and reactors.

VI. PARAMETERS AFFECTING SOLAR FURNACE

The basic principle or functionality of a solar furnace is to achieve high temperature at a point, generally the focal point, by accumulating or concentrating solar radiation at that particular point. For the successful operation of a solar furnace, there are three main parts to be considered, i.e., Heliostat, Parabolic Concentrator, and Shutter.

In designing a solar furnace for a specified heat flux at the target, there are two classes of factors (theoretical and operational) which are to be considered. Both have their fair share in deciding the performance of the furnace. The first class relates to the optical geometry of the concentrator and includes six important design variables:

- Aperture
- The Ratio of aperture to focal length
- Image Diameter
- Average heat flux at the image
- Uniformity of flux distribution at the image
- Efficiency at which energy received by the concentrator is concentrated.

The six design variables considered, define the theoretical characteristics and the operational characteristics, and are the ones that help in estimating the working efficiency. [4]

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:12/December -2020 Impact Factor- 5.354 www.irjmets.com

The operating performance of a furnace depends on all three parts. However, the concentrator is the heart of all the process and determines the operating efficiency of the furnace. Despite continuous support for the revolution paraboloid as the ideal design for the concentrator, the considerable cost in its manufacture and its time-consuming testing are crucial parameters to be looked upon for its commercial application, besides, its extremely high demands in precision. Thus, a practical and economical alternative is using a faceted structure covered with identical curved mirror arrays. This is a more viable option and is commonly used around the globe. The overall performance in these types of concentrators depends on facet sizes and optical errors.

To judge the performance parameters, total heat flux distribution is studied and parallel conclusions are drawn. The Monte Carlo ray tracing (MCRT) method is very flexible and has been proved to have great accuracy and so is generally followed. This simulation gives four verticals of error according to which the performance of the furnace varies.

- First, the tilt error of the heliostat, which originates from heliostat facets deviating from the ideal plane.
- Second, the slope error of the concentrator, which refers to the irregular deformation of reflecting curves of the concentrator facets by the surface adjustment approach.
- Third, the layout error of the concentrator, which is caused by the same or similar shape of facets attached to the concentrator frame.
- Fourth, the tilt error of the concentrator, which arises from concentrator facets tilting.

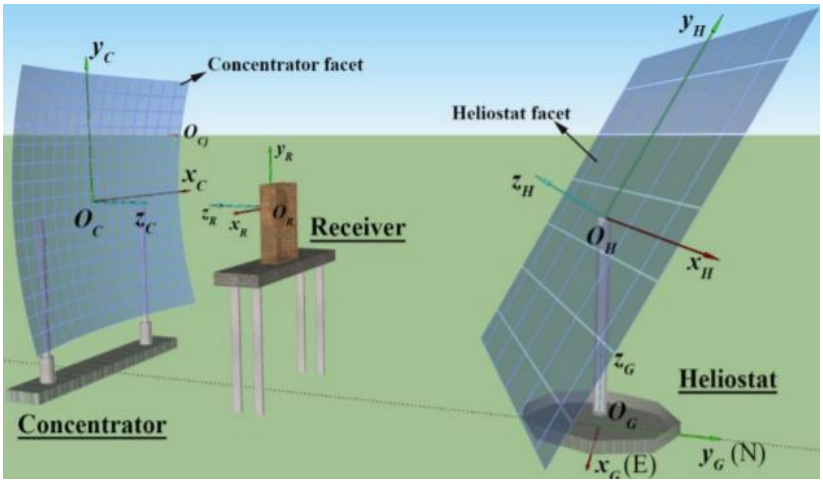


Fig.-1: Schematic structure and placements of components in a solar furnace concerning their respective coordinate systems. [4]

The main components of the optical system and receiver are placed on the same line, joining the center of the concentrator and the center of the heliostat. The tilt error of heliostat affects the non-parallelism degree of the reflected sunlight. As the error increases, both the concentrated heat flux distribution and spot does not meet in a line, i.e., they are diverging.

The FEM (Finite Element Method) is used to obtain the curved surface of the concentrator facet. The slope error of the concentrator results in the mirror unit no longer acting as an ideal surface, but rather as a complex and irregular one under constraints. The error in factors of 1~2 mm could greatly reduce the concentration ability. The layout error occurs when facets attached to the concentrator frame share a similar curvature. Its existence will not enlarge the influence of the slope error. The tilt error of concentrator facets directly affects the focusing effect. As the tilt error increases in a particular direction, heat flux along the direction scatters, while in the orthogonal direction, it maintains in the desired shape, i.e., the gaussian shape but the overall value decreases. [4]

Hence, the deviation in the mountings from the desired axis leads to errors, and as a result, the overall efficiency of the plant decreases. Apart from the mounting parameters, there are certain things regarding the components that should be kept in mind to be able to have smooth functioning.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:12/December -2020 Impact Factor- 5.354 www.irjmets.com

Heliostats are computer-driven devices tracking the changing direction of sunlight as the day progresses, maintaining uniformity in the direction of the rays falling on the concentrator. It is operated in two modes, the open-loop control mode, and the closed-loop control mode. The quantity of light reflected by the heliostat depends on the area and cleanliness of the mirror. The cleanliness of the mirror decreases with time due to dust deposition and requires timely cleaning and maintenance. Depending on the time of day and dust level on the heliostat, the reflected power can vary between 85% to 90% of the available sun's power. [4]

VII. OVERVIEW OF SOLAR FURNACES AROUND THE WORLD

Many operational solar furnaces projects around the world are used for various industrial and research applications. Some of them are covered in this review paper.

7.1 Odeillo Furnace

Odeillo Furnace, the largest solar furnace in the world, is situated and operational in the south of France in Font-Romeu-Odeillo-Via, in the department of Pyrénées-Orientales, region Languedoc-Roussillon. It was built between 1962 and 1968 with operation starting in 1969 and has a power of one megawatt. The measurement facts of the solar furnace are, 54 meters (177 ft) in height and 48 meters (157 ft) width, consisting of 63 heliostats.

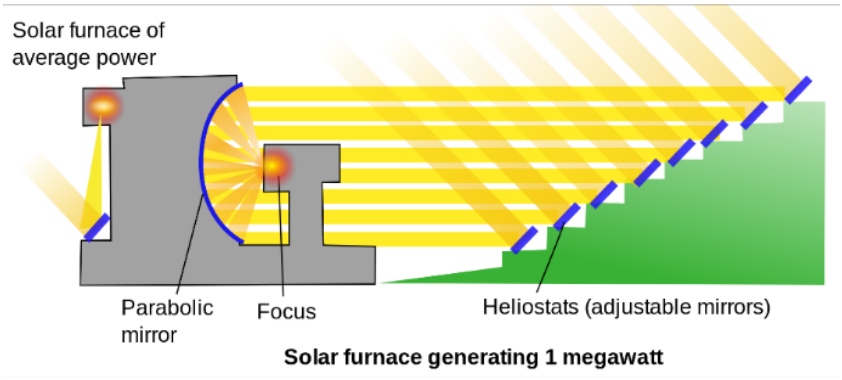


Fig.-2: Depiction of working principle

The reflecting mirrors, Heliostats, 63 in number, each having 45m^2 surface with 180 single mirror panes, placed in a step-wise fashion, reflecting sun rays to the concentrator. The solar furnace consists of a parabolic mirror, acting as a concentrator that converges the light on the specified target known as the "focus".[5]

The heliostats are automatically operated by photoelectric control of their reflected rays. The driving parts that are electronically coupled with the photoelectric control are hydraulic.

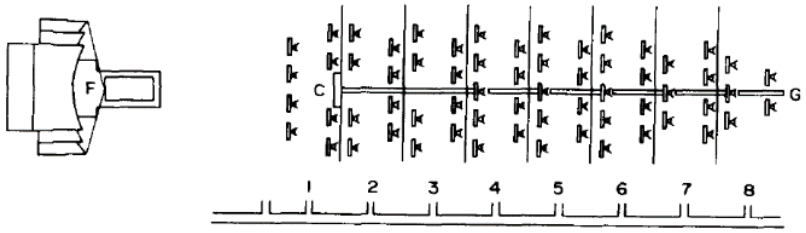


Fig.-3: Schematic diagram of the furnace [5]

The above figure is the top view of the setup with point C representing the station of oil under pressure that acts on the site, and the azimuth of the heliostats through a tunnel, denoted by point G. Each orientator is made up of 180 flat glass panes, back silvered and covered with copper and varnish. The apparent diameter is also taken into consideration while positioning the respective heliostats. This helps in achieving a coherent beam of sunlight that would be comparable to the beam that would be given by a single large orientation.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:12/December -2020 Impact Factor- 5.354 www.irjmets.com

The parabolic mirror is fixed and has a focal length of 18m, 54m wide, and 40m high. It includes 9500 flat plate glasses, curved and adjusted under mechanical constraints which helps in getting an image with minimum dimension. [5] Regarding the distribution of flux density, an experiment was conducted in 1970, in which a vertical plane was located at 18m from the apex of the paraboloid. The conclusions derived, the energy distribution on the plane was rather elliptic than circular as shown in the figure, and that the distribution remained practically unchanged when the receiving plane was tilted up from the vertical position to a 25 degrees inclined position. [5]

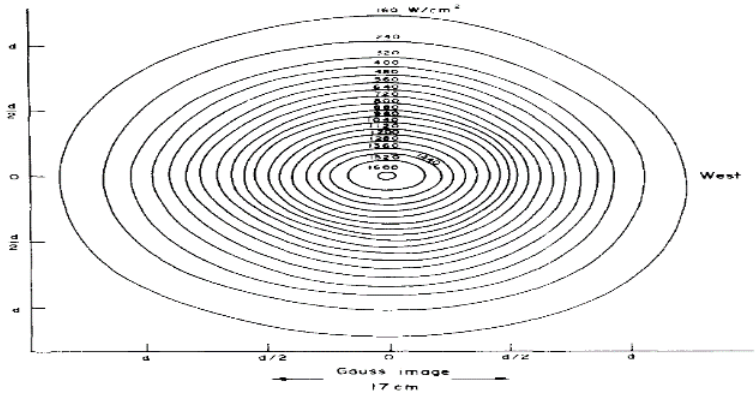


Fig.-4: Distribution of heat flux density on a vertical plane [5]

Owing to the number and the size of heliostats, and the dimensions of the mirror, this furnace concentrates energy that makes it reach a peak of 3200KW. The long-running success of this furnace lies in the fact that the energy densities achieved are comparable to the radiations of a blackbody at a very high temperature. To express this quantitatively, the equilibrium temperature at the center of the focal image would be higher than 4000 K.

7.2 High Radiative Flux Solar Furnace (HRFSF)

A high flux solar furnace facility has been operational in Mexico, located in Temixco, Morelos, Mexico. The installation was officially inaugurated on 18th March 2011. The first stage of the furnace was finished and started up in December 2011. [6]

The design of this furnace considers the intercepted power of approximately 30 KW with a peak concentration of 10,000 suns. The dimension of individual parts are as follows: heliostat of 81 m^2, a shutter made of stainless-steel blades 6.2m^2, and a concentrator of 409 spherical mirrors with hexagonal contours mounted on a spherical frame, and an equivalent focal length of 3.68m. [6]



Fig.-5: Showing Heliostat and Shutter on the left and Concentrator on the right [6]

The left side of the above figure shows the heliostat and shutter and the right side shows the concentrator.[7] As the site of the furnace has a low latitude as compared to the other existing facilities, a large heliostat is required, to illuminate the concentrator for a sufficient number of hours; three hours at least in the summer solstice and eight hours at least in the winter solstice. This facility can reach temperatures as high as 3680 K.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:12/December -2020 Impact Factor- 5.354 www.irjmets.com

7.3 Solar Powered High-Tech Furnace

The solar-powered high-tech solar furnace is operational in Valparaiso University, Indiana, USA. The solar furnace consists of six main parts as shown in the figure below.

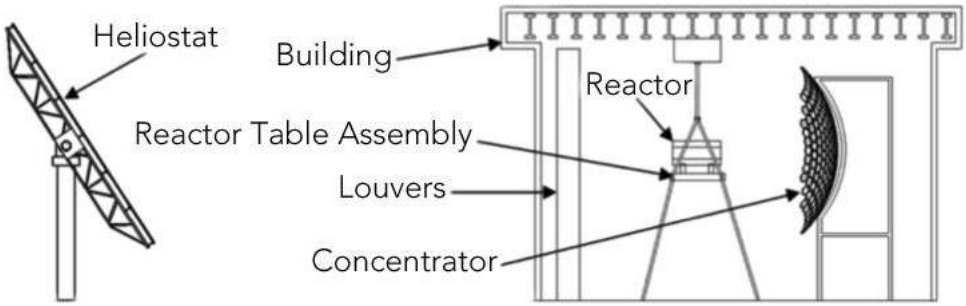


Fig.-6: Schematic diagram of the high-powered tech solar furnace

The heliostat is located in front of the furnace building and is a flat mirror measuring 20-foot by 20-foot, mounted on a pedestal outside the oval-shaped building that houses the furnace. The heliostat tracks the sun reflecting the light to the concentrator which focuses it at the focal point. The concentrator at the Markiewicz Solar Energy Research Facility consists of 306 curved mirrors. The light reaching the concentrator is regulated by a set of louvers placed in between the heliostat and the concentrator.

In this furnace, the point where all the light from the concentrator is focused, a solar thermal reactor is located. The resulting steam drives an electricity-generating turbine. This furnace is capable of generating equal to and beyond 3000 degrees Fahrenheit within the reactor.

Researches on this solar furnace, as of now, are two. First, converting water into the fuel hydrogen - Solar thermal decoupled water electrolysis project being conducted. The second one, splitting zinc oxide into zinc and oxygen, without producing any waste. Zinc can be used in a fuel cell to generate electricity or as a commodity.

7.4 High Flux Solar Furnace (HFSF)

NREL's High flux Solar Furnace (HFSF), a unique 10KW optical furnace has been in operation since December 1989 and is under the Solar Energy Research Institute (SERI). It is situated at the top of NREL's South Table Mountain campus. HFSF has supported several research experiments since it became operational.

In terms of major achievements, in July 1990, SERI researchers produced the first high-temperature superconducting material at the HFSF. Detailed measurements indicated zero resistance at 74.4 K. [8]



Fig.-7: High Flux Solar Furnace

The high performance of this furnace can be attributed to the uniqueness of its design. HFSF is capable of achieving extremely high flux concentrations in a two-stage concentration and has the capacity of generating a

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:12/December -2020 Impact Factor- 5.354 www.irjmets.com

wide range of flux concentrations. It consists of a stationary focal point and is movable off-axis. This off-axis system allows considerable flexibility in size. The figure below depicts the schematics of the HFSF. [8] This furnace consists of a tracking heliostat and 25 hexagonal concave mirrors that concentrate solar radiation to deliver 10 kilowatts of thermal power to a focal area about 4 inches (~10 cm) in diameter within the laboratory control room. The solar furnace can quickly concentrate solar radiation to 10 kilowatts over a 10-cm diameter (2500 "suns"), achieving temperatures of 1800°C and up to peak solar fluxes of 20,000 suns with specialized secondary optics to produce temperatures of up to 3,000°C. Secondary concentrators can modify the focal point, tailor flux levels, and distributions to suit the needs of each research activity. [8]

The HFSF facility also consists of instrumentation to monitor solar radiation levels, automated controls to adjust the power of concentrated sunlight, data acquisition tools, and video monitoring of outside equipment. The HFSF concentrates actual sunlight, it offers an energy-efficient alternative to solar simulators, with tests performed under realistic rather than simulated conditions.

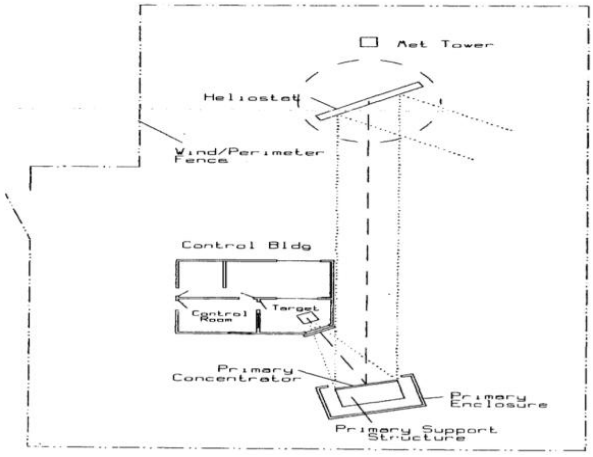


Fig.-8: Schematic Diagram of HFSF [8]

7.5 Jülich Solar Furnace

Jülich Solar furnace is located in Germany and is the first solar furnace constructed for experimental purposes having a power generation capacity of 1.5 MW. The construction of this furnace began in 2007 and it was inaugurated in the year 2009. [9]

In the solar furnace, the open volumetric receiver technology uses ambient air as heat transfer fluid (HTT). Ceramic absorber modules are used which are heated up by the concentrated irradiation. Air is sucked through the ceramic particles heating it to 680° C. This heated air is used further to generate electricity in a steam power plant using Rankine cycle. [10]

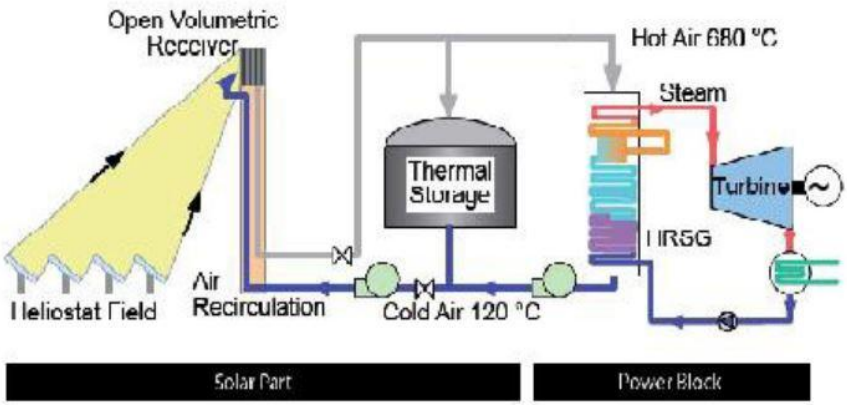


Fig.-9: Schematic Diagram of Jülich Solar Furnace [11]

The additional energy gets stored in a chamber since storing thermal energy is comparatively easy than electricity. This stored energy can be used at night or at unexpected times such as when the weather is cloudy or other factors. The heat transfer coefficient of air is low; therefore, the efficiency isn't too great for this solar furnace. [11]

The solar furnace consists of 2153 heliostats, with an individual aperture area of 8.2 m^2, making a total area of 17650 m^2. There is a 60m tower housing the concentrator, which is connected to the steam turbine to generate electricity. The cooling method used is dry cooling, whereas, the heat storage capacity is 1.5 hours. The turbine used is of 1.5 MW capacity.[9]



Fig.-10: Jülich Solar Furnace [11]

7.6 Recent Development in Concentrated Solar Energy in India

There have not been any large solar furnace projects in India as compared to the likes of Odeillo and Jülich. However, an increase in research and development in this area has led to the development of various sizes of solar furnaces and concentrated solar power projects in the country. One such project is being developed by Shri Shakti Alternative Energy Limited along with Solar Technology Advisers and Swarg Vatika Crematorium. They aim to develop a solution to decrease the consumption of LPG and biomass by using concentrated solar energy.

The operational projects using concentrated solar power principle as an energy source are Tirumala Tirupati Devasthanam kitchen at Tirupati, where 73 parabolic concentrators are used for heating water to a temperature level of 550° to 650° C. This leads to a production of 3600 Kg steam every day, saving about 100,000 kg of cooking gas per day. Another project is ACME solar plant, located in Bikaner, Rajasthan. It consists of 14,280 heliostats with each heliostat having an area of 1.136 m^2. The total Heliostat Solar-Field Aperture Area is 16,222m^2 and the plant has a power generation capacity of 2.5MW. [9]

VIII. INNOVATIVE RESEARCH AND ADVANTAGES OF SOLAR FURNACE

There have been various research projects undertaken in this field. Some of the projects or innovations around the world have been discussed.

8.1 Process Chamber for Thermal Dissociation of ZnO

The group based in Switzerland performed research studies and projects dealing with the thermal dissociation of ZnO. The objective was to perform redox cycle in two steps: the dissociation of ZnO using solar energy at a temperature above 2000K and oxidation of Zn with H_2O/CO_2 to produce H_2/CO. Zn which is recovered can be used to split H_2O and byproducts can be used in power generation in a fuel cell. [12]

8.2 Glass Production and Melting

The raw material of lunar origin contains high silica, content slag, and calcium oxide and is present in abundance on Moon. Due to the high cost of transportation of materials from Earth, NASA in 1979 was motivated to use solar radiation for the onsite production of glass. The research included static and semi-continuous melting experiments. Initially, in the static melting experiment, a powdered soda-lime-silica glass

was used and it was shown that the crystalline raw material transformed successfully into a transparent glass. After the small-scale demonstration, the author concluded that significant research and experimental work is required before commercial feasibility can be evaluated. [12]

8.3 Solar Heat for Cement Production.

Here the author brought the focus on the CO_2 emission in the whole world. In the whole world, CO_2 production can be classified broadly into two categories: Natural and anthropogenic. Considering the anthropogenic category, which is majorly produced due to the burning of fossil fuel, cement production, and farm plowing. The burning of fossil fuel can be decreased by using non-conventional alternatives such as wind, hydro and solar energy. In the cement production, the global production of Portland cement was about 4.05 Gt in 2017 and 4.10 Gt in 2018. The manufacturing of Portland cement results in the production of CO_2 due to the decarburization of limestone.

An interesting approach to reduce the emission of CO2 can be by using a directly irradiated fluidized bed reactor to produce lime for cement production. In the research paper, the author presented a conceptual design of the reactor for the production of Portland cement where concentrated solar energy was used in a conical hole, in a cylindrical reactor. The gases produced can be collected and treated for heat recovery. [12]

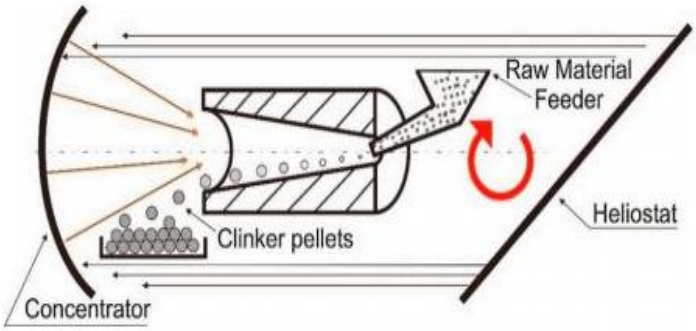


Fig. 11: Conceptual design of reactor [12]

8.4 Rapid Heating and Thermal Cooling

The rapid heating and thermal cycling are undoubtedly one of the significant advantages of the solar furnace over other traditional technologies. The rapid cycles of temperature variation at the surface of the workpiece are hard to achieve by electric resistance furnace, induction furnace, and other furnaces whereas in solar furnace this is achieved easily. Also, the solar furnace does not have any environmental impact like the other methods. [12]

8.5 Wide Spectrum of Radiation

Solar radiation consists of a very wide spectrum, from ultraviolet to infrared radiation, including visible spectrum which is not seen in many artificial sources like X-rays, microwaves, electromagnetic waves in the radar. Solar radiation and solar furnace have extensive applications in studies involving phenomena occurring on materials and structures. [12]

8.6 Difficulties in Concentrated Solar Radiation

- Solar Radiation is dependent on various factors like latitude, weather conditions, and time of the day.
- After concentration, the solar radiation becomes unidirectional thus heating the component in one direction.
- The flux of solar radiation at the focal point is non-homogeneous theoretically. [12]

IX. CONCLUSION

The new vertical of clean and environment-friendly energy sources led the world to shift from conventional sources like fossil fuels to renewable sources like solar, wind, and hydro and the past few decades have seen an unparalleled rise in the solar energy sector. India is blessed geographically that it lies near the equator and as a result, receives one of the highest solar irradiations on earth.

Solar Furnace has been an alternative for higher temperature production in many fields but due to the capital investments and logistics involved, it is yet to be used at its full potential in India, and though the work in this direction is being carried out, we need a Public-Private Partnership (PPP) model that can back up the shortcomings associated with a large-scale solar furnace such as Odeillo Furnace in France and this falls in line with the Paris Agreement that we be more cautious towards the climate change.

The Solar furnace has the potential to replace the furnaces used in metallurgy, cupola furnace, for example, that uses coal as a source of heat and other such applications where high temperatures are involved and this replacement accentuates the fact that the target of 100 GW solar energy by 2022 could be met.

X. REFERENCES

[1] B. Andrade Costaa, J.M. Lemosa, L.G. Rosab, "Temperature control of a solar furnace for material testing", International Journal of Systems Science Vol. 42, No. 8, August 2011, 1253–1264

[2] Johannes Christoph Sattlera, Marc Rögerb, Peter Schwarzbözlc, Reiner Buckd, Ansgar Mackee, Christian Raederf, Joachim Göttschea, "Review of heliostat calibration and tracking control methods", 2020, doi.org/10.1016/j.solener.2020.06.030

[3] Gregory Scott Duncan, Shahin S. Nudehi, Robert D Palumbo, Daniel Blood, Luke Jerod Venstrom, "Student Development of a Five kW Solar Furnace for Solar Thermal Chemistry Research", ASEE Annual Conference and Exposition, 2015

[4] Zhiying Cui, Fengwu Bai, Zhifeng Wang, Fuqiang Wang, "Influences of Optical Factors on the Performance of the Solar Furnace", Energies, 2019, 12, 3933, doi: 10.3390/en12203933

[5] Felix Trombe, Albert Le Phat Vinh, "Thousand-Kilowatt solar furnace, built by the National Centre of Scientific Research in Odeillo (France)", Solar Energy, 1973, Vol. 15, pp. 57-61

[6] A new high-flux solar furnace at Cie-Unam in Temixco, Mexico. First Stage by Claudio A. Estrada, Camilo A. Arancibia-Bulnes, Sergio Vazquez, Carlos A. Pérez-Rábago, David Riveros, Ricardo Perez-Enciso, Jesús Quiñones, Rafael Castrejón, Moises Montiel, Fermín Granados; 2014.

[7] D. Riveros-Rosas, J. Herrera-Va´zquez, C.A. Pe´rez-Ra´bago, C.A. Arancibia-Bulnes, S. Va´zquez-Montiel, M. Sa´nchez-Gonza´lez, F. Granados-Agustı´n, O.A. Jaramillo, C.A. Estrada, "Optical design of a High radiative flux solar furnace for Mexico", 2010, doi: 10.1016/j.solener.2010.02.002

[8] Allan Lewandowski, Carl Bingham, Joseph O'Gallagher, Roland Winston, Dan Sagie, "Performance characterization of SERI High Flux Solar Furnace", Solar Energy Materials, 24, (1991), 550-563

[9] Rajat Malik, Raviratna Subir, Rishabh Goswami, J.P. Kesari, "CONCENTRATED SOLAR POWER FOR MID-DAY MEAL PROGRAM IN DELHI", International Research Journal of Modernization in Engineering Technology and Science, Volume:02/Issue:11/November -2020

[10] Transient Simulation of a Solar-hybrid Tower Power Plant with Open Volumetric Receiver at the Location Barstow, C.RauS, Alexopoulos, G.Breitbach, B.Hoffschmidt, M.Latzke, J.Sattler,2014

[11] Valentina Kronhardt, Spiros Alexopoulos, Martin Reißel, Johannes Sattler, Bernhard Hoffschmidt, Matthias Hänel, Till Doerbeck, "High-temperature thermal storage system for solar tower power plants with open-volumetric air receiver simulation and energy balancing of a discretized model", (2013)

[12] Luís Guerra Rosa, "Solar Heat for Materials Processing: A Review on Recent Achievements and a Prospect on Future Trends", 2019, 3, 83, doi: 10.3390/chemengineering3040083

Journal of Molecular Structure 1228 (2021) 129672

Contents lists available at ScienceDirect

Journal of Molecular Structure

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

Recent developments in *meta*-benziporphodimethene: A new porphyrin analogue

Deepali Ahluwalia, Anil Kumar*, Sudhir Gopalrao Warkar

Department of Applied Chemistry, Delhi Technological University, Delhi-42, India

ARTICLE INFO

Article history:
Received 9 September 2020
Revised 7 November 2020
Accepted 15 November 2020
Available online 8 December 2020

Keywords:
meta-benziporphodimethene
benziporphyrin
Density functional theory
Cell-imaging

ABSTRACT

The article underscores a unique class of benziporphyrin analogues, in which two of its *meso*–carbon atoms are sp^3 hybridized, leading to the formation of a new molecule: *meta*-benziporphodimethene (*meta*-BPDM). This molecule exhibits intriguing structural, spectroscopic and chemical properties. *meta*-benziporphodimethene was first discovered 16 years ago. Although, initially it was an accidental product obtained while synthesizing *meta*- and *para*- benziporphyrin, but the molecule stood-out when acted as a chemo-sensor for Zn^{2+} ions. The discussion here, starts with outlining the importance of porphyrin molecules, followed by the progress in the field of its synthetic modifications. Special emphasis has been given to the new porphyrin analogue: *meta*-BPDM. We have herein covered some recent applications reported like metal-ion sensing, breast cancer detector, UV- turn-on behavior for toxic metals of *meta*-BPDM. Owing to its low synthetic yield, not much research has been done on these systems, despite their beneficial properties. We have also inculcated the factors that may lead to enhancement of product yield of *meta*-benziporphodimethene systems. The X-ray and NMR analysis indicating weak agostic interactions has also been discussed. This article shall provide a better understanding of this class of carbaporphyrinoid family, to its readers and may open new paths for exploring these systems in future.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Porphyrin is a tetra-pyrrolic moiety, very stable, aromatic and is one of the most important molecules in the biological system. Porphyrins are important because of its multiple biological functions and its ability to bind a number of transition metal ions [1–4]. The basic porphyrin-like moieties are found universally in natural pigments like heme, chlorophyll, bacteriochlorophyll, vitamin B_{12} and cytochromes (Fig. 1) [5–8].

Simple porphyrin molecule is highly aromatic in nature. There are 22π-electrons in a simple porphyrin molecule however only 18π-electrons take part in complete conjugation in several possible delocalization pathways, as (represented by bold bonds) in Fig. 2 [9,10].

Over the years, researchers have showed great enthusiasm in exploring this prosthetic molecule employing modifications at several sites in the molecule (Fig. 3) [3]. The major objectives of these investigations is to find the applications parallel to that of natural porphyrins so as to mimic the natural phenomena in-vitro. The electronic structure of porphyrins play a significant role to decide the fate of the system. A vast variety of conjugated and non-conjugated porphyrinoid systems give us a choice for selecting moiety for different applications and investigations [11,12].

The modification in porphyrin molecule can be done in many ways for various purposes. The central region of porphyrin molecule is NNNN core that is surrounded by four pyrrolic nitrogen atoms (Fig. 3) [5–8]. The new synthetic analogues are classified according to their structure and the electronic current they possess, and are generally called modified class of porphyrins [9]. The present study aims to provide a comprehensive and updated review on one such modification in the porphyrinoid system, i.e., *meta*-benziporphodimethene molecules (2e, Scheme 2). The article is divided into 3 major topics to allow the reader to easily maneuver through this review. The first part is concerned with understanding the basic modifications that can be done in a porphyrin system. In the second topic, the readers shall learn briefly about carbaporphyrinoid family. The major focus of this review is the third topic, wherein *meta*-benziporphodimethene systems have been discussed. In this part, methods of synthesis, characterisations and applications of this moiety will be discussed.

1.1. General modification in porphyrin

Generalizations of these modified porphyrins are shown in Scheme 1. Periphery modified porphyrins, which includes modifi-

* Corresponding author.
E-mail address: anil_kumar@dce.ac.in (A. Kumar).

https://doi.org/10.1016/j.molstruc.2020.129672
0022-2860/© 2020 Elsevier B.V. All rights reserved.

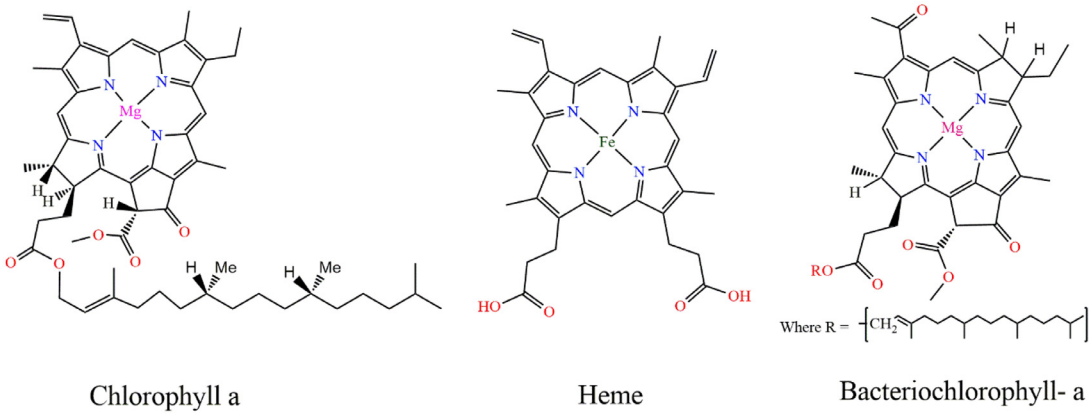


Fig. 1. Porphyrin-like moiety in natural pigments.

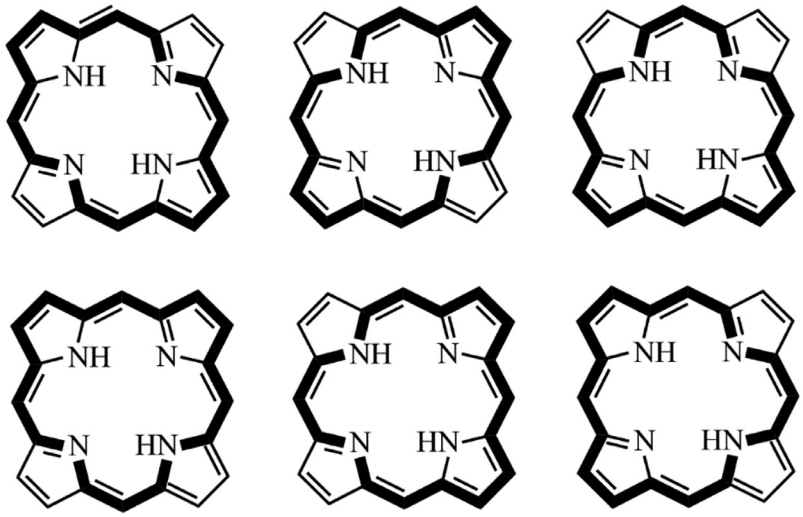


Fig. 2. Possible pathways for the delocalisation of π-electrons in porphyrins. [5].

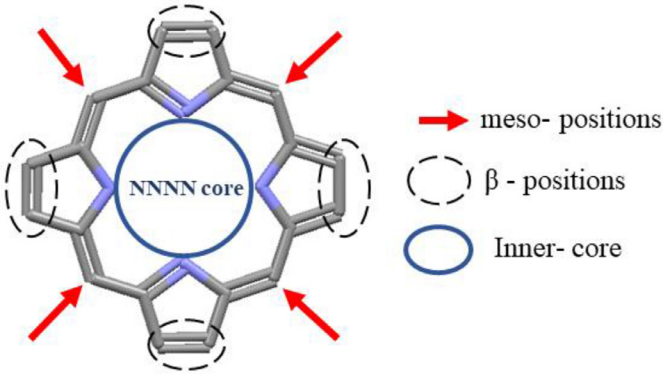
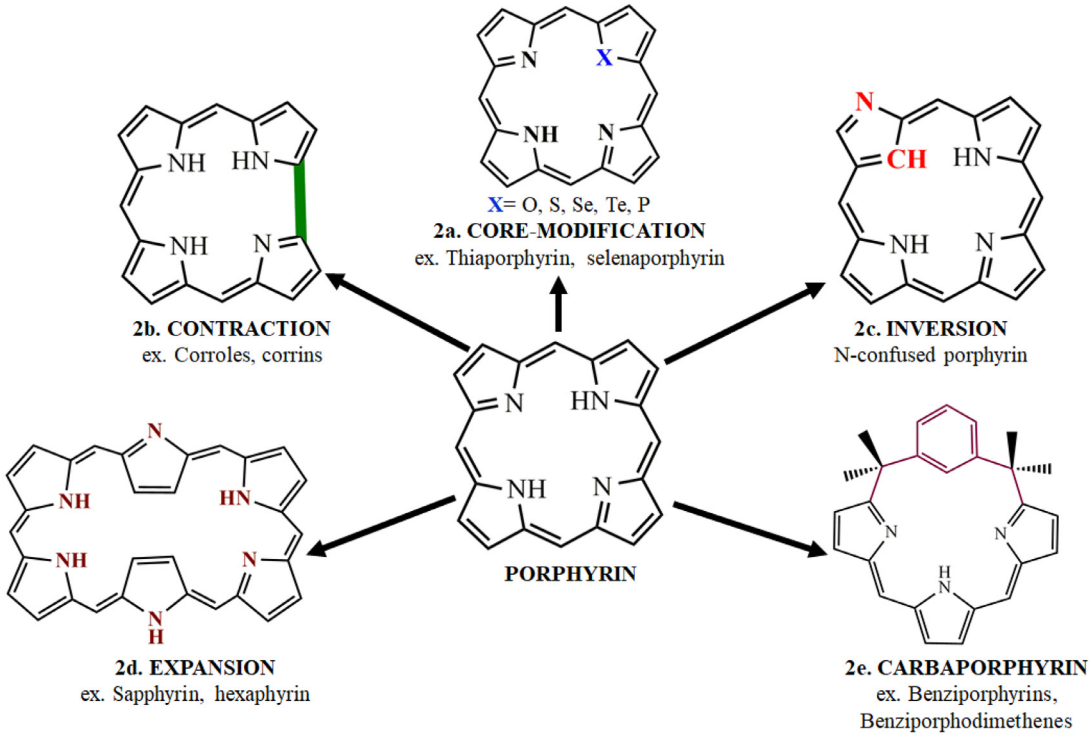
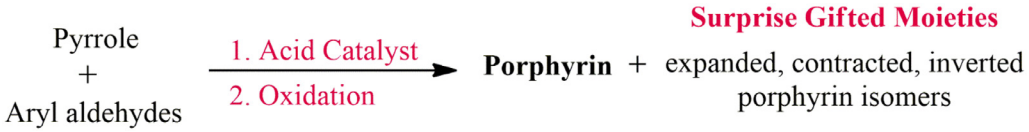


Fig. 3. Structure of porphyrin and the sites of modifications.

cations in the periphery of tetra-pyrrolic moiety and includes β- and *meso*–positons (Fig. 1) [10–14]. Core-modified porphyrins [15], where inner nitrogen atoms of the tetra-pyrrolic moiety are replaced by other elements (generally chalcogens), also called heteroporphyrins, i.e., the NNNN core is replaced by XNNN core (Scheme 2a) The heteroatoms introduced other than nitrogen in the inner core modifies the inner core arrangement of macrocycle and hence changes the binding abilities of these molecules with different transition metal which results the shift in their absorption spectra[16]. Macrocycles containing oxygen [9], sulfur [15,17], tellurium [18,19], selenium [20,21] and recently phosphorus [22–24] provide a platform to constitute a variety of macrocycles able to bind with diffrent transition metals in their unusual oxidation states with excellent properties [16,25]. Contracted porphyrins [29-41], wherein one lesser *meso* carbon is present than the regular porphyrin, for instance, corroles (Scheme 2b). In isomeric porphyrins [39], similar structures and formula, i.e., $C_{20}H_{14}N_4$ by scrambling pyrrolic subunits and bridging carbon atoms. Inverted porphyrins [25], porphyrin isomers having either N or C of pyrrolic units pointing out or inside of the ring, respectively, also called 'N-confused porphyrins' or 'mutant porphyrins' (Scheme 2c) [26–28]. Expanded porphyrins [40], these are the synthetic analogue of porphyrin macrocycles with enlarge core comparative to the other members of the porphyrin family with a minimum of 17 atoms in the inner central core while retaining the extended conjugation of the system (Scheme 2d). The inner core atoms of macrocycle are generally greater than 17 atoms as of regular one and expand the π -electron conjugation. These expanded systems strongly absorb in the red region as compared to 17 atoms and 18π electrons porphyrins. Sapphyrins [41], smaragdyrins [42], isosmaragdyrins [43], pentaphyrins [44], hexaphyrins [45,46], orangarins [47], amethyrins [48], octaphyrins [49,50], rosarins [51], rubyrins [52], ozaphyrins [53], bronzaphyrins [54,55], heptaphyrins [56], turcasarins [57], dodecaphyrins [58] and hexadecaphyrins [58] are the few examples of expanded porphyrins.



Scheme 1. Representation of modifications in the central core and the types of porphyrin analogues.



Scheme 2. Basic route for synthesis of Porphyrin and by-products.

2. Carbaporphyrinoids - a turning point in porphyrin chemistry

The first porphyrin molecule was reported in the year 1867 by Thudichum J. L. W., as hematoporphyrin which was originally named as cruentine [59].

It is very common that the process not only give the desired product but other gifted compounds like expanded, contracted, inverted, and other categories of compounds as well (Scheme 2). It thus became necessary to optimize the conditions of reactions in a systematic way for obtaining the desired product only. Consequently, the one pot synthesis of porphyrins pioneered by Rothemund [60] had been optimized in the past in a number of ways by using various modified catalysts to direct the synthesis towards a particular moiety and to reduce or eradicate the unwanted products [61–66].

A revolutionary point arrived when in 1943, while revisiting Rothemund's synthesis of tetraphenylporphyrinin Aronoff and Calvin reported six-porphyrin-like-compounds ($R_4C_{20}H_{10}N_4$) which may be separated chromatographically. One of these isomers involved one inverted porphyrin ring (Fig. 4).

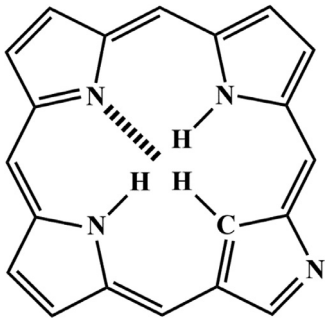


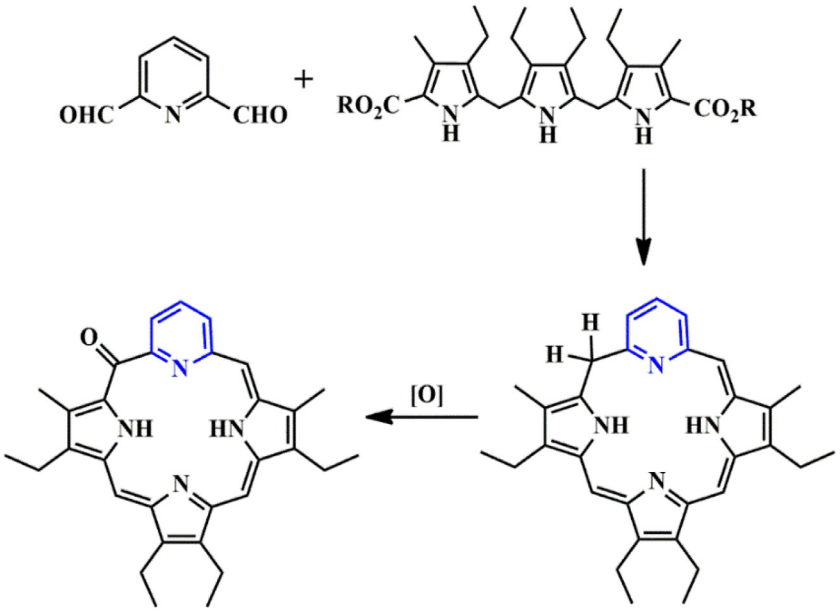
Fig. 4. Inverted Porphyrin isomer by Arnoff and Calvin [60].

In this light, many reports were presented with variety of modifications on porphyrin moiety. In 1994, Berlin and Breitmeir replaced one five-membered ring in the macrocycle with a larger heteroarene (Scheme 3) [67].

In the same year, two research groups independently reported the synthesis of N-confused porphyrins. Furuta et al. synthesized a phenyl-substituted derivative isomer of a porphyrin isomer, wherein they introduced slight change in the reacting species. The acid catalysed condensation of benzaldehyde and pyrrole was carried out. However, instead of using propanoic acid, t-BuOH/ CH_2Cl_2 (1:1) and concentrated HBr was used. The reaction yielded 5% of N-confused porphyrin along with 20% tetraphenylporphyrin (TPP) (Fig. 5) [68].

At the same time, Grazynski and coworkers synthesized a newly core-modified 2-aza-21-carba-5,10,15,20-tetra-p-tolyporphyrin ($CTTPH_2$) with 4% product yield (Fig. 6). The center of this macrocycle has remarkable ability as a tetradentate ligand to form carbon-metal bond. The spectral properties highlighted its potential application as a photosensitizer in photodynamic therapy [69].

These researches opened a new platform for the researchers to explore the *meso*, *β*- and inner-core of the porphyrin molecule. Some of these modifications are discussed in this review.



Scheme 3. Berlin et al. synthesis of modified porphyrin system [67].

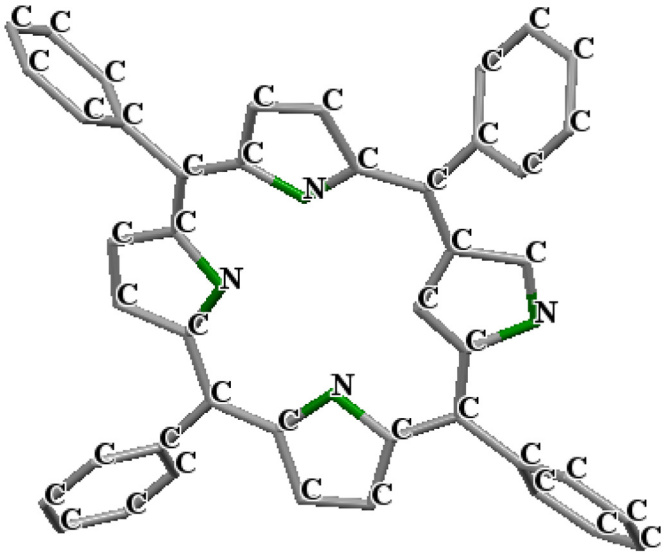


Fig. 5. X-ray crystal structure of N-confused porphyrin reported by Furuta et al. (hydrogens have not been shown for clarity purpose) [68].

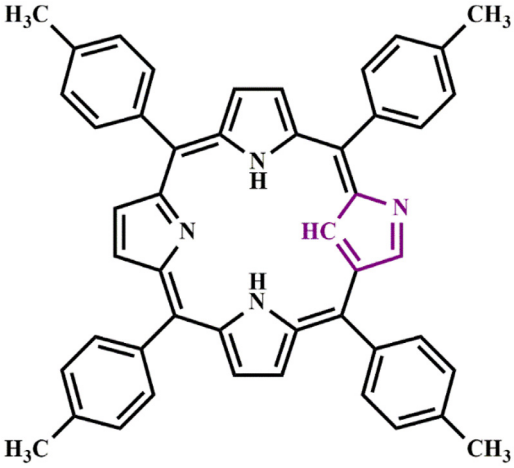


Fig. 6. Structure of N-confused porphyrin reported by Grazynski et al. [72].

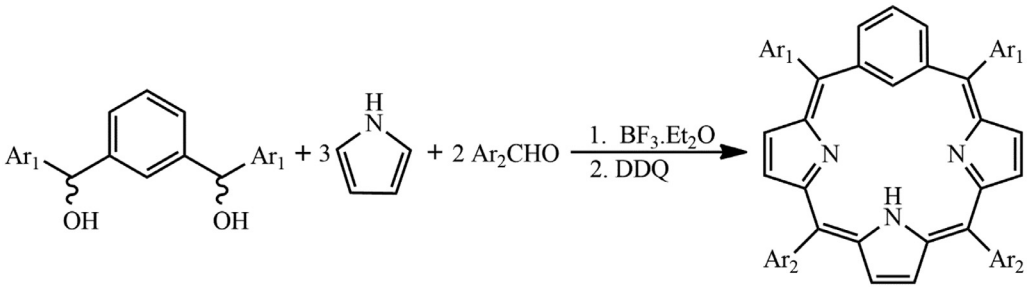
2.1. meta-*Benziporphyrin*

The presence of C–H moiety in the inner core gives higher possibility for stabilization of rare metal carbon interaction [70,71]. It is important to note that in carbaporphyrinoids due to the C–H moiety in the inner core, the π-electrons delocalization is deeply affected and in some cases it ceases completely. The loss of electronic current can be detected by NMR spectroscopy [72]. When a pyrrole ring in porphyrin macrocycle is outplaced by phenylene ring then it creates a subclass called benziporphyrin. Carbaporphyrinoids consisting of atleast one all-carbon ring is often distinguished from N-confused porphyrins by using the prefix true in front of the name of carbaporphyrin. This class of true carbaporphyrins is important to us as it provides feasibility to study the behavior of aromatic rings when surrounded by metal ions [73].

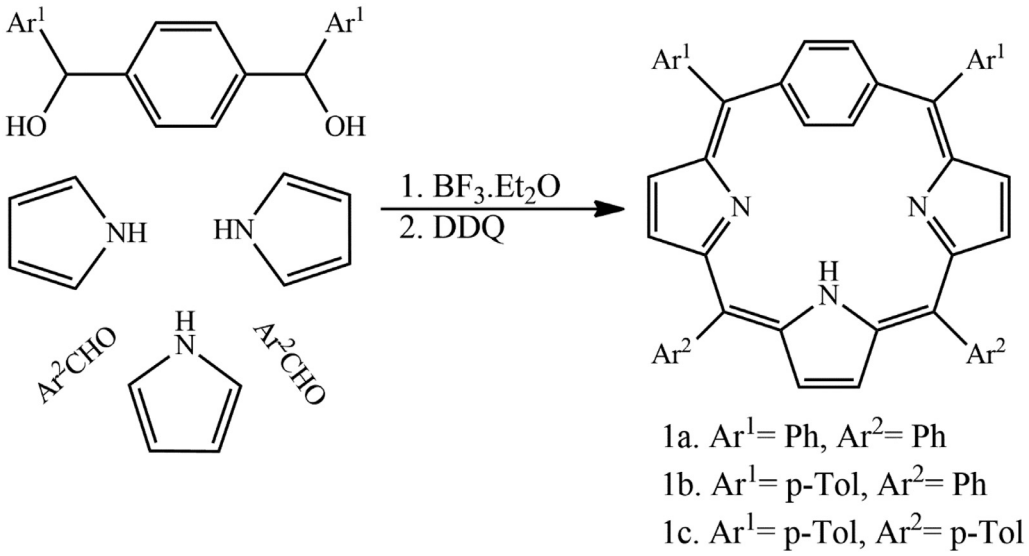
The idea of displacement of one of the pyrrole rings by another aromatic ring was first reported by Berlin et al. as discussed prior in this section. Synthetic analogs like *meso*– substituted benziporphyrins were later prepared for obtaining a pure sample by Martin Stepien, and Latos-Grazynski [74]. The synthesis involved condensation of a pyrrole, aromatic aldehyde and a precursor (α,α'-diol) followed by oxidation with DDQ, and further purification was done by chromatography, yielded 15% product (Scheme 4). The result of this experiment were satisfying considering the easy synthesis of diol precursor.

2.2. para-*Benziporphyrin*

The same was observed when the phenylene ring was incorporated in the system with 1 and 4 position of linkage, resulted in the new class of benziporphrins, called ***para*-benziporphyrins**. *para*-Benziporphyrin was first reported by Martin Stepien and Lechosław Latos-Grazynski in 2001 in synthesis similar to *meta*-benziporphyrin. The α,α'–dihydroxy–1,3–diisopropylbenzene in the synthesis was replaced by 1,4- isomer to



Scheme 4. Synthesis of meta-benziporphyrin analogue by Grazynski et al. [74]..



Scheme 5. Synthesis scheme for *para*-benziporphyrins. [78].

obtain *para*-benziporphyrin. However, the yield was reduced drastically to 1–3% (Scheme 5) [75].

Later, Hung and coworkers synthesized the hetero- analogue by using α, α'–dihydroxy–1, 4–diisopropylbenzene with 16-thiatripyrrane to give a 6% yield [76]. It is worth mentioning here that the *p*-phenylene ring in *para*-benziporphyrin was found out of the plane of three N atoms and is 45° tilted from the plane. A weak conjugation was also indicated by deviation in C–C bonds in *para* phenylene from normal benzene ring. The NMR study reveals that the *para* phenylene ring protons face different magnetic field at low temperatures while the difference end up at high temperature clearly indicating the rotation of *para* phenylene ring in a see-saw motion at higher temperatures. While at low temperature the outer proton 2 and 3 comes at aromatic region, the inside proton 21 and 22 comes upfield due to inner shielding ring current. Further, molecular and spectral study indicates that *para*-benziporphyrin behaves like extended conjugated macromolecules having aromaticity which differentiate it with *m*-benziporphyrins. It is also different from oxy-benziporphyrin as that achieves aromaticity because of semiquinone arrangement [77].

Several modifications in this molecule were then reported by various research groups. But, we now focus our studies on the partially reduced form of *meta*-benziporphyrin, i.e., *meta*-benziporphodimethenes.

3. *meta*-Benziporphodimethene analogues

meta-Benziporphodimethenes (*meta*-BPDM) systems were firstly synthesized by Martin Stępień during his works of Ph.D thesis [75]. These synthesized benziporphodimethenes were alkoxy and hydroxy analogues which were obtained through oxidation of mBPDM. *meta*-Benziporphodimethenes are actually modified *m*-benziporphyrins with two tetrahedral *meso* carbon atoms at 6- and 21- positions (Fig. 7) [69].

3.1. Synthetic route for meta-benziporphodimethene analogues

The prosthetic molecule *meta*-benziporphodimethene was first developed in serendipity by Stępień et al. while spectroscopically manifestating the metal-arene interactions observed in the diamagnetic zinc(II), cadmium(II), and mercury(II) benziporphyrin (both *meta* and *para* isomers) complexes and in the paramagnetic nickel(II) species. The synthesis of 6,6,21,21-tetraphenyl-11,16-bis(4-nitrophenyl)-*meta*-benziporphodimethene involved mixed condensation reaction between pyrrole, 4-nitrobenzaldehyde and α,α'–dihydroxy–1, 3–diphenylbenzene, in fixed ratios, in the presence of lewis acid and oxidants as shown in **Scheme 6**. The reaction yielded 14% of the product. They also reported metalation of the newly synthesized benziporphodimethene analogue, by insertion of Zn(II), Cd(II), Hg(II) and Ni(II) ions, by boiling a mixture of the free base and an appropriate salt, usually anhydrous chloride, in a chloroform solution (**Scheme 7**) [75]. The macrocycles designed in this study aimed to study different orientations of the arene with respect to the coordination center and different linkages between the arene fragment and the rest of the macrocycle, and to draw a comparison of benziporphyrin isomers with the benziporphodimethene molecules on the basis of NMR and UV–Vis analysis, which is discussed in the next Section 5.2.

Hung and co-workers developed a novel *meta*-benziporphodimethene based molecule (Scheme 8) in the year

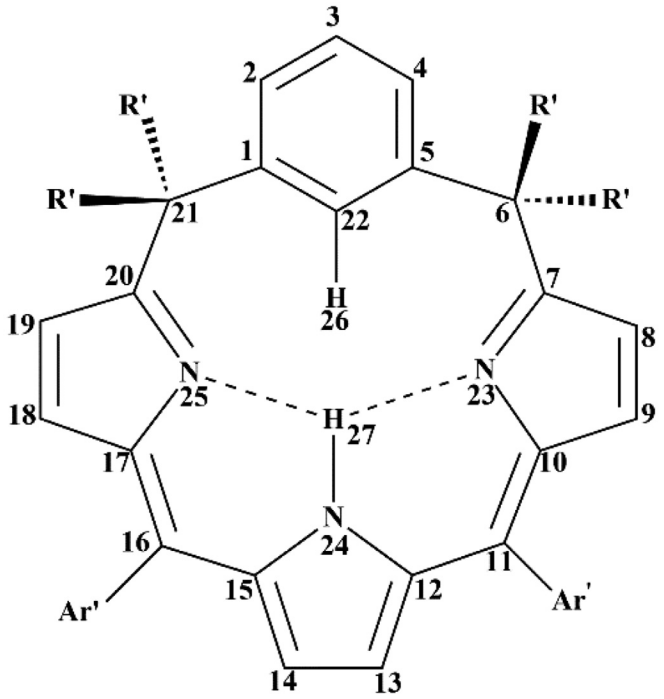


Fig. 7. Structure of freebase *meta*-benziporphodimethene [85].

2008, that was isolated as red solid with 27% product yield, that imparted no background emission and showed a turn-on probe for Zn^{2+} sensing [78]. This report aimed to traverse the applications of the newly synthesized analogue as a chemosensor (discussed in Section 3.3.3.). The synthesis was confirmed by UV–Vis spectroscopy (see Section 5.2). The same procedure as in Scheme 12 has been followed for metalation.

Subsequently, Hung et al. reported the formation of outer α-pyrrolic carbon oxygenated N-confused tetramethyl-*meta*-benziporphodimethenes isomers containing a γ-lactam ring in the macrocycle in addition to the product (**1,** in Scheme 9) formed in Scheme 8 (ref). Two isomeric (Scheme 9) carbonyl group of lactam ring either close to (O-Up, **2**) or away from (O-Down, **3**) the neighboring sp^3 *meso* carbon with product yield of 27%, 4% and 0.8%, respectively [79]. The compounds were characterized using UV–Vis, NMR, X-ray crystallographic analysis and DFT methods, discussed in later sections in this paper.

The synthetic pathway for the *meta*-benziporphodimethene remained the same but researchers attempted several modifications at *meso* carbon atoms (C11, C16 in Scheme 10) of this novel molecule and advancements in its applications were explored. A

Table 1
Consolidated UV data for Scheme 10. [80].

Compound	Wavelength (λ_{max}) in nm			
	Freebase TMBPDM		Zn^{2+} TMBPDM complex	
	Soret Band	Q-Band	Soret Band	Q-Band
1	353	517–548	351	594–639
2	351	513–542	350	594–638
3	321	508–538	348	602–649
4	320	507–533	346	598–642
5	319	508–531	346	600–642
6	342	509–539	347	594–641
7	378	513–545	343	592–643
8	376	513–546	346	598–642

similar synthesis was intended by Kumar et al. in 2016, wherein they used substituted aldehydes, pyrrole and α, α'-dihydroxy-1,3-diisopropylbenzene (Scheme 10) with 15–17% product yield [80]. The research group confirmed the product synthesized using UV–Vis, NMR, mass and X-ray crystallographic analysis. Quantum yield, fluorescence studies and biocell imaging applications were also prospected.

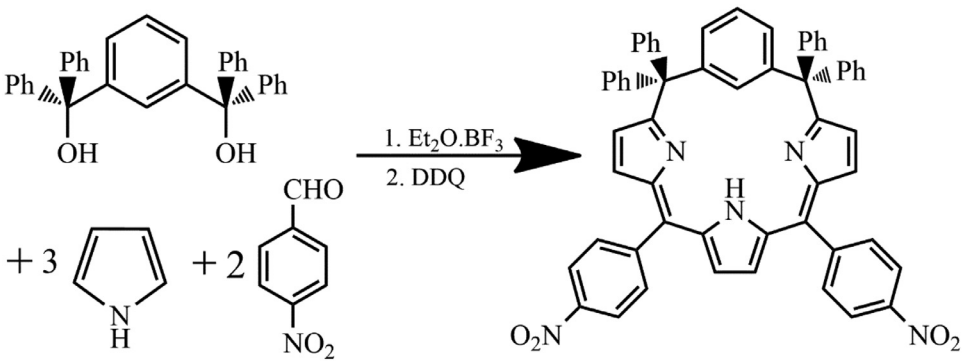
3.2. Spectroscopic analysis

3.2.1. UV–Vis Spectroscopy

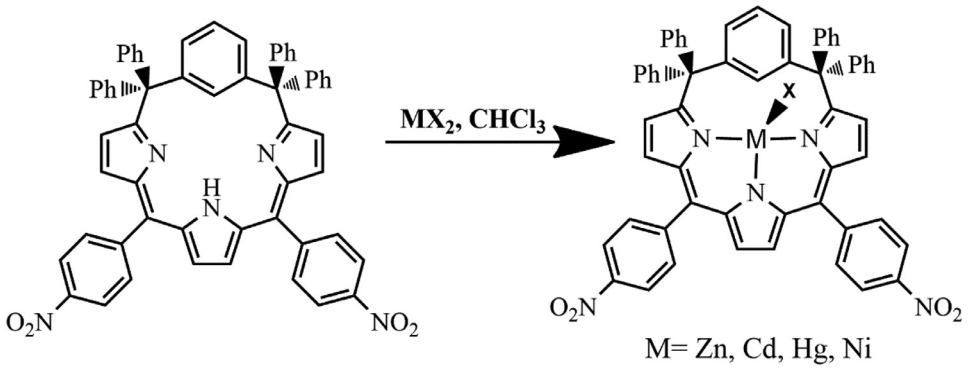
meta-Benziporphodimethenes are generally red color compounds. Owing to its discrete conjugated system, compound shows a broad UV–Vis spectrum. In *meta*-benziporphodimethenes (in Scheme 8), a high energy Soret band is detected at 349 nm and lower energy Q band is found at about 514 and 542 nm approximately (Graph 1) [78]. This compound is non-fluorescent in freebase form and upon addition of Zn^{2+} (Scheme 7), an instant change in solution color from pink to greenish blue was observed (Fig. 8) [80].

The UV–Vis spectra is broad in nature due to the presence of sp^2-sp^3 mixed *meso* carbon atoms rendering non-conjugated system. In the report presented by Kumar et al. (Fig. 9), the Soret like high energy band was observed aro und 342–378 nm for *para* and *ortho*- substituted meso–phenyl groups n freebase *meta*-benziporphodimethenes, **1–8** in Scheme 10. The electron releasing substituent at *para* position showed considerable reduction in Soret like high energy band at 378 and 376 nm, **7,8** in Scheme 10. On the other hand, ortho-halogenation significantly enhance this high energy band emerging at 321, 320 and 319 nm for **3, 4, 5** in Scheme 10 (Table 1) [80].

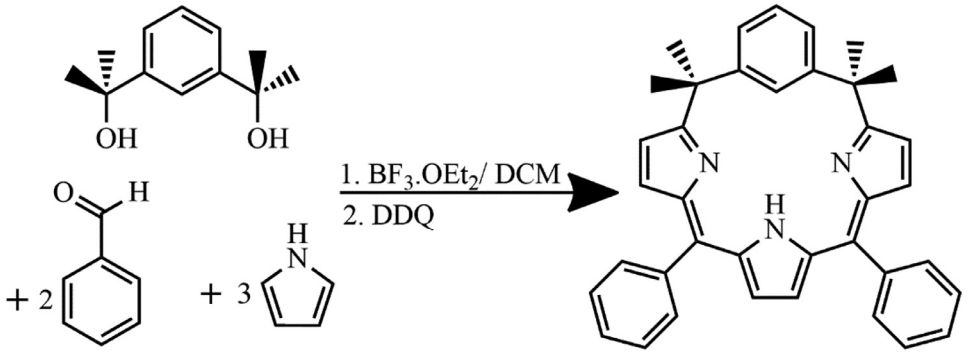
Upon metalation of these freebase meta-benziporphodimethene analogues **1–8** in Scheme 10, a significant red shift is observed which is due to the strong interactions of d-orbital of zinc (II) with the macrocycle in the ground state absorption spectra of zinc



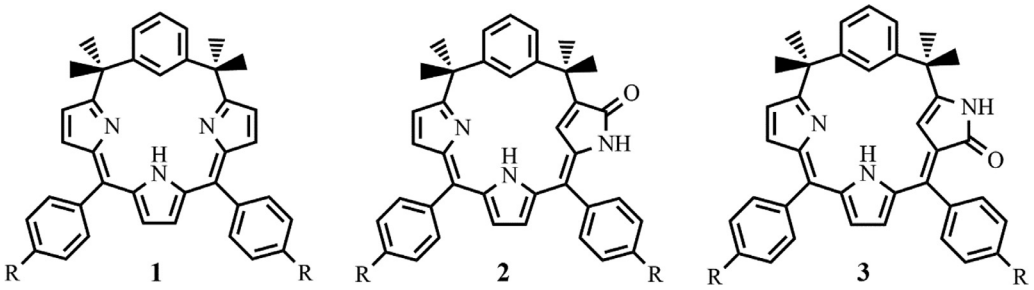
Scheme 6. Synthesis of *meta*-benziporphodimethenes in 2004 [69].



Scheme 7. Metalation of 6,6,21,21-tetraphenyl-11,16-bis(4-nitrophenyl)-*meta*-benziporphodimethene reported in 2004 [69].



Scheme 8. Synthesis of 11,16-Bis(phenyl)−6,6,21,2-tetramethyl-m-benzi-6,21-porphodimthene by Hung et al. [78].



Scheme 9. Tetramethyl-meta-benziporphodimethene (**1**) molecule along with its O-Up (**2**) and O-Down (**3**) isomer, R=H, COOMe [82].

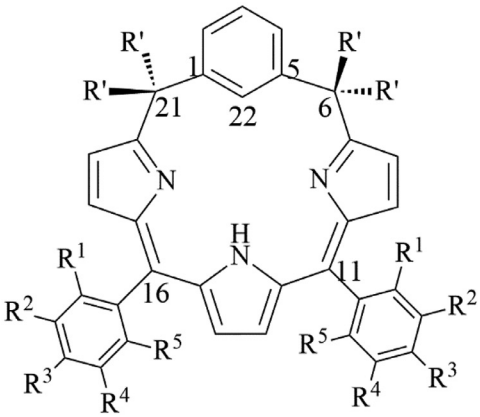
complexes of these compounds with respect to their free-bases (Fig. 10).

3.2.2. Structural analysis

The first ever crystal structure of *meta*-benziporphodimethene zinc complex was reported in 2008 by Hung et al. The analysis monosemously confirmed the coordination of zinc ion with benziporphodimthene through the three pyrrolic nitrogens and an axial chloride (Table 4, 2-ZnCl) [78].

Later, Hung and coworkers came up with crystal structures of freebase tetramethyl-*meta*-benziporphodimethene (TMBPDM) along with its two isomers (Scheme 9) as discussed in Section 3.1. X-ray analysis shows that TMBPDM are more distorted than the other two isomers. The average deviation of 25 atoms on the macrocycle from a mean tripyrrin plane (plane through 17 atoms i.e. atoms of 3 pyrrole rings and 2 sp^2 hybridised meso–carbon atoms) for **1** (TMBPDM) is 1.044 Å while the values for **2** (O-Up Isomer) and **3** (O-Down Isomer) are found to be 0.680 Å and 0.630 Å, respectively. The angle between the CNNN inner-core plane and the macrocyclic phenylene plane is 70.35° for **2** and 69.76° for **3**, but in case of **1**, the angle between phenylene plane and three pyrrolic nitrogen plane is only 52.39° [79]. Another interesting feature in the crystal structure is that the angle between the vector along two sp^3 *meso*–carbons and the plane, defined by the mean plane of 17 atoms on a tripyrrin moiety, is 28.51° for **1**, which is significantly larger thans 15.79° and 14.34° for **2** and **3**, respectively (Fig. 11). These data are in good agreement with the fact that **2** and **3** are less puckered than **1** with phenylene rings oriented vertical to the mean plane of the macrocycle (Fig. 12) [79].

Subsequently, Kumar et al. determined the crystal structure of free base 11,16-Bis(3,4,5-trimethoxy-phenyl)-6,6,21,21-tetramethyl-*meta*-benziporphodimethene (**8** in Scheme 10). The geometry of **8** was found lesser distorted than earlier reported structures of *meta*-benziporphodimethenes (Fig. 13). The average deviation of atoms which constitute the macrocycle was measured to be 0.646 Å which is sufficiently lesser as compared to 1.044 Å for free base *meta*-benziporphodimethenes earlier reported by Hung and co-workers. The angle between the phenylene ring plane and three pyrrolic nitrogens of macrocycle is 46.63° which is remarkably lesser. Important and interesting observation in the crystal structure is that the angle between the plane formed by 17 tripyrrin atoms and the vector along the two sp^3 *meso* carbons is only 15.45°. These observation confirms that obtained structure of com-

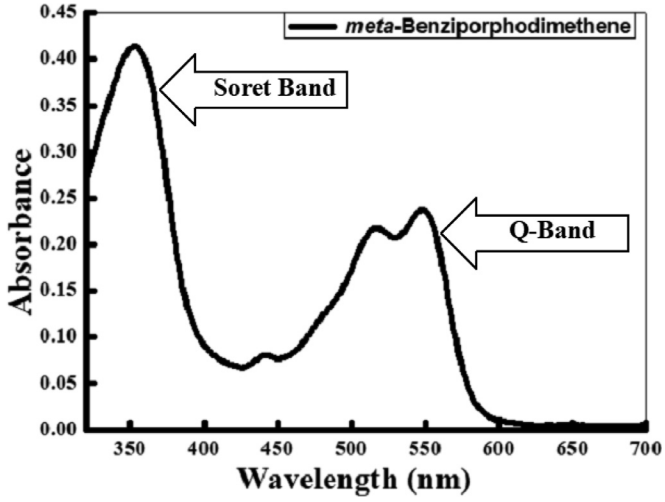


where, R'=CH_3
(1) R^1=R^2=R^4=R^5=H, R^3=Cl
(2) R^1=R^2=R^4=R^5=H. R^3=Br
(3) R^1=R^5=Cl, R^2=R^3=R^4=H
(4) R^1=R^5=F, R^2=R^3=R^4=H
(5) R^1=R^2=R^3=R^4=R^5=F
(6) R^1=R^2=R^4=R^5=H, R^3=CF_3
(7) R^1=R^2=R^4=R^5=H, R^3=OCH_3
(8) R^1=R^5=H, R^2=R^3=R^4=OCH_3

Scheme 10. *Meso*-substituted *meta*-benziporphodimethene derivative by Kumar et al. [80].

Table 2
Consolidated NMR data for Scheme 10. [80].

Compound	N-H(H27,Py_o)	Ar–H(H26)	Ar-H(C2, C3, C4, C11, C16)	β-Py_1, Py_2–H(C8, C9, C18, C19)	β-Py_o-H(C13, C14)	OCH_3(R^2, R^3, R^4)	sp^3-CH_3(C6, C21)
1	12.32 (br, s)	7.93(s)	7.24–7.39 (m)	6.81 (ab quartet)	6.15 (s)	–	1.72 (s)
2	12.32 (br, s)	7.94(s)	7.31–7.57 (m)	6.83 (ab quartet)	6.16 (s)	–	1.74 (s)
3	12.16 (br, s)	8.06(s)	7.28–7.41 (m)	6.59 (d)–6.82 (d)	5.86 (s)	–	1.81 (s)
4	12.18 (br, s)	7.97(s)	6.91–7.45 (m)	6.70 (d)–6.83 (d)	6.00 (s)	–	1.79 (s)
5	12.08 (br, s)	7.89(s)	7.25–7.34 (m)	6.65 (d)–6.90 (d)	5.98 (s)	–	1.78 (s)
6	12.00 (br, s)	7.92(s)	7.68 (d,4H-ArH), 7.56 (d,4H-ArH), 7.29 (m,3H-ArH)	6.75 (d)–6.85 (d)	6.07 (s)	–	1.75 (s)
7	12.38 (br, s)	7.93(s)	7.44 (d, 4 Ar-H), 7.27–7.32 (m, 3H,2,3,4 Ar-H), 6.95 (d, 4 Ar-H)	6.77 (d)–6.88 (d)	6.30 (s)	3.87 (s)	1.69 (s)
8	12.33 (br, s)	7.92(s)	7.30(q, 3H 2,3,4 Ar-H) + 6.74(s, 4Ar-H)	6.80 (d)–6.95 (d)	6.35 (s)	3.92 (s,6H, para OMe) + 3.83 (s,12H, meta OMe)	(s)



Graph 1. UV Spectra of *meta*-benziporphodimethene [78].

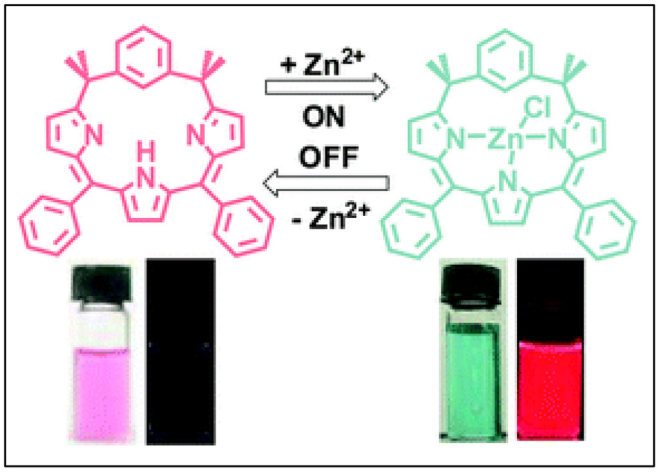


Fig. 8. Color change in *meta*-benziporphodimethene upon metalation using $ZnCl_2$ salt [81].

pound **8** is lesser puckered than the earlier reported structures of *meta*-benziporphodimethenes [80].

In the year 2011, Hung et al. reported crystal structures of group 12 metal complexes (Zn, Cd, Hg) of TMBPDM molecules (Fig. 14) [73]. The comparative studies suggested that the structure of free-base molecule is highly distorted while a flat pyrrin moiety with nearly coplanar pyrrolic nitrogen atoms is observed for all its metal chloride complexes. Further, the central metal ion was found to be tetra-coordinated with three coordinating pyrrolic nitrogen atoms and one chloride ion that sits at axial position of the TMBPDM metal chloride complexes as seen in the crystal structures below. They also reported that the apical deviation value for metal to the N_3 plane (plane consisting of 3 pyrrolic nitrogen atoms, i.e., N23, N24, N25) for TMBPDM core is remarkably lesser compared to benziporphyrin complexes. this in turn reflected the larger central core of TMBPDM relative to benziporphyrins. The authors then studied the geometry of TMBPDM through calculating the angles between the plane passing through the phenylene ring (plane through C1, C2, C3, C4, C5, C22) and N_3 plane. Interestingly, for all the metal complexes, the phenylene ring was nearly perpendicular to the N_3 plane (approx. 81° to 92°) (Fig. 14). The angle increased with the metal ion size. Although, as discussed earlier

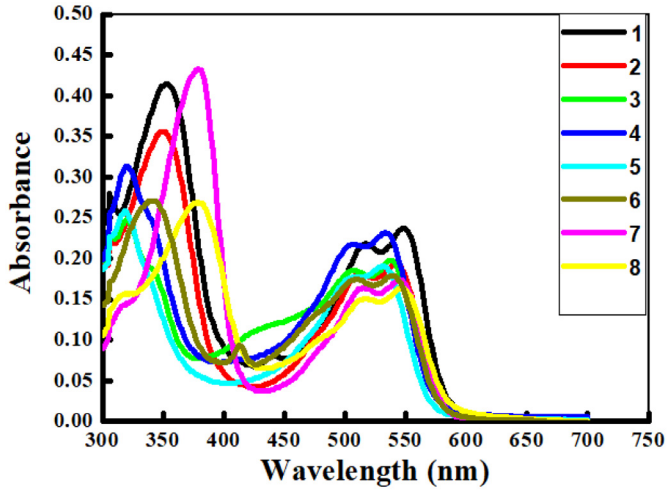


Fig. 9. UV–Vis Spectra for freebase *meta*-benziporphodimethene analogues (Scheme 10).

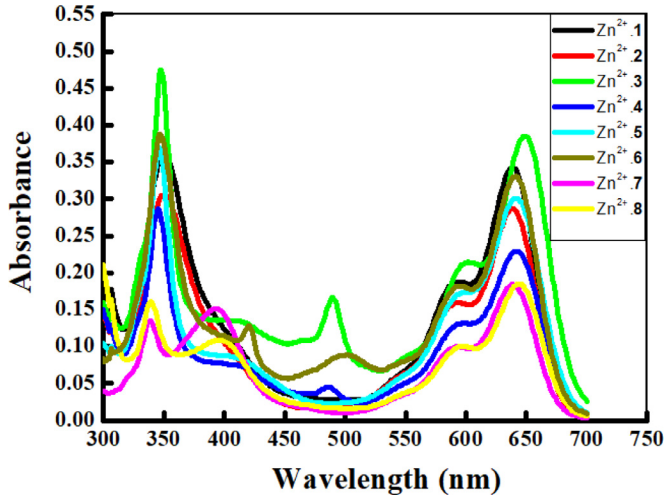


Fig. 10. UV–Vis Spectra of *meta*-benziporphodimethene, **1–8** (Scheme 10) zinc chloride complexes [80].

in this section, this angle lies between 46° to 52° for freebase TMBPDM (Fig. 15). Thus, it may be said that the geometry of freebase TMBPDM is highly distorted relative to its metal complex. The organometallic bond (M-C bond) distance in TMBPDM complexes is much larger than the direct M-C bond linkage, but significantly less than the sum of their vander waal radii. This hinted the presence of metal-arene interactions. The M-C bond distance is also affected by the electronic properties of TMBPDM systems. For example, when an electron withdrawing group is placed at 11,16- position, the organometallic bond distance was relatively lesser than electron donating substituent. Also, the metal-arene interaction was found to be weaker in TMBPDM compared to *m*-benziporphyrins. For instance, the organometallic M•••C bond distance for TMBPDM Cd(II) complexes are nearly 2.831 Å which is longer than that for Cd(II) m-benziporphyrin complexes viz. 2.716 Å. The authors also introduced the *syn* and *anti* conformation in TMBPDM complexes, which are expected to be present in some ratios depending upon the factors stabilizing the conformations. The *syn* form refers to the position of H26 atom (Fig. 7) aligned parallel or in the same direction as axial chloride ion (Fig. 15, (i)), while in *anti* form H26 atom is aligned opposite to the axial chloride ion (Fig. 15, (ii)) [73].

Besides Group 12, they also reported dinuclear and tetranuclear silver (I) TMBPDM complexes. the use of silver nitrate salt for metalation purpose yielded novel dimer and tetramer of TMBPDM silver complexes (Fig. 16). The dimeric TMBPDM-Ag binds together with a weak η^2- type Π-coordination from β-pyrrolic double bond C13=C14 to Ag ion (Fig. 16, a). In Ag_4 complex, apart from Ag–Ag bond, the silver ion is reportedly coordinated by Ag-N (pyrrole), nitrate oxygen atoms and 2,6-lutidine to give this unconventional complex (Fig. 16, b). It is notewrorthy that the bond distance between Ag–C22 in 2-Ag_2 is significantly shorter relative to all other metal chloride complexes of TMBPDM. The 2-Ag_4 fetches an usual triangular Ag–Ag–N moiety. The direct Ag–Ag linkage in this cluster is longest among all the crystal structures with Ag–Ag–N core [73].

3.2.3. NMR shifts in meta-benziporphodimethenes

Interestingly, the two N-confused isomers exists as dimer forming intermolecular hydrogen bonding through lactam-amide bond. The dimerization was confirmed by taking the concentration dependent NMR spectra in $CDCl_3$. It was observed that with increasing concentration, the outer N – H proton comes more downfield as the extent of hydrogen bonding increases. The impact was observed more dominant in the O-Down isomer than in the O-Up isomer. This shows that O-Down isomers have more effective intermolecular hydrogen bonding than the O-Up isomer [79].

meta-benziporphodimethenes are non-aromatic as the ring current attenuated due to tetrahedral meso–carbon atoms. The identical character is noticed in the NMR spectra of these compounds as the chemical shift of inner N–H proton appears at δ ~12.0, signifying the absence of internal shielding by ring current. The methyl

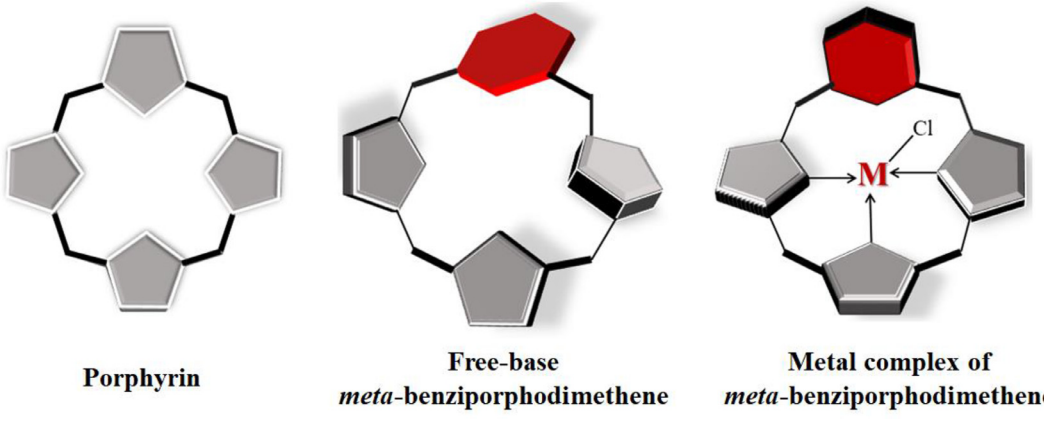
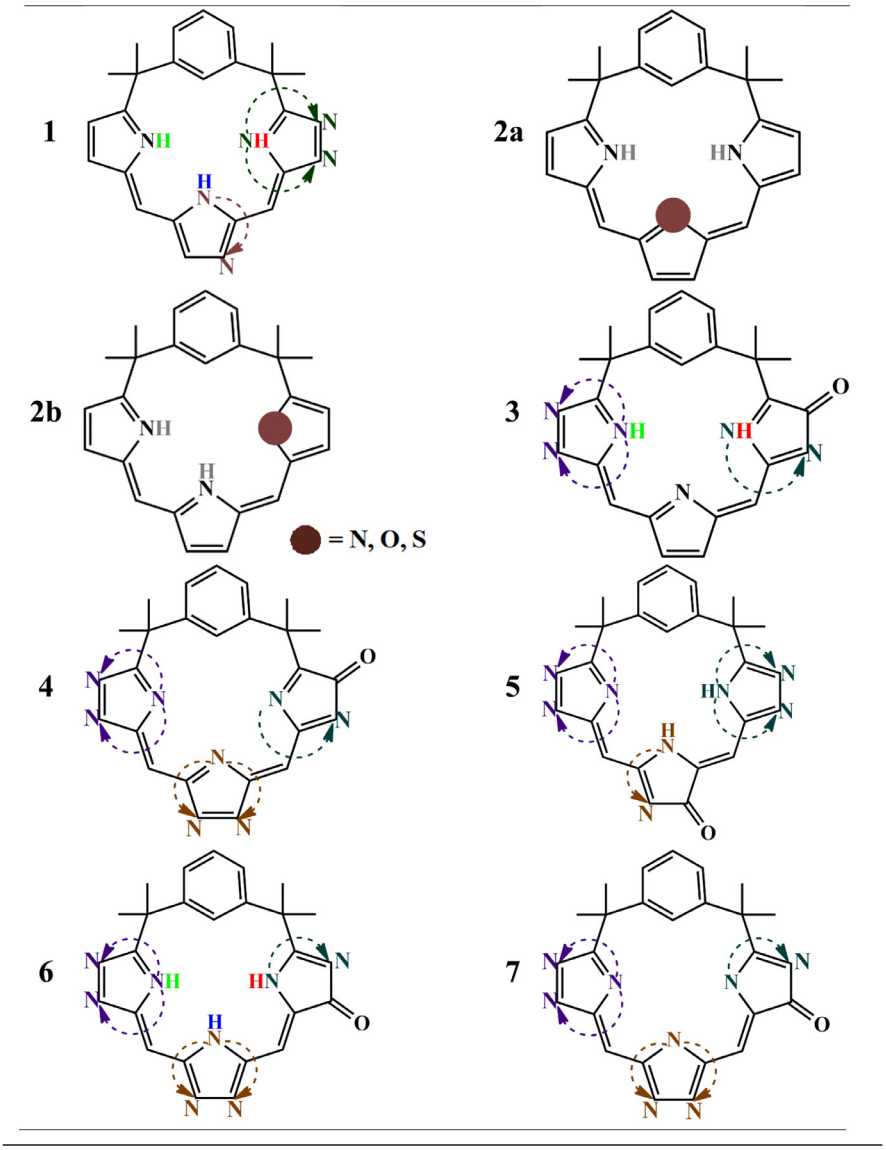


Fig. 11. Depiction of core geometry of porphyrin molecule, non-planar *meta*-benziporphodimethene molecule and its metal complex (M = Zn, Cd, Hg).

Table 3
Overview of TMBPDM analogues studied by Rana et al. [84].



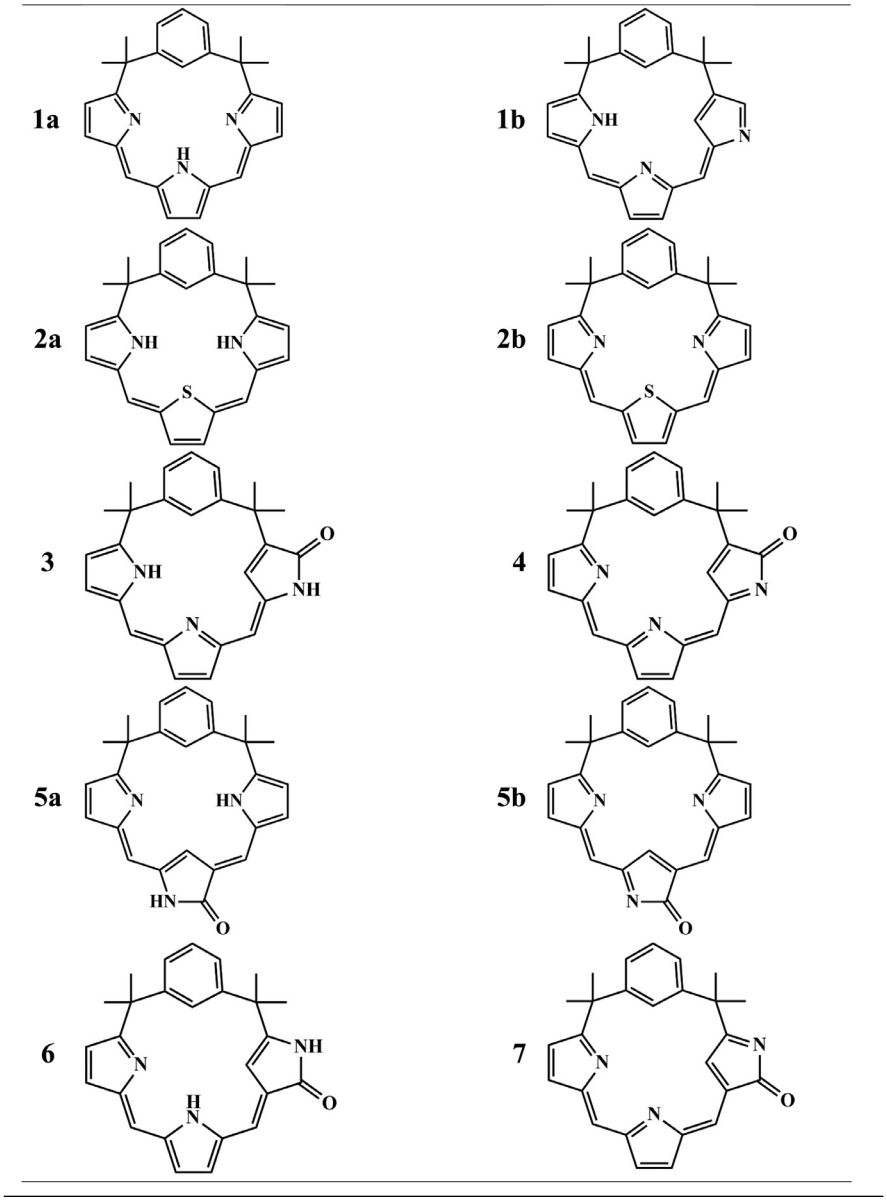
peaks of tetrahedral group appear at ~1.7 ppm. The pyrrolic protons were observed at δ ~6 −7 ppm. The pyrrolic protons 13 and 14 appear as singlets and come at ~6.0 ppm, the other pyrrolic protons 9,18 and 8,19 come slightly downfield at around 6.5 ppm and give a 'ab quartet' pattern. The aromatic protons of *meso* aryl substituents and of *m*-phenylene ring appear in between ~7–8 ppm as multiplet. The C(22)-H comes at around 8.0 ppm and also shows the absence of inner shielding ring current [80].

3.2.4. Electronic structure study

Two major theoretical studies have been reported for *meta*-benziporphodimethenes by Rana et al. in 2019, they presented a DFT overview of core-modified *meta*-BPDM analogues and their N-confused derivatives consisting γ-lactam ring as discussed above. They primarily intended to study the effect on the geometries and relative energies of *meta*-BPDM by varying the positions of N-confused ring and oxygenation at one of the pyrrole rings (Table 3) [81]. For clarity purpose, the phenyl ring at 11, 16- positions have not been shown in the table below.

The tautomer of 1 (in Scheme 9) and its inverted compounds obtained via single pyrrolic α,α- to α,β-linkage isomerization has been discussed. The parenthesis of this study has been shown as structure 1 in Table 3. The dashed arrows signifies the changes in the position of N atoms, which basically is due to inversion of pyrrole ring. Varied colours of H-atoms in the core illustrates only one H is present in the molecule at a time. The position of H on N atoms (N23, N24, N25) have been jumbled relative to N atoms. It has been found when H27 is bonded to N24, the molecule is energetically more favourable (1a, Table 4) compared to when it is attached to N23 and N25 (see Fig. 7). They, thus stated that the inner core H-bonding interactions plays a crucial role in stabilizing the tautomer. Although, the inversion of pyrrole ring leads to increment in relative energy of the molecule, when N23 or N25 are nearer to C11 or C16 centres respectively, the system are said to be more stable by 1.17 kcal/mol than when the respective N atoms are nearer to C6 or C21 centres (or sp^3 hybridised carbon atoms). In other words, N-down configuration is more stable compared to N-up configuration (1b, Table 4). It has also been stated,

Table 4
Most Stable structures of all the tautomers discussed in Table 3. [84].



N-confused isomers denoted by green dashed arrows in 1, are more favourable than N-confused isomers denoted by brown dashed arrow. As per their results, the inverted pyrrole ring opposite to phenylene ring (pyrrole with N24 atom) either breaks the π- electron delocalisation or nullifies the H-bonding interactions. The comparison of furan and thiophene analogues placed in place of N24 pyrrole ring (2a and 2b, Table 3) has also been studied. These modified analogues are said to be more stable due to the presence of O•••NH interactions that lowered the energy. The structure with thiophene ring having absence of hydrogen in the inner core are more symmetrical compared to their oxa and aza analogs (2a, 2b, Table 4). It has also been showed that thia and oxa analogues having unprotonated pyrrole rings have larger dipole moment than their corresponding aza analogue. Structure 3 in Table 3 represents isomers with or without inverted pyrrole ring in the presence of O-up oxygenated pyrrole ring in its reduced form, while structure 4, Table 3 depicts the oxidized form of the same. In the similar fashion, O-down isomers of meta-benziporphodimethene analogues have been compared varying the linkages of pyrrolic N- and H-atoms, denoted by dashed arrows and colours, respectively, in structure 6 and 7, Table 3. As per their research, the O-up isomers are more favourable than O-down (see most stable forms of 3, 4, 6, 7 in Table 4). This further explains the reported yield of O-up and O-down isomers discussed in prior section, as 4% and 0.8% yields, respectively. Also, isomers with inner-protonated pyrrole rings are more favourable and have minimum energy than the unprotonated forms. Another sequence of isomers is formed by oxygenation of N24 pyrrole ring in its oxidized and reduced form (Structure 5, Table 3). the most stable configuration in this case was found when N24 atom is present adjacent to the oxygenated center, in the oxidized form of Structure 5 (no inner H present, 5b, Table 4). This was explained by the doubly CH•••N interactions present in it. While the other tautomers undergo single CH•••N interactions. The enhanced stability of these structures was due to the extended π-conjugate systems (12π electrons) and inner H-bonding network, in case of reduced form of structure 5, Table 3 (5a, Table 4) [81].

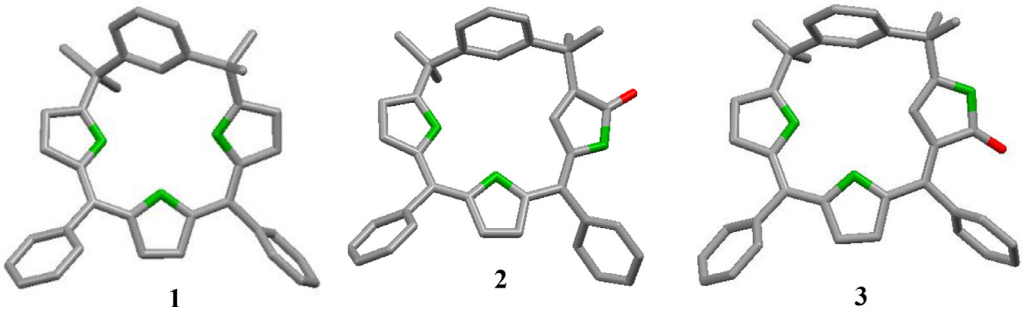
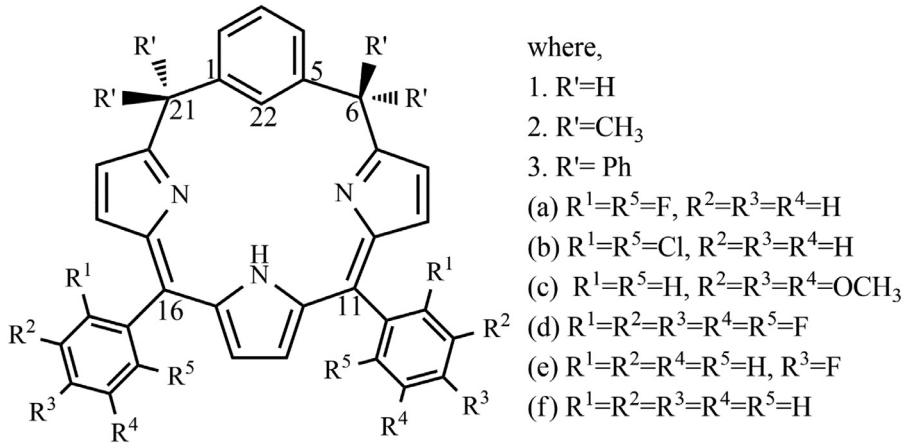


Fig. 12. Molecular model of 11,16-Bis(phenyl)-6,6,21,21-tetramethyl-*m*-benzi-6,21-porphodimethene (**1**) along with its O-Up (**2**) and O-Down (**3**) isomers. Crystal structure coordinates were taken from ref. [79].



Scheme 11. *meta*-benziporphodimethene systems studied [85].

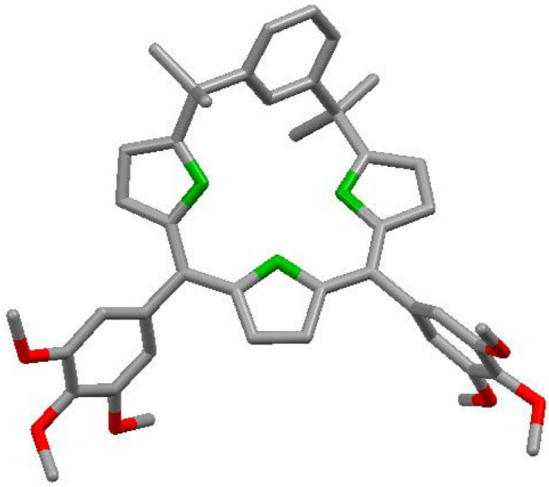


Fig. 13. Molecular model of 11,16-Bis(3,4,5-trimethoxy-phenyl)-6,6,21,21-tetramethyl-*meta*-benziporphodimethene. Crystal structure coordinates were taken from [80].

Further, Ahluwalia et al. recently reported another study to find out the effect of substituents at sp^2 and sp^3 *meso* carbons, on the geometry and energy of these molecules (Scheme 11) [81].

The substitution at these positions affects the inner geometry (cavity size) and the classical hydrogen bond (HB) strengths (Fig. 17)The hydrogen bond strength was estimated using molecular tailoring-based approach (Fig. 18) using density functional theory. Theoretical studies showed that on increasing the bulkiness of sp^3 *meso* positions, the puckered structure undergoes better conventional hydrogen bonding. It is also concluded that electron-withdrawing groups at sp^2 *meso* phenyl stabilize the molecule more than electron-donating ones. This study is useful for perceiving the factors regulating the cavity size in *meta*-benziporphodimethene analogues that can further enhance the ease in metalation of these moieties [82].

The calculated H-bond energies suggested that intramolecular H-bonds in these molecules are moderately stronger (6.5 to 10.5 kcal mol^{-1}). It was also found that the bulky substituents at sp^3 *meso* carbon increased the intramolecular H-bond strength (**Ph>CH_3>H**) and hence strengthen the core of these molecules. On the other hand, ortho-chlorination ((b) in Scheme 11) and penta-fluorination ((d) in Scheme 11) of aromatic ring at sp^2 *meso* position leads to increase in the H-bond strength. It was thus concluded that the presence of electron withdrawing substituents at sp^2 *meso* phenyl strengthens the inner core and stronger intramolecular hydrogen bonding occurs. The obtained HB distances and HB strengths of the optimized molecules are consolidated in the form of a table below (Table 5) [82].

With help of these studies are useful for future designing or predicting experimental work with these new systems.

3.3. Properties of metallated- meta- *benziporphodimethene complexes*

3.3.1. Weak metal-arene interaction

One of the most important property of metal tetramethyl-*meta*-benziporphodimethene complexes is the presence of metal arene interaction in them. *meta*-Benziporphodimethenes have the advantage to activate the C–H bond of *meta* phenylene ring in the vicinity of transition metal atoms and that gave it a new horizon to explore. X-ray structure analysis has confirmed the weak metal–arene interactions in cadmium and nickel complexes of m-

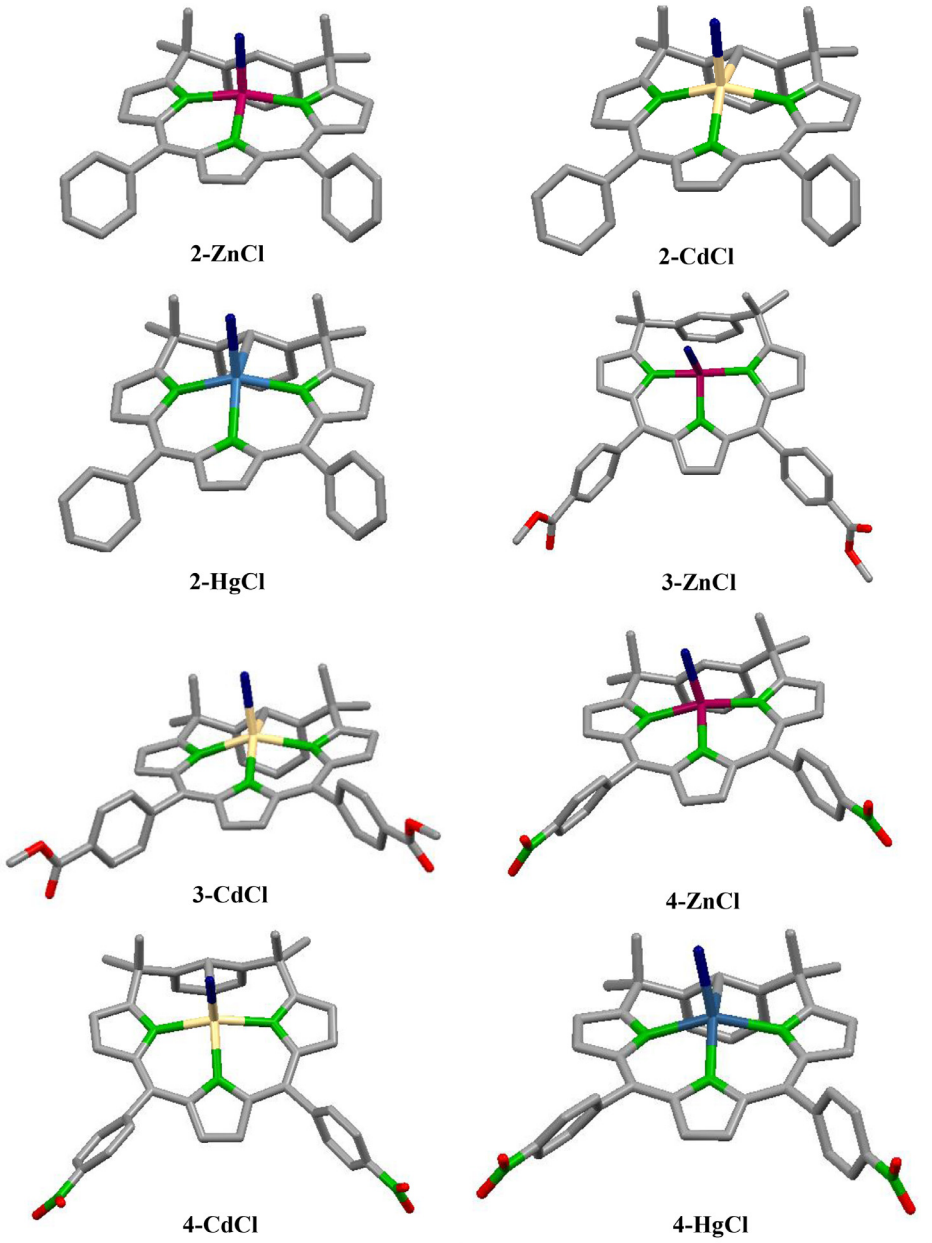


Fig. 14. Crystal structure of Group 12 metal complexes of *meta*-benziporphodimethene analogues [73].

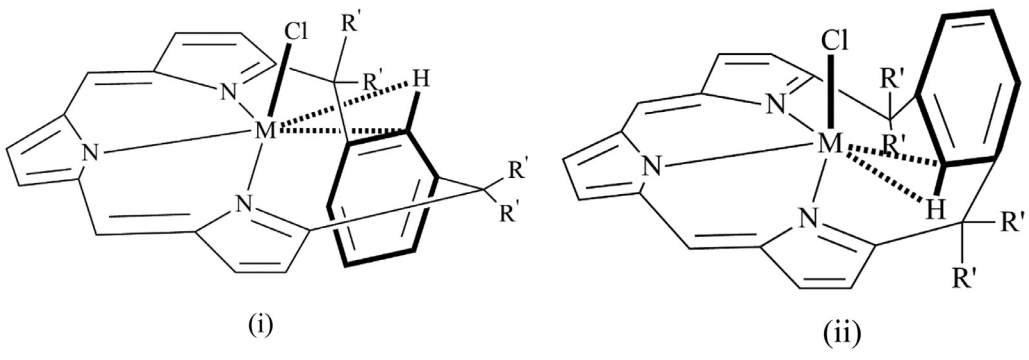


Fig. 15. Representation of (i) *syn*- conformational, and (ii) *anti*- conformational isomers of *meta*-benziporphodimethene [82].

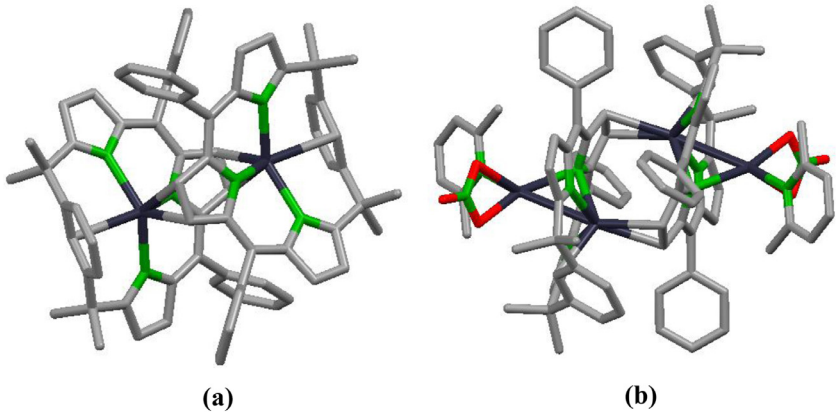


Fig. 16. Overview of theoretical studies performed by Ahluwalia et al. on *meta*-BPDM analogues [82].

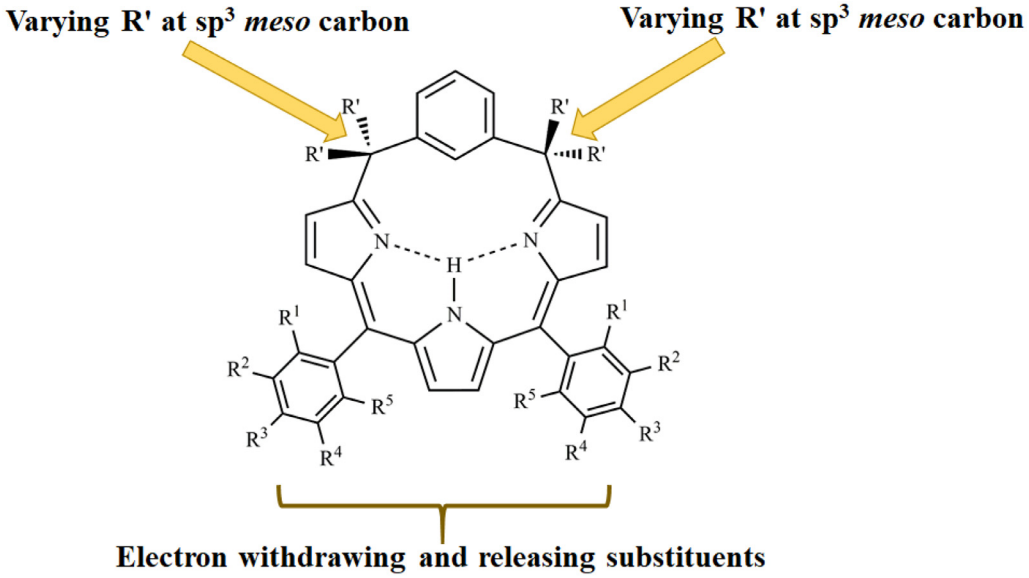


Fig. 17. Fragmentation approach used to analyze the intramolecular Hydrogen Bond strength [81,82].

Table 5
Intramolecular hydrogen bond (H-Bond) energy (in kcal mol^{-1}) and distance (in Å). See Scheme 11 for geometries of the molecules [82].

Molecule	H-bond Distance (Å)		H-Bond Energy (kcal mol^{-1})	
	HB1[a]	HB2[b]	HB1[a]	HB2[b]
1(a)	2.234	2.254	8.64	7.96
1(b)	2.228	2.259	8.83	8.21
1(c)	2.264	2.235	7.13	7.69
1(d)	2.231	2.256	8.66	8.04
1(e)	2.2	2.264	7.83	7.22
1(f)	2.131	2.389	10.05	5.76
2(a)	2.467	2.206	8.05	10.15
2(b)	2.303	2.303	7.83	7.83
2(c)	2.216	2.217	6.67	6.65
2(d)	2.206	2.206	7.51	7.51
2(e)	2.216	2.215	6.75	6.76
2(f)	2.420	2.045	4.36	10.14
3(a)	2.212	2.212	7.64	7.63
3(b)	2.206	2.206	8.64	8.64
3(c)	2.065	2.382	10.33	5.69
3(d)	2.212	2.212	6.79	6.82
3(e)	2.371	2.078	5.10	0.78
3(f)	2.341	2.095	6.22	10.07

[a] HB1 corresponds to N25–H27 bond.
[b] HB2 corresponds to N23-H27 bond.

benziporphyrin, with distances shorter than the sum of van der Waals radii [69]. The atypically large intramolecular through-space scalar couplings between the cadmium nucleus and the arene protons in *m*-benziporphyrin Cd(II) complexes as well as the downfield chemical shifts as a result of spin transfer from nickel center to an agostic proton in the paramagnetic *m*-benziporphyrin Ni(II) complexes provide solid supporting evidences for the presence of the metal-arene interaction. Fascinatingly, the strength of the metal-arene interaction is extremely dependent on the conformation of the arene moiety. As in the case of *meta*-benziporphyrins, this kind of interaction not only helps in the activation of phenylene ring but also in the stabilization of metal complexes [69,73].

3.3.2. Fluorescence enhancement

An important property of these compounds is that they are non-fluorescent in free-base form but turn on fluorescence upon zinc metallation with no background emission [54]. The most important part of this finding is that *meta*-benziporphodimethenes act as a specific sensor for zinc ions [80,83]. Very less fluorescence is observed for cadmium and mercury ions comparatively. Since the fluorescence "switch on" upon zinc metallation and hence this kind of fluorescence comes under the category of "**chelation enhanced fluorescence"(CHEF)**. CHEF sensors are mostly based on fluorophores, such as fluorescein [84,85], dansyl and anthracene,

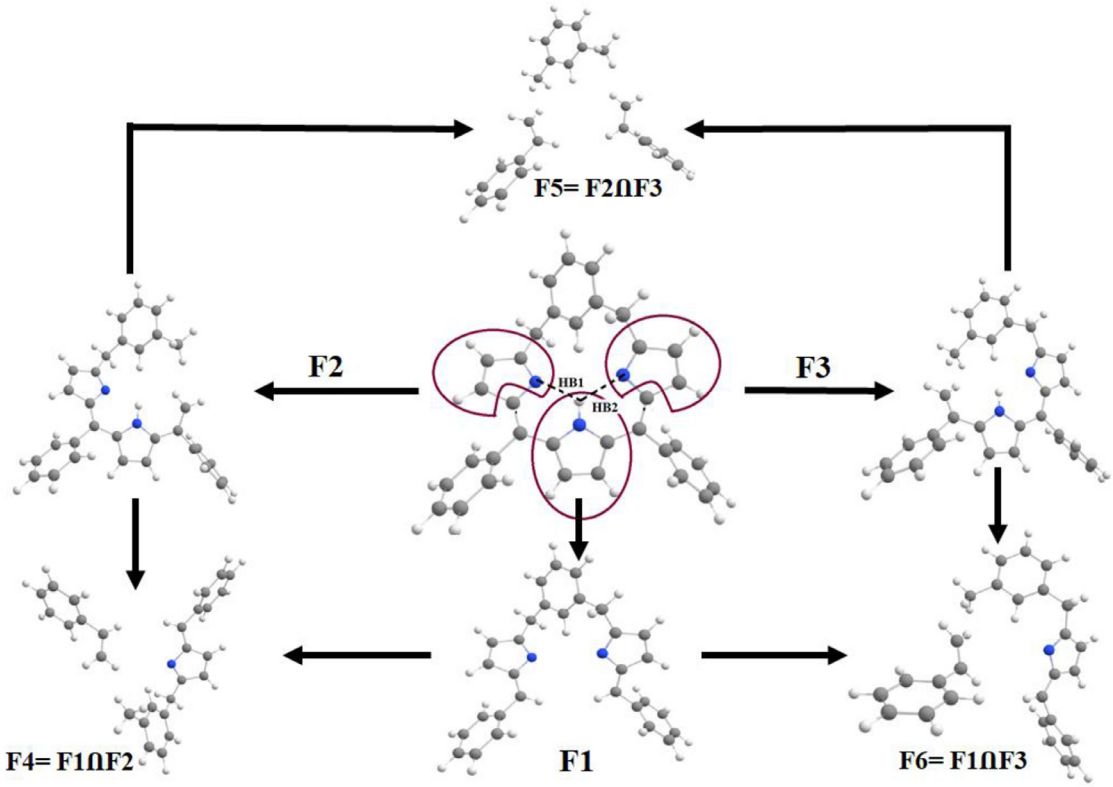


Fig. 18. MDA-MB-468 cells incubated with different *meta*-benziporphodimethenes and Zn^{2+} ions for 30 min and respective bright field marked with A-D as well as fluorescence images (just down to its respective bright field image) were captured at 10X objective lens magnification [93].

that emits at wavelength shorter than 600 nm. Those sensors which are effective above 600 nm, where background emission is very little, minimizes light induced tissue damages, penetrates better and scatters less in optically diffuse samples. It is interesting to note that these kind of moieties do not form complexes with alkali, alkaline earth and other some transition metals which are physiologically important as shown by Hung and coworkers. They also proved that the stability constant for zinc complexation with meta-benziporphodimethene is greater than the cadmium and mercury respectively and hence, Zn^{2+} have a tendency to replace $Cd^{2+}$and Hg^{2+} ions from their corresponding complexes [78].

3.3.3. Fluorescence phenomena: structure co-relation

It is to be noted that *meta*-benziporphodimethene is a selective Zn^{2+} ion sensor and the intensity of fluorescence is comparatively higher than in case of cadmium or mercury. The most important reason for this phenomena is planarity of the macrocycle in the bind state. The free-base macrocycle exists as highly distorted non-planar molecule, but on metallation it has been observed that the tripyrrin moiety becomes planar. The average deviation of the three nitrogen atoms of the tripyrrin plane get reduced from 0.27 Å to 0.10 Å upon metallation. The metal ions in the complex occupy a tetrahedral coordination which involves three nitrogen of tripyrrin moiety and a chlorine which sits at the axial position. It has also been found that the zinc has the least apical deviation of 0.48–0.50 Å from the tripyrrine plane as compared to that of 0.62–0.67 Å for cadmium and 0.70–0.71 Å for mercury, and the orientation of *meta*-phenylene ring in metal-*meta*-benziporphodimethene changed from 81.4° in Zn-*meta*-benziporphodimethene to 89.3° for Cd and 92.1° for Hg. The data is highly in line with the efficient transfer of electron density from ligand to metal and plots a better resonance tendency in case of zinc as binding metal ion [73].

3.3.4. Cell-imaging applications

Zinc ions in the human body are involved in various biological processes. Zinc in the human body can be found in bound state with protein as structural cofactors [86] and gene expression regulator [87]. Labile pool of zinc ions may also be developed in the vesicles of neurons. Besides, Zn^{2+} is also a chief regulator of the cellular apoptosis [88], and it is pertinent to severe diseases like Alzheimer's disease [89,90] and Parkinson's disease [91,92] when a disorder of zinc metabolism has occurred. Patients suffering from inflammatory diseases and tumors also show abnormal Zn^{2+} concentration in sera and tissues. [65] As both vital roles and toxic influences of Zn^{2+} are commonly observed, detection of Zn^{2+} may help to reveal its physiological and pathological roles. Thus, a small-sized semiconductor optoelectronic device is needed for a real-time and indicator-free identification of Zn^{2+} in liquid medium.

With such commercial application in mind, Hung and coworkers have designed an organic hydrogel film with micron-sized pillar array for real-time and indicator-free recognition of zinc ions in the solution by introducing a fluorescent indicator 11,16-bis(phenyl)-6,6,21,21-tetramethyl-*m*-benzi-6,21-porphodimethene in a hydrogel host poly(2-hydroxyethyl methacrylate). High stability and selectivity to Zn^{2+} ions is shown by the sensing film. The responsiveness of this film was increased by fabricating a micron-sized pillar array on the surface of the sensing film to increase the surface area. For Zn^{2+} concentrations of 10^{-4} M and 10^{-3} M, the response time was estimated around 30 s and 3 s, respectively [83].

Recently, Kumar and his coworkers explored these *meta*-benziporphodimethenes for cellular imaging of zinc ions in MDA-MB-468 breast carcinoma cells and concluded that these molecules are lesser cytototoxic for longer time and can internalize successfully (Fig. 19) The electron releasing and withdrawing group

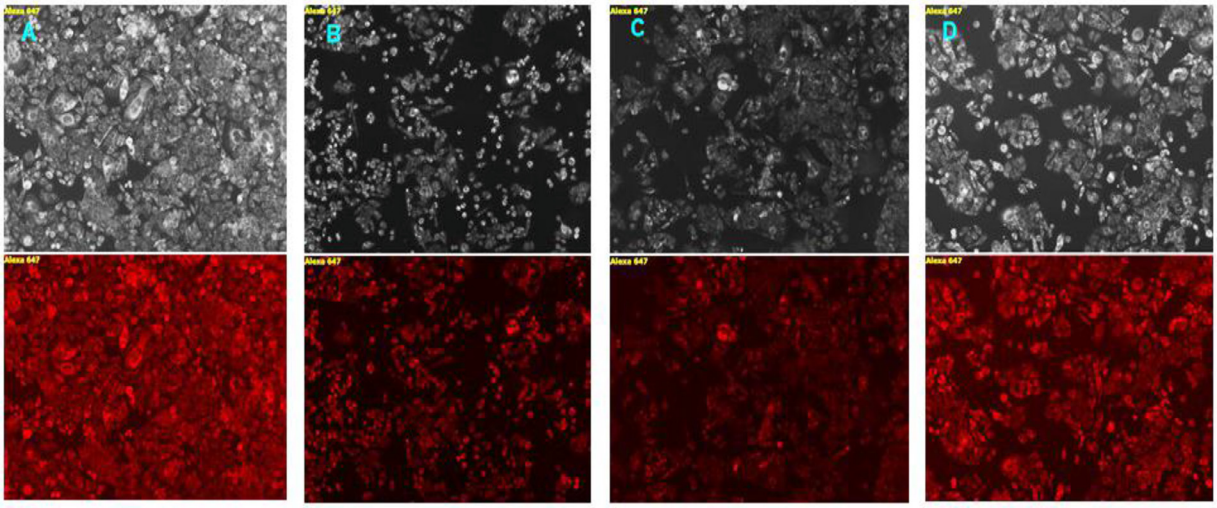


Fig. 19. The structures of dinuclear (2-Ag_2 TMBPDM, (a)) and tetranuclear (2-Ag_4 TMBPDM, (b)) complexes, with 11,16- bisphenyl groups (**2**), omitted for clarity purpose [73].

does not impact on internalization. Further, these molecules can be used for sensing intracellular zinc ions. Though, *meta*-benziporphodimethenes respond to Cd^{2+} and Hg^{2+} ions but the concentration of these ions are rarely present in healthy cells so will not cause any interference in imaging of Zn^{2+} ions in healthy cells [93]. DFT studies were also carried out and suggested the symmetrical geometry of zinc complex of 11,16-bis(2,6 di-fluoro phenyl)-6,6,21,21-tetramethyl-*meta*-benziporphodimethene and the values are comparable to the earlier reports [67].

4. Summary and outlook

Carbaporphyrinoids are one of the promising synthetic porphyrin analogs including corroles, benziporphyrins and *meta*-benziporphodimethenes. Recently, developed *meta*-benziporphodimethenes analogs shows great potential in estimation and detection of spectroscopically silent Zn^{2+}ions both in *in-vitro* and *in-vivo* analysis. These molecules has the advantage of low background emission, secondly the fluorescence occurs at high wavelength in visible region so less photo damage. The effect of meso–phenyl substituents in meta-benziporphodimethenes was found significant in zinc metal complexation and hence in detection. These different meso–sustituted meta-benziporphodimethenes shows good compatibility in estimation of zinc ions selectively among other physiologically available metal ions with good binding coefficient inside the cells. Electronic structure properties of these molecules forwards the advancement in this area of research. Analysis of accurate zinc ion estimation inside the cells can be explored further. It is also note worthy that only aromatic aldehydes have been used for successful synthesis of meta-benziporphodimethene systems. No research group till date have reported the synthesis via condensation of a typical aliphatic aldehyde ($C_nH_{2n+1}CHO$) with pyrrole and a diol (using any methodology). This strongly connects to the idea of inherent electron-richness of the macrocycle, which can be supported by the probable idea that the oxidation potential of the electron donating aliphatic substituents is too low to permit the product to survive (Figs. 11, 16, 18, 19).

Declaration of Competing Interest

Authors declare no competing interest.

Acknowledgements

DA, AK and SGW thank Delhi Technological University for infrastructural facilities. All authors thank Dr. Milind M. Deshmukh for his Molecular Tailoring Approach methodology in our earlier work. DA also thank UGC for providing SRF.

References

[1] G. Jori, C. Fabris, M. Soncin, S. Ferro, O. Coppellotti, D. Dei, L. Fantetti, G. Chiti, G. Roncucci, Photodynamic therapy in the treatment of microbial infections: basic principles and perspective applications, Lasers Surg. Med. 38 (2006) 468–481, doi:10.1002/lsm.20361.
[2] M. Yedukondalu, M. Ravikanth, Core-modified porphyrin based assemblies, Coord. Chem. Rev. 255 (2011) 547–573, doi:10.1016/j.ccr.2010.10.039.
[3] David Dolphin, The Porphyrins, 6, Academic Press, New York, 1978, doi:10.1016/b978-0-12-220103-5.50014-3.
[4] R.G. Karl, M. Kadish, K.M. Smith, The Porphyrin Handbook, 6 A, Academic Press, 2000.
[5] T.D. Lash, Benziporphyrins, a unique platform for exploring the aromatic characteristics of porphyrinoid systems, Org. Biomol. Chem. 13 (2015) 7846–7878, doi:10.1039/c5ob00892a.
[6] T.D. Lash, Out of the Blue! Azuliporphyrins and related carbaporphyrinoid systems, Acc. Chem. Res. 49 (2016) 471–482, doi:10.1021/acs.accounts.5b00523.
[7] T.D. Lash, Carbaporphyrinoid systems, Chem. Rev. 117 (2017) 2313–2446, doi:10.1021/acs.chemrev.6b00326.
[8] B. Szyszko, L. Latos-Grażyński, Expanded Carbaporphyrinoids, Angewandte Chemie - Int. Ed. 59 (2020) 16874–16901, doi:10.1002/anie.201914840.
[9] P.J. Chmielewski, L. Latos-graiynski, M.M. Olmstead, A.L. Balch, Nickel complexes of 21-oxaporphyrin and 21, 23-dioxaporphyrin, Chem. Eur. J. 3 (1997) 268–278, doi:10.1002/chem.19970030216.
[10] P.J. Chmielewski, Synthesis and characterization of a directly linked N-confused porphyrin dimer, Angew. Chem. 116 (2004) 5773–5776, doi:10.1002/ange.200461361.
[11] M. Siczek, P.J. Chmielewski, Synthesis, characterization, and chirality of dimeric N-confused porphyrin–zinc complexes: toward the enantioselective synthesis of bis(porphyrinoid) systems, Angew. Chem. 119 (2007) 7576–7580, doi:10.1002/ange.200701860.
[12] P.J. Chmielewski, Synthesis and characterization of a cyclic bis-silver(I) assembly of four 2-aza-21-carbaporphyrinatosilver(iii) subunits, Angew. Chem. 117 (2005) 6575–6578, doi:10.1002/ange.200502208.
[13] T. Ishizuka, A. Osuka, H. Furuta, Inverted N-confused porphyrin dimer, Angewandte Chemie - Int. Ed. 43 (2004) 5077–5081, doi:10.1002/anie.200460017.
[14] T. Ishizuka, S. Ikeda, M. Toganoh, I. Yoshida, Y. Ishikawa, A. Osuka, H. Furuta, Substitution, dimerization, metalation, and ring-opening reactions of N-fused porphyrins, Tetrahedron 64 (2008) 4037–4050, doi:10.1016/j.tet.2008.02.041.
[15] A. Ulman, J. Manassen, F. Frolow, D. Rabinovich, Synthesis of new tetraphenylporphyrin molecules containing heteroatoms other than nitrogen. III. Tetraphenyl-21-tellura-23-thiaporphyrin: an internally-bridged porphyrin, J. Am. Chem. Soc. 97 (1975) 6540–6544, doi:10.1016/S0040-4039(01)94699-4.
[16] I. Gupta, M. Ravikanth, Recent developments in heteroporphyrins and their analogues, Coord. Chem. Rev. 250 (2006) 468–518, doi:10.1016/j.ccr.2005.10.010.

[17] J. Latos-Grażyński, Lisowski Lechosław, 21-Thiatetra-p -tolylporphyrin and its copper(i1) bicarbonate complex. structural effects of copper-thiophehe binding, J. Am. Chem. Soc. 109 (1987) 4428–4429.

[18] R.D. Ulman, J. Abraham, Manassen, F. Frolow, Synthesis of new tetraphenylporphyrin molecules containing heteroatoms other than nitrogen. III. Tetraphenyl-L1-Tellura-23-Thiaporphyrin: an internally-bridged porphyrin, Tetrahedron Lett. 19 (1978) 1885–1886, doi:10.1016/S0040-4039(01)94699-4.

[19] L. Latos-Grażyński, E. Pacholska, P.J. Chmielewski, M.M. Olmstead, A.L. Balch, Alteration of the reactivity of a tellurophene within a core-modified porphyrin environment: synthesis and oxidation of 21-telluraporphyrin, Angewandte Chemie Int. Ed. Eng. 34 (1995) 2252–2254, doi:10.1002/anie.199522521.

[20] R.D. Ulman, J. Abraham, Manassen, F. Frolow, Synthesis of new tetraphenylporphyrin molecules containing heteroatoms othetr than nitrogen: II. Tetraphenyl-21-selena-23-thiaporphyrin and tetraphenyl-21,23-diselenaporphyrin, Tetrahedron Lett. 19 (1978) 167–170, doi:10.1016/S0040-4039(01)85074-7.

[21] L. Latos-Grazyński, E. Pacholska, P.J. Chmielewski, M.M. Olmstead, A.L. Balch, 5, 20-Diphenyl-10, 15-bis(p-tolyl)-21-selenaporphyrin and Its Nickel(II) Complexes, Inorg Chem 35 (1996) 566–573, doi:10.1021/ic950329z.

[22] Y. Matano, T. Nakabuchi, T. Miyajima, H. Imahori, Phosphole-containing hybrid calixpyrroles: new multifunctional macrocyclic ligands for platinum(II) ions, Organometallics 25 (2006) 3105–3107, doi:10.1021/om060300a.

[23] Y. Matano, T. Nakabuchi, T. Miyajima, H. Imahori, H. Nakano, Synthesis of a phosphorus-containing hybrid porphyrin, Org. Lett. 8 (2006) 5713–5716, doi:10.1021/ol0622763.

[24] Y. Matano, M. Nakashima, T. Nakabuchi, H. Imahori, S. Fujishige, H. Nakano, Monophosphaporphyrins: oxidative π-extension at the peripherally fused carbocycle of the phosphaporphyrin ring, Org. Lett. 10 (2008) 553–556, doi:10.1021/ol7029118.

[25] Y. Matano, H. Imahori, Phosphole-containing calixpyrroles, calixphyrins, and porphyrins: synthesis and coordination chemistry, Acc. Chem. Res. 42 (2009) 1193–1204, doi:10.1021/ar900075e.

[26] Z. Gross, I. Saltsman, R.P. Pandian, C.M. Barzilay, The first metal chelation by a neutral porphyrin analogue, Tetrahedron Lett. 38 (1997) 2383–2386, doi:10.1016/S0040-4039(97)00357-2.

[27] A. Kumar, I. Goldberg, M. Botoshansky, Y. Buchman, Z. Gross, Oxygen atom transfer reactions from isolated (Oxo)manganese(V) corroles to sulfides, J. Am. Chem. Soc. 132 (2010) 15233–15245, doi:10.1021/ja1050296.

[28] Z. Gross, N. Galili, I. Saltsman, The first direct synthesis of corroies from pyrrole, Angewandte Chemie – Int. Ed. 38 (1999) 1427–1429. doi:10.1002/(SICI)1521-3773(19990517)38:10<1427::AID-ANIE1427>3.0.CO;2-1.

[29] N. Galili, Z. Gross, N-substituted corroles : a novel class of chiral ligands, Angew. Chem. Int. Ed. Engl. 38 (1999) 2366–2369.

[30] Z. Gross, L. Simkhovich, N. Galili, T. Israel, First catalysis by corrole metal complexes : epoxidation, hydroxylation, and cyclopropanation the first ever application 4of corroles shows that their metal complexes are good catalysts, almost as potent as the corresponding metalloporphyrins in the oxy, Tetrahedron (1999) 599–600.

[31] L. Simkhovich, A. Mahammed, I. Goldberg, Z. Gross, Synthesis and characterization of germanium, tin, phosphorus, iron, and rhodium complexes of tris(pentafluorophenyl)corrole, and the utilization of the iron and rhodium corroles as cyclopropanation catalysts, Chem. Eur. J. 7 (2001) 1041–1055. doi:10.1002/1521-3765(20010302)7:5<1041::AID-CHEM1041>3.0.CO;2-8.

[32] L. Simkhovich, Z. Gross, Iron(IV) corroles are potent catalysts for aziridination of olefins by Chloramine-T, Tetrahedron Lett. 42 (2001) 8089–8092, doi:10.1016/S0040-4039(01)01717-8.

[33] A. Mahammed, Z. Gross, Iron and manganese corroles are potent catalysts for the decomposition of peroxynitrite, Angew. Chem. 118 (2006) 6694–6697, doi:10.1002/ange.200601399.

[34] I. Aviv, Z. Gross, Corrole-based applications, Chem. Commun. (2007) 1987–1999, doi:10.1039/b618482k.

[35] A.H. Iris, Z. Gross, Aura of corroles, Chem. Eur. J. 15 (2009) 8382–8394, doi:10.1002/chem.200900920.

[36] J. Bendix, H.B. Gray, G. Golubkov, Z. Gross, High-field (high-frequency) EPR spectroscopy and structural characterization of a novel manganese(III) corrole, Chem. Commun. (2000) 1957–1958, doi:10.1039/b006299p.

[37] J.H. Choy, S.Y. Kwak, Y.J. Jeong, J.S. Park, Epoxidation catalysis by a manganese corrole and isolation of an oxomanganese(V) corrole, Angewandte Chemie - Int. Ed. 39 (2000) 4045–4047. doi:10.1002/1521-3773(20001117)39:22<4045::AID-ANIE4045>3.0.CO;2-P.

[38] G. Golubkov, J. Bendix, H.B. Gray, A. Mahammed, I. Goldberg, A.J. DiBilio, Z. Gross, High-valent manganese corroles and the first perhalogenated metallocorrole catalyst, Angewandte Chemie - Int. Ed. 40 (2001) 2132–2134. doi:10.1002/1521-3773(20010601)40:11<2132::AID-ANIE2132>3.0.CO;2-5.

[39] S.K. Pushpan, S. Venkatraman, V.G. Anand, J. Sankar, H. Rath, T.K. Chandrashekar, Inverted porphyrins and expanded porphyrins: an overview, Proc. Indian Acad. Sci. 114 (2002) 311–338, doi:10.1007/BF02703823.

[40] J.L. Sessler, D. Seidel, Synthetic expanded porphyrin chemistry, Angewandte Chemie - Int. Ed. 42 (2003) 5134–5175, doi:10.1002/anie.200200561.

[41] V.J. Bauer, D.L.J. Clive, M.M. King, D. Dolphin, F.L. Harris, J. Loder, S.W.C. Wang, J.B. Paine, R.B. Woodward, Sapphyrins: novel aromatic pentapyrrolic macrocycles, J. Am. Chem. Soc. 105 (1983) 6429–6436, doi:10.1021/ja00359a012.

[42] M.J. Broadhurst, R. Grigg, The synthesis of 22 =-electron macrocycles. Sapphyrins and related compounds, J. Chem. Soc. Perkin Trans. (1972) 2111–2116 1. 1, doi:10.1039/P19720002111.

[43] J.L. Sessler, J.M. Davis, V. Lynch, Synthesis and characterization of a stable smaragdyrin isomer, J. Org. Chem. 63 (1998) 7062–7065, doi:10.1021/jo981019b.

[44] Hans Rexhausen, Albert Gossauer, The synthesis of a new 22 pi-electron macrocycle: pentaphyrin, J. Chem. Soc. Chem. Commun. (1983) 275 –275, doi:10.1039/C39830000275.

[45] A. Gossauer, Syntheses of some unusual polypyrrole macrocycles, Bull. SOC. Chim. Belg. 92 (1983) 793–795, doi:10.1002/bscb.19830920905.

[46] A. Gossauer, N. Engel, Chlorophyll catabolism- structures, mechanisms, conversions, J. Photochem. Photobiol. B 32 (1996) 141–151, doi:10.1016/1011-1344(95)07257-8.

[47] J.L. Sessler, S.J. Weghorn, Y. Hiseada, V. Lynch, Hexaalkyl terpyrrole: a new building block for the preparation of expanded porphyrins, Chemistry 1 (1995) 56–67, doi:10.1002/chem.19950010110.

[48] E. Vogel, M. Broring, J. Fink, D. Rosen, H. Schmickler, J. Lex, K.W.K. Chan, Y. Wu, A. Dietmar, M. Nendel, K.N. Houk, From porphyrin isomers to octapyrrolic "figure eight" macrocycles, Angewandte Chemie (Int. Ed. Eng.). 34 (1995) 2511–2514, doi:10.1002/anie.199525111.

[49] M. Broring, J. Jendrny, L. Zander, H. Schmickler, J. Lex, Y. Wu, M. Nendel, J. Chen, D.A. Plattner, W.R. Roth, Octaphyrin-(1.0.1.0.1.0.1.0), Angew. Chem. 107 (1995) 2709–2711, doi:10.1002/anie.199525151.

[50] A. Werner, M. Michels, L. Zander, J. Lex, E. Vogel, Figure eight" cyclooctapyrroles: enantiomeric separation and determination of the absolute configuration of a binuclear metal complex, Angewandte Chemie - Int. Ed. 38 (1999) 3650–3653. doi:10.1002/(SICI)1521-3773(19991216)38:24<3650::AID-ANIE3650>3.0.CO;2-F.

[51] J.L. Sessler, S.J. Weghorn, T. Morishima, M. Rosingana, V. Lynch, V. Lee, Rosarin: a new, easily prepared hexapyrrolic expanded porphyrin, J. Am. Chem. Soc. 114 (1992) 8306–8307, doi:10.1021/ja00047a061.

[52] J.L. Sessler, T. Morishima, V. Lynch, Rubyrin: a new hexapyrrolic expanded porphyrin, Angewandte Chemie Int. Ed. Eng. 30 (1991) 977–980, doi:10.1002/anie.199109771.

[53] J.A.I. Douglas C. Miller, M.R. Johnson, J.J. Becker, Synthesis and characterization of the new 22-pi-aromatic furan containing macrocycle, "ozaphyrin, J. Heterocycl. Chem. 30 (1993) 1485–1490, doi:10.1002/jhet.5570300604.

[54] D.C. Miller, M.R. Johnson, J.A. Ibers, Synthesis and characterization of new "expanded" thiophene- and furan-containing macrocycles, J. Org. Chem. 59 (1994) 2877–2879, doi:10.1021/jo00089a035.

[55] M.R. Johnson, D.C. Miller, K. Bush, J.J. Becker, J.A. Ibers, Synthesis and characterization of a new 26π-aromatic thiophene-containing macrocyclic ligand, J. Org. Chem. 57 (1992) 4414–4417, doi:10.1021/jo00042a019.

[56] J.L. Sessler, D. Seidel, V. Lynch, Synthesis of [28]heptaphyrin(1.0.0.1.0.0.0) and [32]octaphyrin(1.0.0.0.1.0.0.0) via a directed oxidative ring closure: the first expanded porphyrins containing a quaterpyrrole subunit [18], J. Am. Chem. Soc. 121 (1999) 11257–11258, doi:10.1021/ja992934x.

[57] J.L. Sessler, S.J. Weghorn, V. Lynch, M.R. Johnson, Turcasarin, the largest expanded porphyrin to date, Angewandte Chemie Int. Ed. Eng. 33 (1994) 1509–1512, doi:10.1002/anie.199415091.

[58] J. Setsune, Y. Katakami, N. Iizuna, [48]Dodecaphyrin-(1.0.1.0.1.0.1.0.1.0.1.0) and [64]Hexadecaphyrin-(1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0): the largest cyclopolypyrroles, J. Am. Chem. Soc.. 121 (1999) 8957–8958. 10.1021/ja991669c.

[59] P.J. Chmielewski, L. Latos-grazynski, K. Rachlewicz, T. Glowiak, Tetra-p-tolylporphyrin with an inverted pyrrole ring: a novel isomer of porphyrin, Angew. Chem. Int. Ed. Engl. 33 (1994) 779–781, doi:10.1002/anie.199407791.

[60] M.C.S. Arnoff, The porphyrin-like products of the reaction of pyrrole with benzaldehyde, J. Org. Chem. 8 (1943) 205–223, doi:10.1021/jo01191a002.

[61] C.F. Koelsch, H.J. Richter, The synthesis of bis-2,2′-(1,3-diphenylindenol-3), J. Am. Chem. Soc. 57 (1935) 2010, doi:10.1021/ja01313a509.

[62] P. Rothemund, A new porphyrin synthesis. The synthesis of porphin, J. Am. Chem. Soc. 157 (1936) 625–627.

[63] A.D. Adler, F.R. Longo, W. Shergalis, Mechanistic investigations of porphyrin syntheses. I. Preliminary studies on ms-tetraphenylporphin, J. Am. Chem. Soc. 86 (1964) 3145–3149, doi:10.1021/ja01069a035.

[64] J.B. Kim, J.J. Leonard, F.R. Longo, A mechanistic study of the synthesis and spectral properties of mero-tetraarylporphyrins, J. Am. Chem. Soc. 94 (1972) 3986–3992, doi:10.1021/ja00766a056.

[65] A.D. Adleb, F.R. Longo, J.D. Finarelli, J. Goldmacher, J. Assour, L. Korsakoff, A simplified synthesis for meso-tetraphenylporphine, J. Org. Chem. 32 (1967) 476, doi:10.1021/jo01288a053.

[66] J.S. Lindsey, I.C. Schreiman, H.C. Hsu, P.C. Kearney, A.M. Marguerettaz, Rothemund and adler-longo reactions revisited: synthesis of tetraphenylporphyrins under equilibrium conditions, J. Org. Chem. 52 (1987) 827–836, doi:10.1021/jo00381a022.

[67] K. Berlin, E. Breitmaier, Benziporphyrin, a benzene-containing, nonaromatic porphyrin analogue, Angewandte Chemie Int. Ed. Eng. 33 (1994) 1246–1247, doi:10.1002/anie.199412461.

[68] H. Furuta, T. Asano, T. Ogawa, N-confused porphyrin": a new isomer of tetraphenylporphyrin, J. Am. Chem. Soc. 116 (1994) 767–768, doi:10.1021/ja00081a047.

[69] M. Stępień, L. Latos-Grazyński, L. Szterenberg, J. Panek, Z. Latajka, Cadmium(II) and nickel(ii) complexes of benziporphyrins. a study of weak intramolecular metal-arene interactions, J. Am. Chem. Soc. 126 (2004) 4566–4580, doi:10.1021/ja039384u.

[70] G.P. Moss, Nomenclature of tetrapyrroles, 1979. 10.1351/pac199668101919.

[71] G.P. MOSS, Nomenclature of tetrapyrroles: recommendations 1986, Eur. J. Biochem. 178 (1988) 277–328, doi:10.1111/j.1432-1033.1988.tb14453.x.

[72] M. Stępień, L. Latos-Grazyński, Benziporphyrins: exploring arene chemistry in a macrocyclic environment, Acc. Chem. Res. 38 (2005) 88–98 10.1021/ar040189+.
[73] G.F. Chang, C.H. Wang, H.C. Lu, L.S. Kan, I. Chao, W.H. Chen, A. Kumar, L. Lo, M.A.C. Dela Rosa, C.H. Hung, Factors that regulate the conformation of m-benziporphodimethene complexes: agostic metal-arene interaction, hydrogen bonding, and Î• 2,π coordination, Chem. Eur. J. 17 (2011) 11332–11343, doi:10.1002/chem.201100780.
[74] M. Stępień, L. Latos-Grazyński, Tetraphenylbenziporphyrin - a ligand for organometallic chemistry, Chem. Eur. J. 7 (2001) 5113–5117. doi:10.1002/1521-3765(20011203)7:23<5113::AID-CHEM5113>3.0.CO;2-V.
[75] M. Stępień, L. Latos-Grazyński, Tetraphenyl-p-benziporphyrin: a carbaporphyrinoid with two linked carbon atoms in the coordination core, J. Am. Chem. Soc. 124 (2002) 3838–3839, doi:10.1021/ja017852z.
[76] C.H. Hung, C.Y. Lin, P.Y. Lin, Y.J. Chen, Synthesis and crystal structure of core-modified benziporphyrin: thia-p-benziporphyrin, Tetrahedron Lett. 45 (2004) 129–132, doi:10.1016/j.tetlet.2003.10.090.
[77] T.D. Lash, Oxybenziporphyrin, a fully aromatic semiquinone porphyrin analog with pathways for 18π-electron delocalization, Angewandte Chemie Int. Ed. Eng. 34 (1995) 2533–2535, doi:10.1002/anie.199525331.
[78] C.H. Hung, G.F. Chang, A. Kumar, G.F. Lin, L.Y. Luo, W.M. Ching, E. Wei-Guang Diau, m-Benziporphodimethene: a new porphyrin analogue fluorescence zinc(II) sensor, Chem. Commun. (2008) 978–980, doi:10.1039/b714412a.
[79] G.F. Chang, A. Kumar, W.M. Ching, H.W. Chu, C.H. Hung, Tetramethyl-m-benziporphodimethene and isomeric α,β-unsaturated γlactam embedded N-confused tetramethyl-m-benziporphodimethenes, Chem. Asian J. 4 (2009) 164–173, doi:10.1002/asia.200800279.
[80] R.K. Sharma, L.K. Gajanan, M.S. Mehata, F. Hussain, A. Kumar, Synthesis, characterization and fluorescence turn-on behavior of new porphyrin analogue: meta-benziporphodimethenes, Spectrochimica Acta Part 169 (2016) 58–65, doi:10.1016/j.saa.2016.06.020.
[81] A. Kumar, C.H. Hung, S. Rana, M.M. Deshmukh, Study on the structure, stability and tautomerisms of meta-benziporphodimethene and N-confused isomers containing γ–lactam ring, J. Mol. Struct. 1187 (2019) 138–150, doi:10.1016/j.molstruc.2019.03.064.
[82] D. Ahluwalia, A. Kumar, S.G. Warkar, M.M. Deshmukh, Effect of substitutions on the geometry and intramolecular hydrogen bond strength in meta-benziporphodimethenes: a new porphyrin analogue, J. Mol. Struct. 1220 (2020) 128773–128783, doi:10.1016/j.molstruc.2020.128773.
[83] Y.C. Chao, Y.S. Huang, H.P. Wang, S.M. Fu, C.H. Huang, Y.C. Liang, W.C. Yang, Y.S. Huang, G.F. Chang, H.W. Zan, H.F. Meng, C.H. Hung, C.C. Fu, An organic hydrogel film with micron-sized pillar array for real-time and indicator-free detection of Zn2+, Org. Electron. 12 (2011) 1899–1902, doi:10.1016/j.orgel.2011.07.001.
[84] P. Chavez-Crooker, N. Garrido, G.A. Ahearn, Copper transport by lobster hepatopancreatic epithelial cells separated by centrifugal elutriation: measurements with the fluorescent dye Phen Green, J. Exp. Biol. 204 (2001) 1433–1444.
[85] W. Breuer, S. Epsztejn, P. Millgram, I.Z. Cabantchik, Transport of iron and other transition metals into cells as revealed by a fluorescent probe, Am. J. Physiol. (1995) 268, doi:10.1152/ajpcell.1995.268.6.c1354.
[86] E.H. Cox, G.L. McLendon, Zinc-dependent protein folding, Curr. Opin. Chem. Biol. 4 (2000) 162–165, doi:10.1016/S1367-5931(99)00070-8.
[87] G.K. Andrews, Cellular zinc sensors: MTF-1 regulation of gene expression, Biometals 14 (2001) 223–237, doi:10.1023/A:1012932712483.
[88] A.Q. Truong-Tran, J. Carter, R.E. Ruffin, P.D. Zalewski, The role of zinc in caspase activation and apoptotic cell death, Biometals 14 (2001) 315–330, doi:10.1023/A:1012993017026.
[89] M.P. Cuajungco, G.J. Lees, Zinc and Alzheimer's disease: is there a direct link? Brain. Res. Rev. 23 (1997) 219–236, doi:10.1016/S0165-0173(97)00002-7.
[90] C. Devirgiliis, P.D. Zalewski, G. Perozzi, C. Murgia, Zinc fluxes and zinc transporter genes in chronic diseases, Mutat. Res. 622 (2007) 84–93, doi:10.1016/j.mrfmmm.2007.01.013.
[91] D.T. Dexter, A. Carayon, F. Javoy-agid, Y. Agid, F.R. Wells, S.E. Daniel, A.J. Lees, P. Jenner, C.D. Marsden, Alterations in the levels of iron, ferritin and other trace metals in parkinson's disease and other neurodegenerative diseases affecting the basal ganglia, Brain 114 (1991) 1953–1975, doi:10.1093/brain/114.4.1953.
[92] G.A. Qureshi, A.A. Qureshi, S.A. Memon, S.H. Parvez, Impact of selenium, iron, copper and zinc in on/off Parkinson's patients on L-dopa therapy, J. Neural Transm. Suppl. (2006) 229–236, doi:10.1007/978-3-211-33328-0_24.
[93] R.K. Sharma, A. Maurya, P. Rajamani, M.S. Mehata, A. Kumar, meta-Benziporphodimethenes: new cell-imaging porphyrin analogue molecules, ChemistrySelect 1 (2016) 3502–3509, doi:10.1002/slct.201600812.

Rehoboam at the NTCIR-15 SHINRA2020-ML Task

Tushar Abhishek
IIIT-Hyderabad, India
tushar.abhishek@research.iiit.ac.in

Ayush Agarwal
Delhi Technological University, India
ayush286@gmail.com

Anubhav Sharma
IIIT-Hyderabad, India
anubhav.sharma@research.iiit.ac.in

Vasudeva Varma
IIIT-Hyderabad, India
vv@iiit.ac.in

Manish Gupta*
IIIT-Hyderabad, India
manish.gupta@iiit.ac.in

ABSTRACT

Maintaining a unified ontology across various languages is expected to result in effective and consistent organization of Wikipedia entities. Such organization of the Wikipedia knowledge base (KB) will in turn improve the effectiveness of various KB oriented multilingual downstream tasks like entity linking, question answering, fact checking, etc. As a first step toward a unified ontology, it is important to classify Wikipedia entities into consistent fine-grained categories across 30 languages. While there is existing work on fine-grained entity categorization for rich-resource languages, there is hardly any such work for consistent classification across multiple low-resource languages. Wikipedia webpage format variations, content imbalance per page, imbalance with respect to categories across languages make the problem challenging. We model this problem as a document classification task. We propose a novel architecture, RNN_GNN_XLM-R, which leverages the strengths of various popular deep learning architectures. Across ten participant teams at the NTCIR-15 Shinra 2020-ML Classification Task, our proposed model stands second in the overall evaluation.

Keywords: Document Classification, Named-Entity Recognition, XLM Roberta, Multi-lingual Modeling, RNN_GNN_XLM-R

TEAM NAME

Rehoboam (RH312)

TASK

Shinra 2020-ML Classification Task

1 INTRODUCTION

Wikipedia consists of a large number of entity-centric articles. Entity mining on such comprehensive resource has been useful in many natural language processing (NLP) tasks. To maximize the use of such knowledge, resources created from Wikipedia need to be structured for inference, reasoning, or any other purposes in many NLP applications. The current structured knowledge bases such as DBpedia, Wikidata, Freebase, YAGO, and Wikidata among others, are created mostly by bottom-up crowdsourcing, which may cause a significant amount of undesirable inconsistency in the structure of the knowledge base.

To resolve the issues and with a final goal to structure the knowledge in Wikipedia and create the KB as accurately as possible, Shinra ML Task was defined as to classify 30 language Wikipedia entities into 219 categories defined in Extended Named Entity (ENE) (ver.8.0)[1] (a four-layer ontology for names, time, numbers and concepts), using categorized Japanese Wikipedia pages and the inter-language links to the corresponding pages in target languages.

Language	Entity Name	English Translation	Fine-grained category
Japanese	東京都	Tokyo	Province
Hindi	प्रशांति निलयम	Prasanthi Nilayam	Worship_Place
Russian	Звонящий в полночь	Caller at midnight	Broadcast_Program
Spanish	Sala de crisis de la Casa Blanca	White House Crisis Room	Facility_Part
French	Sneaker Pimps	Sneaker Pimps	Show_Organization

Table 1: Entity Categorization Examples

A few examples of such entities with their fine-grained categorization are shown in Table 1.

Recently, several Transformer [15]-based multi-lingual models across a wide range of NLP tasks have been proposed. This includes models for multi-lingual translation, summarization, question generation, sentiment analysis, etc. In this work, we focus on multi-lingual text categorization. Models like mBERT [7], XLM [10], XLM-R [6] and Unicoder [8], InfoXLM [4] are therefore relevant to our problem. However, to the best of our knowledge, this is the first work on multi-lingual categorization of Wikipedia entities using Transformer-based models.

We initially experimented with mono-lingual entity categorization for Hindi using various models like MPAD (Message Passing Attention Networks for Document Understanding) [12], MAGNET (Multi-Label Text Classification using Attention-based Graph Neural Network) [13], DTMT (Deep Transition RNN-based Architecture for Neural Machine Translation) [11] and a named entity recognition (NER) based model. Next, we investigated the performance of multi-lingual models like mBERT and XLM-R on entity categorization for 29 languages[1]. Finally, we propose a novel multi-lingual approach, RNN_GNN_XLM-R, which helps us get the best reported results.

Across 30 languages, our proposed model achieves a micro-F1 average score of 73.7. Across ten participant teams at the NTCIR-15 Shinra 2020-ML Classification Task [14], our proposed model stands second in the overall evaluation.

*The author is also an applied scientist at Microsoft

[1] All 30 languages except for Greek because of input dataset format differences

We discuss details of the dataset in Section 2. We discuss our baseline methods and our best proposed method, RNN_GNN_XLM-R, in Section 3. We present detailed evaluation results in Section 4. Finally, we conclude with a brief summary in Section 5.

2 DATASET

2.1 Original Dataset

The original dataset comprises on 5.07 million Wikipedia entities pages categorized into one of the 219 categories across 30 languages. For more details, please refer the overview of SHINRA2020-ML Task [14].

2.2 Preprocessing

We use the Wikipedia Cirrus Dump provided for each language for training, validation and testing. We use the 'text' section of the dump for experimenting with our models as an input. We perform masking of numerical values, removing white-spaces and punctuation characters. We also remove characters of other languages from the 'text' field, and the hyperlinks.

2.3 Taxonomy

The dataset has been annotated in a hierarchy of four levels. Each data sample can have multiple labels. The label can belong to any of the four levels. The topmost level consists of five categories Name, Timex, Numex, Concept, and Ignored. Name, Timex and Numex are further sub-categorized to three more levels while the last two topmost categories are leaf nodes in the type hierarchy.

2.4 Multi-lingual dataset selection

We carefully selected a set of 3.8M entity pages covering a subset of 18 languages out of 30 to compose multi-lingual dataset for faster training. To maintain language diversity, we used syntactic word orders as the primary criteria of selection and the number of training data available in each language as secondary criteria. Within the syntactic word ordering, only subject-verb-object (SVO) and subject-object-verb (SOV) order were preferred as they cover most of the world's spoken languages. Obtained multi-lingual dataset was randomly shuffled without replacement to maintain a proper mix of different languages in the training batch. We used a stratified split ratio of 80:20 to create training and validation datasets. Languages selected for creation of multi-lingual datasets are listed in Table 2

SVO order		SOV order	
Languages	**Training Data**	**Languages**	**Training Data**
Chinese	267,107	Dutch	199,983
English	439,354	German	274,732
French	318,828	Hindi	30,547
Finnish	144,750	Hungarian	120,295
Italian	270,295	Korean	190,807
Polish	225,552	Persian	169,053
Portuguese	217,896	Turkish	111,592
Russian	253,012		
Spanish	257,835		
Swedish	180,948		
Vietnamese	116,280		

Table 2: Languages selected for creation of multi-lingual dataset along with training data statistics.

3 FINE-GRAINED ENTITY-TYPE CLASSIFICATION METHODS

We tried the following state-of-the-art methods as our baselines. Since most (~98%) of the entity pages (especially in Hindi) have single labels, we experiment with Message Passing Attention Networks for Document Understanding (MPAD) [12] which is a state-of-the-art approach for single label document classification. Next, we experiment with multi-label text classification methods like Attention-based Graph Neural Network (MAGNET) [13]. Further, since the task is very entity-centric, we designed an entity-specific CNN-RNN model. All of the above approaches used Gated Recurrent Units (GRUs) [5], so we moved to the DTMT encoder approach which further acts as an enhancement over our RNN models. Finally, we propose a new architecture, RNN_GNN_XLM-R, for the task which is a combination of RNN, GNN and Transformer modules. In the following, we provide a brief description of these models.

3.1 MPAD

Message Passing Attention Networks for Document Understanding (MPAD) [12] represent documents as word co-occurrence networks. They used GNNs (graph neural networks) with an enhanced COMBINE step to create intermediate representation for each node in the graph. These node representations then pooled in the READOUT step to create document representation. MPAD performs best on 5 out 10 standard text classification datasets and has competitive results with the state-of-the-art on others.

3.2 DTMT

Meng et al. [11] enhance the RNN based Neural Machine Translation (NMT) by increasing the transition depth between consecutive hidden states and build a novel Deep Transition RNN-based Architecture for Neural Machine Translation. It reinforces the hidden-to-hidden transition with multiple non-linear transformations and additionally captures linear transformation path throughout this deep transition to avoid the gradient vanishing problem. We used the bi-directional encoder from this architecture and passed the final hidden state of the final encoder token to a linear layer to classify sentences into multi-label classes.

3.3 MAGNET

Multi-Label Text Classification using Attention-based Graph Neural Network (MAGNET) [13] uses feature matrix and correlation matrix obtained from GNNs in order to capture dependencies between the labels and classifiers for the downstream task. Each classification label is represented as nodes in the graph and dependencies among them are learned through GNNs implicitly. It achieves state-of-the-art performance over 5 different multi-label classification datasets.

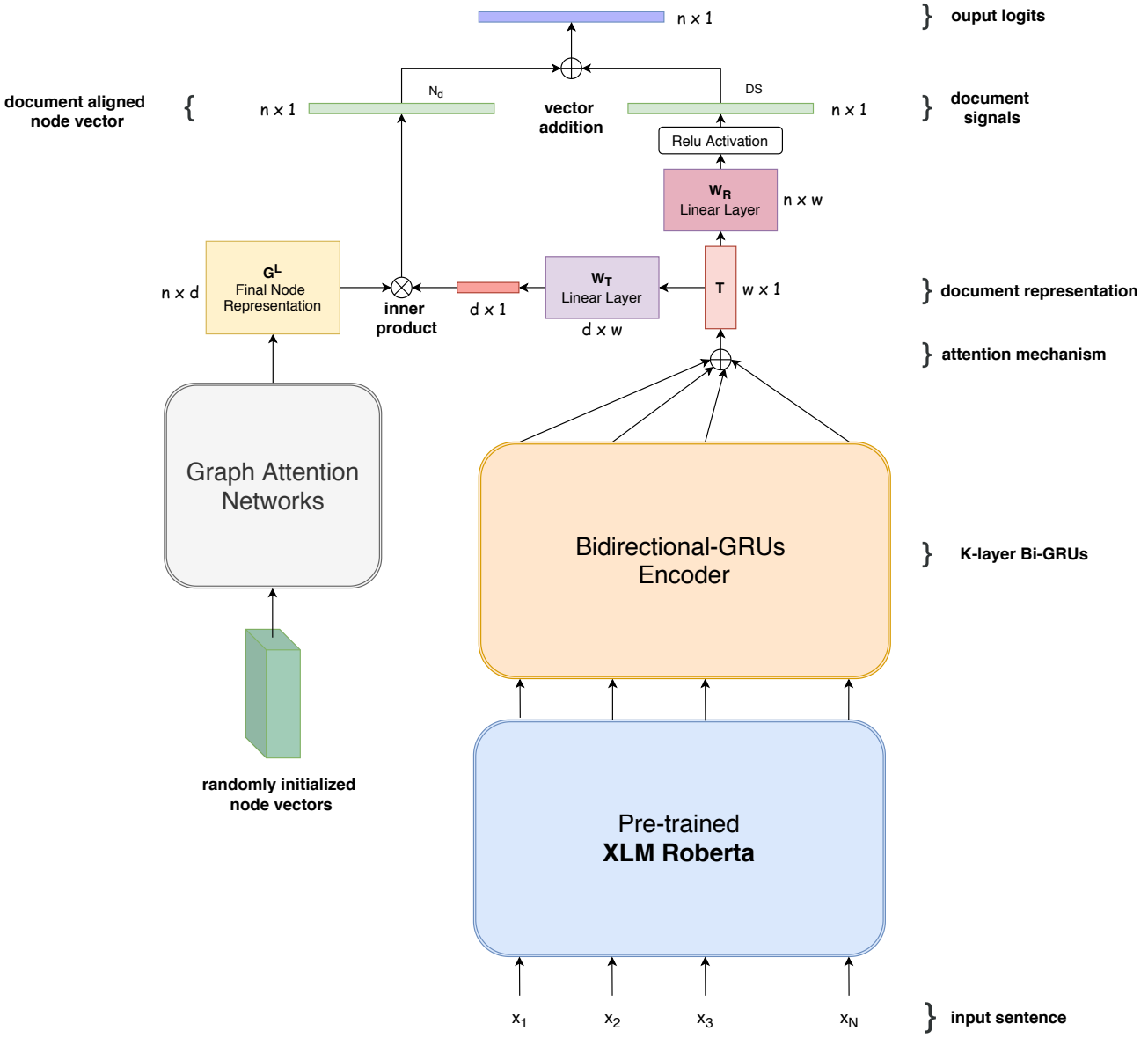


Figure 1: Architecture of our proposed system, RNN_GNN_XLM-R

3.4 NER Based Model

During the task, we also experimented with a model based on named entities present in the text of a Wikipedia Page. We extract the named entities and the entity type of each one of them from the 'text' field of the Wikipedia dump. We then conducted an experimentation study with a CNN-RNN architecture to combine the text, named entity, and entity type for each Wikipedia article. We trained an embedding layer for the named entities and another for the named entity types. We fine-tune the 300D fastText [9] word embeddings for our input text. For every input text token, we thus have a 900D vector (300D fastText, 300D entity embedding and 300D entity type embedding). This 900D vector was transformed using 3 CNN layers to obtain a 100D representation for each token. These were then passed as input to an LSTM whose last layer was connected to an output softmax layer for final prediction.

3.5 Multi-lingual Models

We also experimented with popular pre-trained multi-lingual Transformer architectures. We fine-tuned mBERT-base [7] (12 layers) and XLM-R-base [6] (12 layers). The vector representation of the [CLS] token from last layer is attached to a linear classification head and

the model is trained over the multi-lingual dataset mentioned in Section 2.4.

3.6 Proposed Method: RNN_GNN_XLM-R

Each Wikipedia article is represented as sequences of words of length m, $D' = [b_1, b_2, \ldots, b_m]$ where b_i represented the i^{th} word. Input sub-word token is obtained after passing document D' to sentence-piece tokenizer, to get $D = [x_1, x_2, \ldots, x_n]$ where x_i represents the i^{th} subword. The document D is passed to XLM-Roberta and the output from its last layer is used as input feature representation of each sub-word vector $U = [u_1, u_2, \ldots, u_n]$. Input representation U is then fed to a K-layer attentional bidirectional-GRU to get enhanced contextual representation, and the output vectors from both directions are concatenated. Similar to [13], our model learns the relatedness among the labels through Graph Attention Networks (GAT) [16]. To preserve document level signals for classification we pass the document vector T through a non-linear transformation with ReLU activation. Final score for each label is obtained by adding it with node representation, and then connecting to an output softmax layer. Figure 1 shows the overall architecture of the proposed model.

4 EXPERIMENTS AND RESULTS

4.1 Evaluations

4.1.1 Metric. The model is expected to classify each page into one or more of the 219 taxonomy categories correctly. If the estimated category is not an exact match, the model does not get a score for that. The model is evaluated for the multi-label classification using the micro-averaged F1 measure, i.e., the harmonic mean of micro-averaged precision and micro-averaged recall.

4.1.2 Leaderboard. We test the performance of our baseline models as described in Sections 3.1-3.5 in addition to the proposed model,RNN_GNN_XLM-R, on the leaderboard [2] of the Shinra2020 ML Task. The leaderboard dataset consists of 2000 data samples whose distribution was independent of the training dataset. The organizers provided the performance of the model using the micro-averaged F1 metric. Results from the leaderboard are described in Tables 3 and 4.

4.1.3 Final evaluation by Shinra organisers. For the final evaluations, we were required to submit the model predictions on the entire test set for any number of languages. The detailed statistics of the entire test set are unknown to the participants. The model performance was evaluated by the organizers using micro-averaged F1 metric. We present the results obtained on the test set for our proposed approach in Section 4.3. Results using the entire test set for final evaluation are described in Table 5.

Methods	Training Data	**Micro-Averaged F1-score**
MPAD [12]	Mono-lingual	59.5
DTMT [11]	Mono-lingual	61.9
MAGNET [13]	Mono-lingual	61.3
NER based model (Ours)	Mono-lingual	61.1
mBERT-base [7]	Multi-lingual	66.5
XLM Roberta [6]	Multi-lingual	68.1
RNN_GNN_XLM-R (Ours)	Multi-lingual	**72.1**

Table 3: The F1-score are reflected from Shinra leaderboard submission for Hindi evaluation dataset.

4.2 Hindi dataset results (leaderboard)

On the Hindi dataset we tried all the approaches (both monolingual and multi-lingual). The models mentioned in Sections 3.1-3.4 are trained on the Hindi language dataset only. For implementing the MPAD we extracted all the datapoints with single label and tested it on them. Whereas for MAGNET, NER based Model and DTMT encoder, all the datapoints were considered while testing. Table-3 shows results only for Hindi language obtained from Shinra leaderboard submission. mBERT-base and XLM Roberta have been trained on the multi-lingual training dataset but evaluated on the Hindi leaderboard dataset.

The table clearly validates that RNN involving approaches don't perform that well as compared to Transformer-based approaches. Also we observe that MAGNET outperforms MPAD. The enhanced RNN encoder in DTMT outperforms the traditional RNN encoder models (MPAD and MAGNET). Surprisingly, our entity-aware NER based model does not perform that well. As expected, multi-lingual training outperforms mono-lingual training. Our proposed system, RNN_GNN_XLM-R, is better in comparison to other Transformer-based models. This is because our model not just uses text semantics but also leverages correlations between class labels for improved predictions.

4.3 Multi-lingual results (leaderboard)

We populated the results from the Shinra leaderboard [2] for our proposed system, RNN_GNN_XLM-R, and multi-lingual baselines along with some of the other leading participant models. We trained RNN_GNN_XLM-R on 18 languages and did zero shot inference on the remaining 11 languages – Arabic, Bulgarian, Catalan, Czech, Danish, Hebrew, Indonesian, Norwegian, Romanian, Ukrainian and Thai. Table 4 is the official leaderboard result dated 30^{th} September 2020.

4.4 Official evaluations results

We have populated the official results for six languages. We submitted the predictions to the Shinra[3] evaluation system only for those six languages. Table-5 contains the best micro-averaged F1-score of our model as well as some of the other competitive models in the respective languages.

5 CONCLUSION

In this paper, we studied the problem of multi-lingual fine-grained entity type classification. We investigated the effectiveness of various state-of-the-art monolingual RNN encoders as well as multi-lingual Transformer-based models. We proposed a novel adaptation of the XLM-R model where we augment the XLM-R base with RNN and GNN modules. We experimented with these models on the datasets provided by the NTCIR-15 Shinra2020 ML task organizers. We evaluated the effectiveness of our models on Hindi

Languages	Models					
	HUKB	**PribL**	**uomfj**	**mBERT-base**	**XLM-R-base**	**RNN_GNN_XLM-R**
English (en)	51.2	73.9	74.5	72.1	74.4	**76.0**
Spanish (es)	56.7	75.1	73.2	70.7	73.1	**75.4**
French (fr)	48.9	73.5	70.4	73.3	73.1	**74.3**
German (de)	61.1	**75.8**	71.3	72.5	70.8	75.3
Chinese (zh)	60.1	75.4	72.5	70.5	73.0	**75.8**
Russian (ru)	55.0	**74.5**	69.8	70.9	70.3	73.4
Portugese (pt)	51.9	71.0	69.5	69.5	68.9	**72.4**
Italian (it)	49.0	73.4	72.4	70.0	71.9	**74.3**
Arabic (ar)	51.0	70.7	70.2	67.1	69.6	**72.3**
Indonesian (id)	54.0	-	71.2	72.9	74.4	**74.5**
Turkish (tr)	58.2	73.2	69.5	72.6	**73.9**	**73.9**
Dutch (nl)	60.3	**73.8**	70.8	72.3	72.9	72.9
Polish (pl)	61.6	**76.6**	73.5	71.8	74.2	75.1
Persian (fa)	60.9	-	73.8	70.0	72.9	**73.9**
Swedish (sv)	59.6	-	69.7	70.9	69.9	**73.4**
Vietnamese (vi)	61.7	72.2	72.1	74.4	**75.4**	74.7
Korean (ko)	53.8	**74.6**	72.3	71.2	71.2	70.8
Hebrew (he)	52.2	-	68.6	68.9	68.0	**72.1**
Romanian (ro)	53.2	-	69.8	72.3	71.7	**75.0**
Norwegian (no)	50.2	71.7	70.8	68.5	70.5	**72.2**
Czech (cs)	53.3	69.2	68.1	68.0	69.5	**72.2**
Ukrainian (uk)	58.2	69.6	70.5	70.5	70.0	**71.5**
Hindi (hi)	44.1	60.5	65.9	66.5	68.1	**72.1**
Finnish (fi)	51.9	-	73.1	**73.2**	73.0	72.9
Hungarian (hu)	54.9	-	71.4	71.7	72.8	**74.6**
Danish (da)	52.7	72.2	73.2	70.4	71.4	**73.8**
Thai (th)	60.3	-	50.3	66.8	70.9	**75.2**
Catalan (ca)	43.9	-	71.6	**72.1**	70.9	71.9
Bulgarian (bg)	56.7	-	74.7	75.2	73.9	**76.7**
Average	54.7	72.5	70.5	70.9	71.7	**73.7**

Table 4: Micro-averaged F1 score comparison across various models on the Leaderboard test set. HUKB, PribL and uomfj are the other top performing teams on the Shinra Challenge leaderboard. mBERT-base and XLM-R-base are our other baselines. RNN_GNN_XLM-R is our best proposed model. Best results for each language are highlighted in bold.

Languages	Teams				
	ousia	**uomfj**	**LIAT**	**PribL**	**RH312 (ours)**
Bulgarian	-	**83.07**	75.2	-	82.13
French	**81.01**	78.21	76.88	78.52	80.31
Hindi	69.75	66.67	16.49	-	**71.7**
Indonesian	-	**78.51**	72.44	-	77.55
Thai	76.36	65.02	49.58	-	**76.77**
Turkish	-	**84.85**	77.19	84.36	83.28

Table 5: Micro-averaged F1 score comparison across various models on the official evaluation dataset for 6 languages. ousia, uomfj, LIAT and PribL are the other top performing teams on the Shinra Challenge leaderboard. RH312 (i.e., RNN_GNN_XLM-R) is our best proposed model. Best results for each language are highlighted in bold.

dataset as well as on datasets for six languages. Our proposed model, RNN_GNN_XLM-R, outperforms other methods we experimented with. We believe this is because XLM Roberta extracts rich semantics from the document, attentional-RNN module enhances the contextual information and the GNN module helps learn classification label correlations leading to improved generalization. The leaderboard and final evaluation results demonstrate the effectiveness of our models.

REFERENCES

[1] Extended Named Entity. *Homepage.* https://ene-project.info/

[2] Shinra leaderboard. *Homepage.* https://www.nlp.ecei.tohoku.ac.jp/projects/AIP-LB/task/shinra2020-ml

[3] Shinra project. *Homepage.* http://shinra-project.info/?lang=en

[4] Zewen Chi, Li Dong, Furu Wei, Nan Yang, Saksham Singhal, Wenhui Wang, Xia Song, Xian-Ling Mao, Heyan Huang, and Ming Zhou. 2020. InfoXLM: An Information-Theoretic Framework for Cross-Lingual Language Model Pre-Training. *arXiv preprint arXiv:2007.07834* (2020).

[5] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase

representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* (2014).

[6] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised Cross-lingual Representation Learning at Scale. In *ACL*. Association for Computational Linguistics, 8440–8451.

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-HLT (1)*. Association for Computational Linguistics, 4171–4186.

[8] Haoyang Huang, Yaobo Liang, Nan Duan, Ming Gong, Linjun Shou, Daxin Jiang, and Ming Zhou. 2019. Unicoder: A universal language encoder by pre-training with multiple cross-lingual tasks. *arXiv preprint arXiv:1909.00964* (2019).

[9] Armand Joulin, Edouard Grave, Piotr Bojanowski, Matthijs Douze, Hérve Jégou, and Tomas Mikolov. 2016. Fasttext. zip: Compressing text classification models. *arXiv preprint arXiv:1612.03651* (2016).

[10] Guillaume Lample and Alexis Conneau. 2019. Cross-lingual language model pretraining. *arXiv preprint arXiv:1901.07291* (2019).

[11] Fandong Meng and Jinchao Zhang. 2019. DTMT: A novel deep transition architecture for neural machine translation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 224–231.

[12] Giannis Nikolentzos, Antoine J.-P. Tixier, and Michalis Vazirgiannis. 2020. Message Passing Attention Networks for Document Understanding. In *AAAI*. AAAI Press, 8544–8551.

[13] Ankit Pal, Muru Selvakumar, and Malaikannan Sankarasubbu. 2020. MAGNET: Multi-Label Text Classification using Attention-based Graph Neural Network. In *ICAART (2)*. SCITEPRESS, 494–505.

[14] Satoshi Sekine, Masako Nomoto, Kouta Nakayama, Asuka Sumida, Koji Matsuda, and Maya Ando. 2020. Overview of SHINRA2020-ML Task. In *Proceedings of the NTCIR-15 Conference*.

[15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*. 5998–6008.

[16] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2017. Graph Attention Networks. *CoRR* abs/1710.10903 (2017).

Journal Pre-proofs



Research paper

Remote Mid IR Photoacoustic Spectroscopy for the Detection of Explosive Materials

Mohit Mann, A.S. Rao, Ramesh C. Sharma

PII: S0009-2614(20)31131-3
DOI: https://doi.org/10.1016/j.cplett.2020.138231
Reference: CPLETT 138231

To appear in: *Chemical Physics Letters*

Received Date: 15 July 2020
Revised Date: 10 November 2020
Accepted Date: 26 November 2020

Please cite this article as: M. Mann, A.S. Rao, R.C. Sharma, Remote Mid IR Photoacoustic Spectroscopy for the Detection of Explosive Materials, *Chemical Physics Letters* (2020), doi: https://doi.org/10.1016/j.cplett.2020.138231

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.

© 2020 Published by Elsevier B.V.

Journal Pre-proofs

Remote Mid IR Photoacoustic Spectroscopy for the Detection of Explosive Materials

Mohit Mann[1, 2], A.S. Rao[2], Ramesh C Sharma[1]*

Laser Science and Technology Centre, DRDO, Delhi - 110054, India[1]

Department of Applied Physics, Delhi Technological University, Bawana Road,

Delhi - 110042, India[2]

Abstract:

Standoff photo acoustic detection of explosives using quantum cascade laser has been studied. The laser beam has been focused on to the target molecules adsorbed on metallic surfaces. A 25 kHz microphone, a resonator cavity in form of a cylinder, and a parabolic acoustic reflector have been used to demonstrate the method. The experimental results have been compared with the theoretical calculations in three different experimental conditions. The photo acoustic signals have been found to decrease with increase in the standoff range. The sensitivity of 2.0 μg/cm^2 has been achieved at ~ 4.0 meter.

Keywords: *Tunable Quantum Cascade Laser, Ultrasonic Detector, Resonator Cavity, Photoacoustic Spectroscopy, Parabolic Reflector.*

***Corresponding Author: Rameshsharma@lastec.drdo.in**

1. Introduction:

In the recent past, the terrorists activities around the globe have been increased many folds making the human life vulnerable [1]. Therefore, the early warning against the use of hazardous chemicals, nerve gases and explosives is extremely desirable for timely action to mitigate the threat and safeguard human life. Standoff Photoacoustic (PA) technique has emerged as a successful method in detection of hazardous chemicals and explosives. Several scientists have described the sensing of high energy molecules, hazardous chemicals and explosive materials using Mid – IR vibrational spectroscopy [2-9]. In recent years, the Quantum Cascade Laser (QCL) has grabbed the attention of researchers due to better quality output power, mode quality, wide range tunability and room temperature operations as compared to other diode lasers [10-12]. The efforts towards the development of QCL based detection systems using PA spectroscopic technique has progressed significantly [13-18]. Every explosive material has its

Journal Pre-proofs

own characteristic signature in terms of optical, acoustic, physical and thermal properties [19]. Laser Photoacoustic Spectroscopy (LPAS) offers superior capability in identifying molecular signatures from standoff distances in comparison to FTIR spectroscopy [20]. It offers a unique technique to differentiate between explosives and similar non-explosive materials using phase-shift analysis of spectrum [21].

The present work has been carried out to study the detection and identification of hazardous chemicals and explosives adsorbed on surfaces of the suspicious objects by screening them. In the study, a QCL, emitting electromagnetic radiation in mid infrared wavelength region, has been used as the laser source. An ultrasonic microphone coupled with resonator tube of cylindrical shape, has been employed as a standoff sensor. The sensor has peak sensitivity at 25 kHz. An acoustic reflector has been utilized to enhance the acoustic signal collection. The developed system is capable to detect < 2.0 μg/cm^2 of explosives/hazardous materials from standoff distance of ~ 4.0 meters.

2. Experimental:

The schematic of experimental setup is shown in Fig 1. The ultrasonic microphone was been used as a PA detector. A tunable Quantum Cascade Laser (make: Daylight Solutions; model no. UberTuner -8) operating in the wave number range 1130 cm^{-1} to 1420 cm^{-1} (i.e. 7 μm to 9 μm in wavelength) was utilized as the laser source. The step size of the tuning was kept at 1.0 cm^{-1}. A function generator was used to modulate the laser beam of QCL at 25 kHz with pulse duration of 500 ns. The average peak power of the laser was nearly 600 mW. The working principle of the system is as follows. The laser beam is allowed to fall on sample adsorbed on surface. The electromagnetic radiation is absorbed by the sample resulting in excitation of molecules followed by non-radiative decay which in turn generates the acoustic/pressure waves. These waves are collected by the acoustic reflector and focused on to the cylindrical resonator and sensed by the coupled microphone. The microphone used has a narrow frequency response centered at ~25 kHz. The acoustic signals received by microphone are then amplified by pre-amplifier. After amplification the signals are fed to lock-in-amplifier (USB 4716). A reference signal at the same frequency (generated by a function generator) is also fed to lock-in amplifier. The output of the lock-in amplifier is processed by data acquisition system and PA signal is

plotted with the wave number (in the entire tuning range of the QCL) using a Lab View based GUI developed in house.

The samples of TNT and PETN were prepared using pure acetone solvent. The samples dissolved in Acetone were spread on the target plate and allowed to dry for few hours. The Acetone got vaporized and the dried uniform residue layers were left on the target plate. The concentration of sample on the surface of target plate was ~5 µg/cm^2. Experimental spectra were recorded in the three conditions when: (1) only microphone was used, (2) microphone with cylindrical resonator tube was used and (3) microphone, cylindrical resonator tube and parabolic acoustic reflector were used. The resonator tube placed at the focal point of parabolic reflector (as shown in Fig 1). The aperture diameter of parabolic reflector was 25 cm.

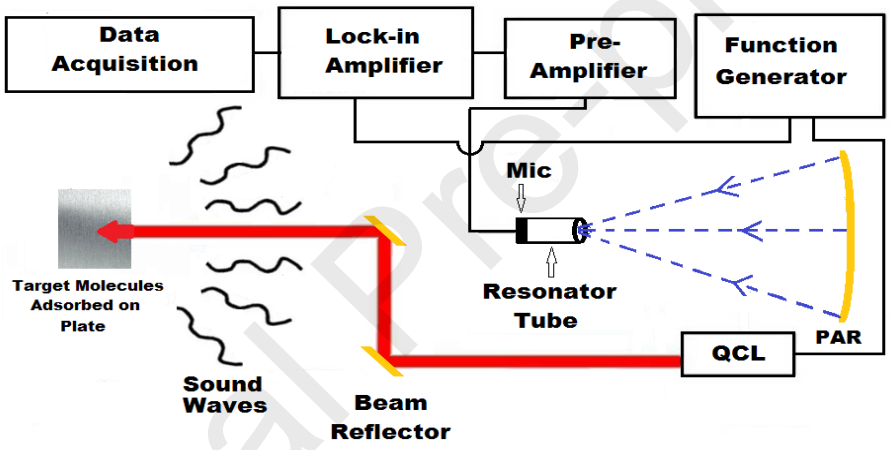


Fig. 1*: Schematic of experimental setup for Standoff laser photo acoustic Sensor for hazardous materials/explosives adsorbed on surfaces of suspicious objects with microphone, cylindrical resonator tube and parabolic acoustical reflector (PAR).*

3. Theoretical Photoacoustic Calculations

Acoustic cylindrical resonator tube used in the experiments was made up of stainless steel material. Simulations were carried out for the pressure waves generated inside the resonator cavity with one side open ended cylindrical geometry containing the ultrasonic Eigen frequencies concentrated at ~25 kHz. The developed resonator tube has a length of 2.40 cm and inner diameter of 1.70 cm [16]. The resonator tube was attached to the ultrasonic piezo microphone detector from one side and other side was left open. The microphone used operates at a resonant

frequency of 25 kHz. The Eigen frequencies of acoustic resonator open cylinder can be obtained theoretically by equation (1),

$$f = n.s/4L \quad (1)$$

where 'n' stands for any odd number, 's' is speed of sound in air and 'L' is the length of cylinder. COMSOL Multi-physics modeling software was used to simulate the Eigen frequencies of the resonator using Finite-element based method. The Eigen modes and Eigen frequency are calculated by solving the following partial differential equation,

$$\nabla^2[p(r)/\rho] + [(\omega/s)^2.(p(r)/\rho] = 0 \quad (2)$$

where $p(r)$ is the acoustic pressure and is given by,

$$p = p_0 \exp(i\omega t)\ N/m^2 \quad (3)$$

is a harmonic measure for the pressure acoustics frequency field of study. 'ρ' is the density of air (1.2 kg/m^3), 'ω' is the angular frequency, and 's' is the speed of sound in air (~340 m/s) at STP. The surface of the cylinder is considered to be hard surface boundaries with no surface at the opening. The pressure inside the cell is taken ~10.13 Pascal (~0.076 Torr). The theoretical calculations are also made based on the acoustic waves received by cylindrical resonator tube coupled with microphone. The absorption of the modulated transmitted mid IR laser radiation by a sample. The general photo-acoustic equation is described as equation (4) by [16-18]

$$p_0 = \Gamma\, \alpha_{abs}\, F \quad (4)$$

where p_0 is the local pressure, Γ is the Gruneisen parameter of target sample, α_{abs} is the absorption coefficient of sample, and F is the local light fluence. The Gruneisen parameter can be further expressed as; $\Gamma=\beta v_s^2/c_p$, where β is the isobaric volume expansion coefficient, v_s is the acoustic speed in medium (340m/s), and c_p is the specific heat. For a short laser pulse equation(4) can be solved to get local pressure rise p_0, which is related to volumetric absorption A_e (J/cm^3), compressibility k, the density of medium as ρ, the specific heat capacity,c_p and $v_s$be the acoustic velocity in the propagation median and by the relation; $v_s^2 = (1/ k.\rho)$

$$p_0 = (\beta/k.\rho.c_p).A_e = \Gamma.A_e \quad (5)$$

Journal Pre-proofs

In the equation, only Γ and α_{abs} are dependent on absorbers in target sample, Γ be the dimensionless unit.

$$A_e = \alpha_{abs}(cm^{-1}) \, . \, F \, (J/cm^2) \quad (6)$$

The pressure wave propagating from the source can be described to have amplitude at a distance 'r' from it, is as given in equation [17, 18];

$$P\,(r) = (p_0 / r).\, exp\,(-8.68.\alpha_{atm}.\, r) \quad (7)$$

where α_{atm} is the attenuation coefficient of the atmosphere (dB/m) and the factor 8.68 is for unit conversion to m^{-1}. The generated PA pressure amplitude, reaching at a distance 'r' from the source is having '1/r' dependence. Attenuation coefficients at 30% humidity with ultrasonic frequencies between 20 kHz to 80 kHz are plotted in Figure 2. Using the equation (7) PA signals with standoff range are plotted in Fig.3 at resonant ultrasonic frequency 25 kHz. LPAS signal is achieved up to 4.0 meter distance at modulation at 25 kHz frequency. Parameters used for PETN and TNT explosive molecules for theoretical calculation is given in Table 1.

Table 1: Different parameters used for PETN and TNT molecules for theoretical calculations. Coefficient of Atmospheric Attenuation (α_{atm}) at 25 kHz = 0.821.

Target Sample	Iso-thermal compressi-bility κ (GPa^{-1})	Thermal coefficient of volume expansionβ (K^{-1})	Density ρ (Kg/m^3)	Heat Capacity c_p (J/Kg.K)	F (mJ/m^2)	μ	Volumetric Absorption A_e (J/m^3)	P_0
PETN	1/8.52	$3.9*10^{-5}$	1654	1071.19	4.5	$1.95*10^6$	$7.8*10^6$	1463
TNT	1/13.51	$6*10^{-5}$	1974	2702.86	4.5	$1.95*10^6$	$7.8*10^6$	1185

4. Results & Discussion

Fig. 2 demonstrates the graph between attenuation coefficient in dB/m v/s frequency in kHz at relative humidity of 30% and the air temperature of 30ºC. From this graph, the value of attenuation coefficient obtained at 25 kHz is 0.821 dB/m.

Journal Pre-proofs

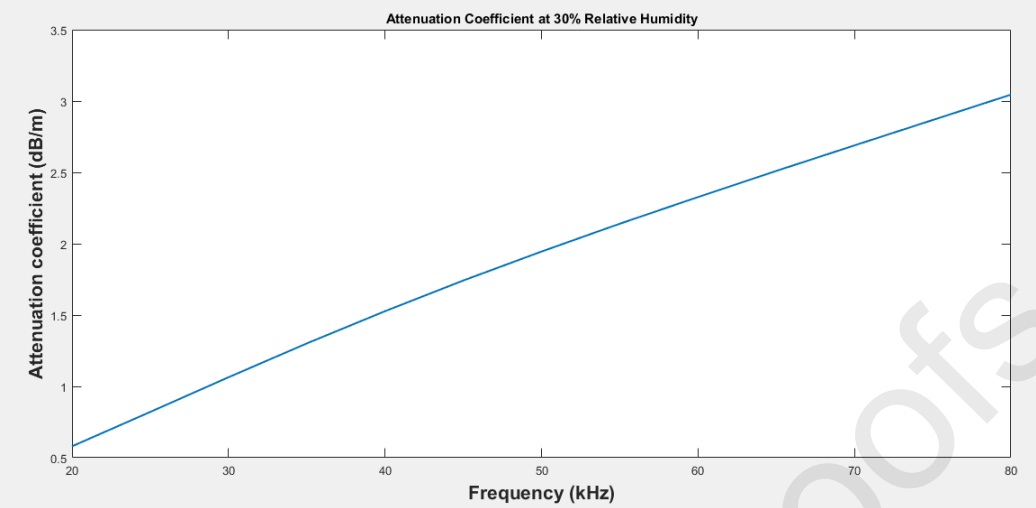


Fig. 2: Ultrasonic Frequency dependent atmosphere attenuation of acoustic wave at Relative Humidity (30%). ISO 9613-1:1993

The absorption cross section of explosive molecules like RDX, TNT, PETN, TATP, are of the order of 10^{-18} cm^2/molecule in the wavelength region 7.0 to 9.0 µm. For the case of PETN molecule, the absorption cross section at wavelength 7.7 µm (wave number 1286 cm^{-1}) is of the order of 5.4 x10^{-18} cm^2/molecule. For PETN very strong absorption peak at 1286 cm^{-1} is due to NO_2 stretch, CH bend and C_5 skeletal. The absorption coefficient can be determined by multiplying the absorption cross section of the molecule with the density of the molecules. The product of absorption coefficient (α) and absorption length (l) gives absorbance (A).

In our experimental work laser peak power was ~ 600 mW with a spot size of 1.0 cm^2 area on a uniform film of samples deposited on the target surface. Pulse width of laser was 500ns. The resultant fluence 'F' was 3.80 mJ/m^2. The literature values of κ-isothermal compressibility and the β-Thermal coefficient of volume expansion of PETN are (1/11.83) GPa^{-1} and 17.4 x 10^{-5} K^{-1} respectively. Other properties such as density ρ, specific heat capacity c are 1620 kg/m^3 and 1250 J/kg. K respectively, resulting into a Grueneisen parameter Γ~1, using equation 5 [9]. The absorption depth is approximately 0.80 µm at 1286 cm^{-1}. Therefore, for film thickness greater than 1.0 µm, the whole of the laser fluence is absorbed by PETN film. From equation 5, for PETN molecules, we inferred the local pressure rise P_0 ~ 6,500 Pascal, above the atmospheric pressure. The pressure wave propagating from the source can be described to have amplitude at a distance 'r' from it, is as given in equation (7) (shown in Fig.3).

Journal Pre-proofs

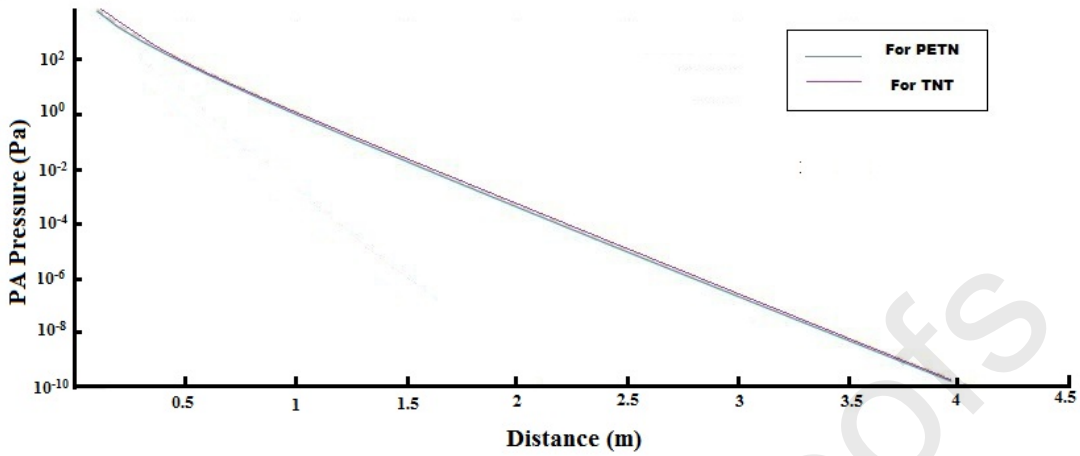


Fig. 3: Calculated Standoff PA Pressure (Pa) v/s Distance (m) at ultrasonic frequency of 25 kHz.

The PA pressure of 10^{-10} Pa could be measured up to a distance of 4.0 meters for both the analysts at 25 kHz frequency. Fig. 4 and Fig. 5 depict the standoff PA spectra of TNT and PETN molecules respectively. The background of the signal is represented by Fig.4 (a) and Fig. 5(a) respectively for TNT and PETN when there was no sample adsorbed on the target surface. When the experiment was performed with microphone, the characteristic peak of TNT molecule was obtained at 1360 cm^{-1} (Fig.4(b)) whereas three characteristics peaks appeared at 1276 cm^{-1}, 1286 cm^{-1}, 1310 cm^{-1}in case of PETN molecule (Fig.5(b)). Figures 4(c) and 5(c) represent the photo-acoustic spectra recorded using microphone coupled with the resonator tube (i.e. without using acoustic reflector). The two fold enhancement in the PA signal was observed in comparison to the spectra recorded with microphone only. To further enhance the signal, acoustic reflector was used in conjunction with resonator tube. Spectra recorded with both resonator tube and acoustic reflector are shown in Fig.4 (d) and Fig. 5 (d) for TNT and PETN respectively. In this case, the PA signal enhanced to a greater extent.

Journal Pre-proofs

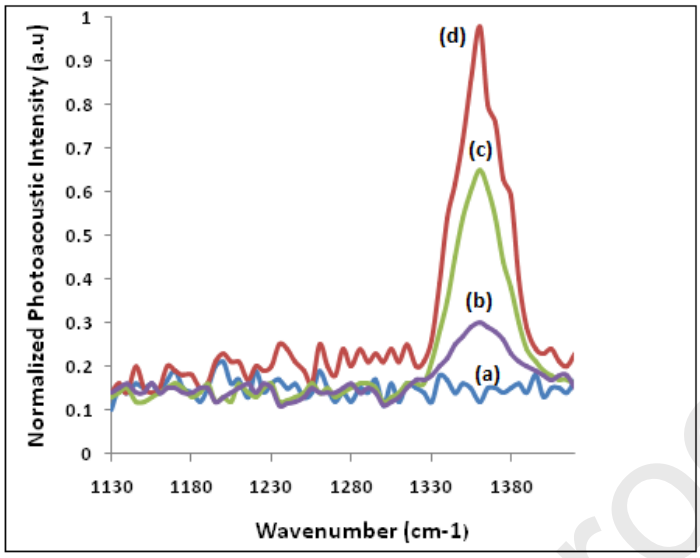


Fig. 4: Standoff Photoacoustic Spectra of TNT (a) Background without Sample (b) Only Microphone(c) At 25 kHz with microphone and resonator tube (d) At 25kHzwith microphone and resonator tube coupled with acoustic reflector. (Quantity of each case 5µg/cm² and Standoff Distance of 1m for each case)

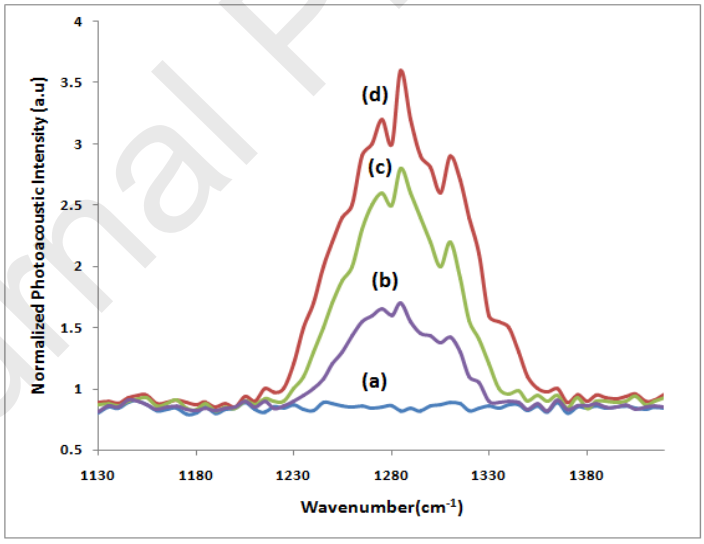


Fig. 5: Standoff Photoacoustic Spectra of PETN (a) Background without Sample (b) Only Microphone (c) At 25kHz with microphone and resonator tube (d) At 25kHzwith microphone and resonator tube coupled acoustic reflector. (Quantity of each case 5µg/cm² and Standoff Distance of 1m for each case)

Journal Pre-proofs

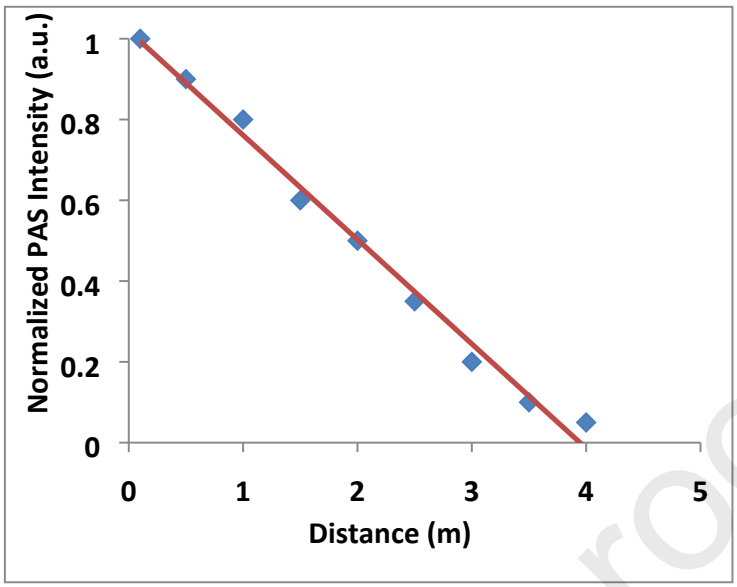


Fig. 6: Photoacoustic Signal of TNT in quantity 5µg/cm² adsorbed on surface at different distances.

Figure 6 shows variation of normalized PAS signal of TNT (concentration of 5µg/cm² on adsorbed on target surface) with the distance. The detection limit obtained experimentally from the graph is up to 4.0 meters at SNR > 1.0 with cylindrical resonator tube coupled microphone detector. Sensitivity in the case of microphone coupled with resonator tube and parabolic reflector was ~ 2.0 µg/cm² for standoff distance of 4.0 m.

4. Conclusions

Standoff PA detection of explosives using tunable quantum cascade laser, an ultrasensitive microphone as the acoustic detector, a resonator tube and acoustic parabolic reflector has been demonstrated. The study indicated that the PA signal strength is proportional to laser pulse energy and inversely proportional to standoff detection distance. With the use of resonator tube the signal strength can be enhanced or measurement range can be extended. A parabolic acoustic reflector has been observed to extend the detection range further. The theoretically obtained photo acoustic signal observed to be in good agreement with the experimental results. The present set up is very capable of detecting/identifying explosives and hazardous materials by screening of suspicious objects. It has the potential to be used as screening device for Defence and homeland security. With the development of higher power QCLs it is possible to achieve a longer detection range and trace amount detection capability.

Journal Pre-proofs

Declarations

Availability of data and material - The data is provided in the form of table in the paper. Sample/Material used is PETN and TNT. No Funding is received from external source.

Acknowledgments

The authors are highly thankful to the Dr. Subodh Kumar for his valuable suggestions during all the stages of research work. Mr.Sudhir Kumar, Mr. Umesh Sah and Mrs. Usha Kumari for their help during the experimental work, Mr Mayank and Mr Mukul Sharma, Delhi Technological University helped for Theoretical Calculation.

References:

1. Victoria A. Greenfield**,** Reducing the Threat of Improvised Explosive Device Attacks by Restricting Access to Explosive Precursor Chemicals, The National Academies Press, Washington, DC, 23-26, **(2018).**

2. Y. A. gruzdkov and Y. M. Gupta, J. Phys. Chem. A, 105, 6197, **(2001).**

3. K. McCann and J. Laane, Journal of Molecular Structure, 890, 36, **(2008).**
4. J. J. P. Stewart, S. R. Bosco, W. R. Carper, SpectrochimicaActa Part A: Molecular Spectroscopy, 42, 13,**(1986).**

5. C. M. Ramos, L. F. Alzate, N. M. Hernaádez, S. P. Hernaádez, N. Min, Journal of Molecular Structure: Theochem, 769, 69,**(2006).**

6. F. Pristera, M. Halik, A. Castelli, W. Fredericks, Anal. Chem., 32, 495,**(1960).**

7. A. Banas, K. Banas, M. Bahua, H. O. Moser, L. Wen, P. Yang, Z. J. Li, M. Cholewa, S. K. Lim, Ch. H. Lim, Vibrational Spectroscopy, 51, 168-176, **(2009).**

Journal Pre-proofs

8. D. L. Perkins, C. R. Lovell, B. V. Bronk, B. Setlow, P. Setlowm and M. L. Myrick, Applied Spectroscopy, 59, 893, **(2005).**

9. S. V. Ingale, P. B. Wagh, P. U. Sastry, A. K. Patra, R. Tiwari, I. K. Singh, S. B. Phapale, R. D. Wasnik, A. S. Rao and S. C. Gupta, Defence Sci. Journal, 61, 534,**(2011).**

10. F.Xie, C. Caneau, H. P. Leblanc, D. P. Caffey, L. C. Hughes, T. Day, C-en. Zah, "Watt-level room temperature continuous-wave operation of Quantum cascade laser with λ>10μm", IEEE Journal of Selected Topics in Quantum Electronics, 19, 1200407, **(2013).**

11. A. Talukdar, M. F. Khan, D. Lee, S. Kim, T. Thundat & G. Koley, "Piezotransistive transduction of femto scale displacement for photo acoustic spectroscopy", Nature Communications, 6,**(2015).**

12. J. S. Li, B. Yu, H. Fischer, W. Chen and A. P. Yalin "Quantum cascade laser based Photoacoustic detection of explosives", Rev. Sci. Instrum. 86, 031501,**(2015).**

13. Ramesh C. Sharma, Surya Gautam, Saurabh Gupta, Hari B. Srivastava "Photoacoustic sensor for trace detection of post-blast explosive and hazardous molecules". Sensors and Actuators B: Chemical; 243, 59-63,**(2017).**

14. Ramesh C. Sharma, Subodh Kumar, Surya Kumar Gautum, Saurabh Gupta, Deepak Kumar, Hari B. Srivastava, "Detection of Ultrasonic Waves Using Resonant Cylinderical Cavity for Defense Application", IEEE Sensor Journal, 17, 1681,**(2017).**

15. D. C. Dumitras, D. C. Dutu, C. Matei, A. M. Magureanu, M. Petrus, C. Popa, Laser Photoacoustic Spectroscopy: principles, instrumentation and characterization, Journal of Optoelectronics and Advanced Materials, 9, 3655-3701, **(2017).**

16. Ramesh C. Sharma, Subodh Kumar, Sudhir Kumar, Mohit Mann, Mayank and Mukul Sharma, "Photoacoustic Remote Sensing of suspicious object for defence and forensic

Journal Pre-proofs

applications", SpectrochimicaActa Part A: Molecular and BiomolecularSpectroscopy, 224, 117445,**(2020).**

17. Chen Xing , Guo D, Choa FS, Wang CC, Trivedi S, Snyder AP, Ru G, Fan J., "Standoff photoacoustic detection of explosives using quantum cascade laser and an ultrasensitivemicrophone", Applied Optics 52 (12), 2626-32, **(2013).**

18. Lihong V. Wang, Tutorial on Photoacoustic Microscopy and Computed Tomography IEEE Journal of Selected topics in Quantum Electronics 14, 1, 171, **(2008).**

19. Yasser H. El-Sharkawy, Sherif Elbasuney, "Novel laser induced photoacoustic spectroscopy for instantaneous trace detection of explosive materials", Forensic Science International 277, 215–222, **(2017).**

20. Sherif Elbasuney, Yasser H. El-Sharkawy, "Instant identification of explosive material: Laser induced photoacoustic spectroscopy versus fourier transform infrared", Trends in Analytical Chemistry 108, 269-277, **(2018).**

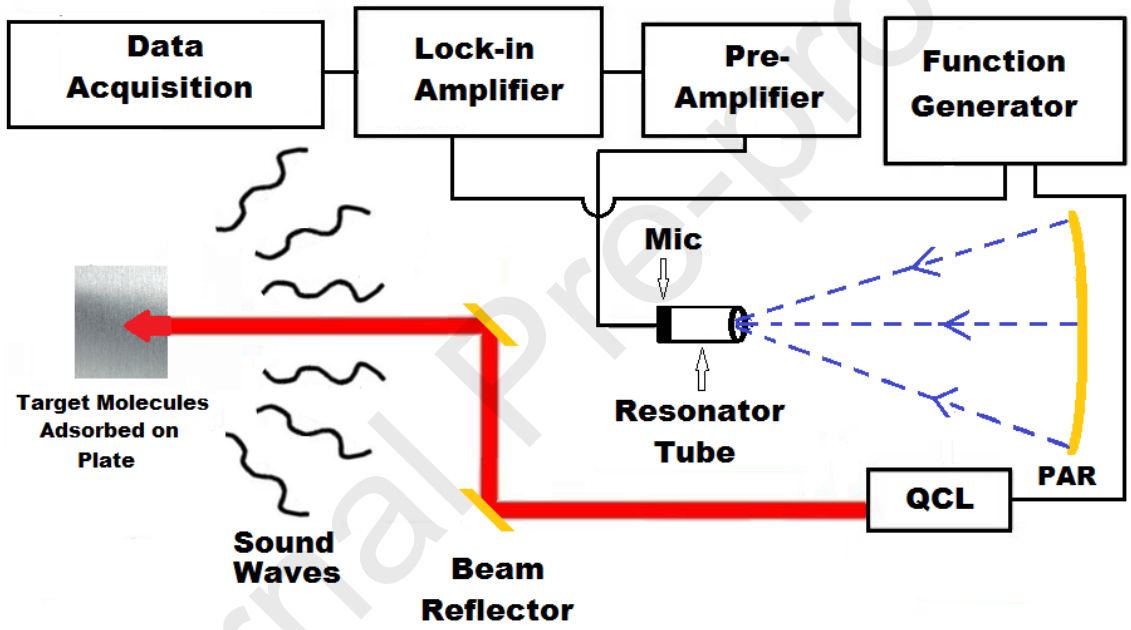
21. Yasser H. El-Sharkawy, Sherif Elbasuney, "Real time recognition of explosophorous group and explosive material using laser induced photoacoustic spectroscopy associated with novel algorithm for time and frequency domain analysis", Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 204, 25-32, **(2018).**

Journal Pre-proofs

Declaration of Interest Statement

We declare that there is no conflict of interest.**No Funding** is received from external source.

Graphical Abstract



Schematic experimental setup of standoff laser PA Sensor with Explosive adsorbed on plate (suspicious objects) using microphone, resonator tube and parabolic acoustical reflector.

Journal Pre-proofs

Journal Pre-proofs

Highlights of the Research Paper

- Detection of hazardous Chemicals/ Explosives using Standoff Laser Photoacoustic Spectroscopy (LPAS).
- Acoustical Signal collecting efficacy increased by using different components such as microphone, resonator tube and a parabolic acoustic reflector.
- Detection limit for sample having concentration of 5 $\mu g/cm^2$ is determined experimentally.
- A lower detection limit of 2.0 $\mu g/cm^2$ is obtained with coupled devices at the standoff distance of ~4.0 meter.

Journal Pre-proofs

Arabian Journal for Science and Engineering
https://doi.org/10.1007/s13369-020-05112-2

RESEARCH ARTICLE-COMPUTER ENGINEERING AND COMPUTER SCIENCE

Research on Understanding the Effect of Deep Learning on User Preferences

Garima Gupta[1] · **Rahul Katarya**[1]

Received: 25 May 2020 / Accepted: 2 November 2020
© King Fahd University of Petroleum & Minerals 2020

Abstract
Recommender systems are becoming more essential than ever as the data available online is increasing manifold. The increasing data presents us with an opportunity to build complex systems that can model the user interactions more accurately and extract sophisticated features to provide recommendations with better accuracy. To construct these complex models, deep learning is emerging as one of the most powerful tools. It can process large amounts of data to learn the structure and patterns that can be exploited. It has been used in recommender systems to solve cold-start problem, better estimate the interaction functions, and extract deep feature representations, among other facets that plague the traditional recommender systems. As big data is becoming more prevalent, there is a need to use tools that can take advantage of such explosive data. An extensive study on recommender systems using deep learning has been performed in the paper. The literature review spans in-depth analysis and comparative study of the research domain. The paper exhibits a vast range of scope for efficient recommender systems in future.

Keywords Recommender systems · Machine learning · Deep learning

1 Introduction

A recommender system's goal is to present items to a user, which are most likely to lead to conversion. Conversion might relate to different things in different contexts. For example, for e-commerce, it might mean purchasing the product, and for Netflix, it might mean viewing the content. To achieve this goal, recommender systems have to study the underlying data, consisting of items, users, and their interactions. To study these interactions, we need to extract relevant features and create a system that can learn and model such interactions.

Based on the properties of the network that the system exploits, recommender systems can be categorized into content-based, collaborative filtering-based, and hybrid systems. Recommender systems based on content exploit the item's features to find similar items that the user might like. A collaborative filtering recommender system uses the user–user interaction. The philosophy being user similar to each other might have similar choices. Systems that combine both these aspects in one or the other way are termed as hybrid systems.

Recommender systems use machine learning capabilities at its core to learn the interaction functions and predict items that are most likely to lead to conversion. As the data present online is increasing every day, and the concept of big data is harnessing attention, the dimensionality and modality of data are growing explosively. Hence, there arises a need for a tool that can exploit big data to provide better results. Deep learning is essentially an extension of machine learning that is specifically designed to manage big data. Deep learning algorithms are based on large neural networks with multiple hidden layers. These models can exploit data to learn complex interactions and feature representations. This can help us improve the accuracy of recommendations and bestow new meaning to tackle problems like cold start that plagues traditional recommender systems.

Because of the many advantages, there is a sudden boom in implementing deep learning in RS. For this reason, ACM has been holding a workshop DLRS (deep learning in recommender systems) with RecSys since the year 2016. Several

✉ Rahul Katarya
rahulkatarya@dtu.ac.in

Garima Gupta
garimacsdtu@gmail.com

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

Published online: 26 November 2020

deep learning techniques are implemented in recommender systems, as discussed in the paper. Some papers have used only a single technique, while others have used an amalgamation of different techniques to improve the results. Deep learning is a powerful tool and is being used to solve a plethora of problems in recommender systems. It has also proved to be effective in improving accuracy in various domains. In this paper, the research work done in RS using deep learning techniques has been explored. With the advancement in big data, practitioners will need better tools to tackle the complexities that come with it, and at the same time, exploit it for better results.

To make useful recommendations, there are various aspects to be kept in mind like the application domain, type of dataset, problem statement, kind of recommender system used, type of deep learning technique used, the side information such as gender of the user, user's login sessions, text semantics, and much more.

This paper will report the following research questions restricted to deep learning in recommender systems, and the answers to these questions have been provided in Sect. 3:

.

RQ 1	Which application domains have been included in the study?
RQ 2	What are the different pre-defined and self-generated datasets other authors have worked on in the past?
RQ 3	What types of recommender systems have been studied and implemented in past studies?
RQ 4	Which deep learning techniques have been implemented in the previous research?
RQ 5	Which metrics have been used to analyze the results?

The paper has been organized as follows. Section 2 furnishes the background study of the literature survey, i.e., a short description of the recommender systems, machine learning, and deep learning. Section 3 is the Research Methodology in which the readers can understand the division of research papers studied for this survey paper. Here, the papers have been classified based on various parameters. Section 4 is the Literature Review, which consists of a detailed study of the research work and encompasses the conclusion and the research gaps observed in the mentioned field of study. The paper concludes with References.

This paper is written to explain the application of deep learning in recommender systems and discuss how recommender systems can be improved and extended by applying different deep learning techniques. To perform this survey, we conducted an extensive study of previous research done in recommender systems, employing deep learning. We surveyed top conferences and journals in deep learning in recommender systems and collated the papers relevant to our research area. The original contribution of the survey is presented in Sect. 4 focusing on a systematic flow of information. First, we explained the basic concept of deep learning models. Afterward, we explored its applications in recommender systems and provided detailed analysis of key recommender systems using those techniques.

2 Background

This section expounds on the major concepts introduced in the paper, namely recommender systems, machine learning, and deep learning.

2.1 Recommender Systems

Recommender systems use machine learning and artificial intelligence algorithms to predict items to users. In today's world, the availability of such a vast pool of data makes it difficult for e-commerce websites to identify user-centric items [1]. If a user intends to select an item, he might not be able to find the item most suitable to him due to the presence of multiple options. To make this task easy, the concept of recommender systems was introduced. RS uses machine learning and artificial intelligence techniques to recommend relevant items to the user.

2.1.1 Content-Based Filtering

In content-based filtering, the recommendations are performed based on the previous choices of the user. As the name suggests, this type of filtering technique depends upon the various parameters of the user items. These parameters, when taken into account, reveal the granular level information about the user's preferences. Content-based filtering aims to exploit such a detailed analysis of the user's liking and recommends items in the present. The recommender system studies the choices user made in the past in different domains and depending upon those choices, the user is recommended items [2]. Such choices made by the user depend on various parameters or features of the items. Since these recommendations match the preferences the user made in the past, hence they are more user-centric. To explain with the help of an example of content-based recommendations, let us say that a user likes to watch movies by the director *Quentin Tarantino*, so the movies which will be recommended to the user will feature the same director. This was content-based filtering based on just one feature. Similarly, we can take multiple features to determine the recommendation list. Mathemati-

cally, content-based RS for multiple users can be represented using Eq. 1:

$$\min_{\theta^1,\ldots,\theta^n} \frac{1}{2}\sum_{j=1}^{n}\sum_{i:r(i,j)=1}\left(\left(\theta^j\right)^T x^i - y^{(i,j)}\right)^2 + \frac{\lambda}{2}\sum_{j=1}^{n}\sum_{k=1}^{n}\left(\theta_k^j\right)^2. \tag{1}$$

Here, θ^j represents the parameter vector for user *j*, *r(i,j)= 1* if user *j* has rated the movie *i*, else 0, x^i represents the feature vector for movie *i*, and $y^{(i,j)}$ represents the rating for movie *i* by user *j*.

2.1.2 Collaborative Filtering

The second type of recommender system is collaborative filtering-based Recommender System. In this type of filtering, the recommendations are affected by the neighbors of the user. It is seen that if a user's friend buys a particular product, the user is likely to buy the same or a similar product. Based on this principle, collaborative filtering adopts a neighborhood-based approach. In this technique, the user's neighbors' choices are studied, and based on those choices, similar items are recommended to the user. User-based collaborative filtering is essentially a neighborhood-based approach. To explain with an example of how collaborative filtering works, let us say that *user a* has a neighbor *user b. User b* likes a particular product, say *product p*. Now in this filtering technique, the preference of *user a* will be subjected to the preferences of *user b* and vice versa. Hence, *user a* will be recommended *product p* and products similar to *product p*.

Collaborative filtering is further of two types, namely, item-based collaborative filtering and user-based collaborative filtering.

a) Item-based Collaborative Filtering

In this type of collaborative filtering technique, the recommendations are based on the similarity between the items. If a user purchases *product p* and *product q* is similar to *product p* in one or more ways, i.e., if they share some common features, the user is recommended *product q* and other products similar to *product q*. The similarity between the items is calculated using several similarity measures. It can be *cosine-based similarity,* which can be calculated using Eq. 2. It is also known as *vector-based similarity,* and the similarity is determined by calculating the cosine of the angle between the two vectors.

$$\text{sim}(a,b) = \cos\left(\vec{a},\vec{b}\right) = \frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\| * \|\vec{b}\|}. \tag{2}$$

Here, $\vec{a}$ and $\vec{b}$ are the two item vectors,

Another similarity measure that can be used to deduce the similarity between different items is *Pearson correlation-based similarity.* It can be given by Eq. 3.

$$\text{sim}(a,b) = \frac{\Sigma_{u\varepsilon U}\left(R_{u,a}-\bar{R}_a\right)\left(R_{u,b}-\bar{R}_b\right)}{\sqrt{\Sigma_{u\varepsilon U}\left(R_{u,a}-\bar{R}_a\right)^2}\sqrt{\Sigma_{u\varepsilon U}\left(R_{u,b}-\bar{R}_b\right)^2}}. \tag{3}$$

Here, $R_{u,a}$ denotes the rating given by *user u* for *item a*, $\bar{R}_a$ denotes the average rating of *item a.*

Similarly, $R_{u,b}$ denotes the rating given by *user u* for *item b*, and $\bar{R}_b$ denotes the average rating of *item b.*

b) User-Based Collaborative Filtering

In this type of collaborative filtering technique, the recommendations take place depending upon the preferences of the user's neighbors. If *user u* has a neighbor *user v*, and *user v* likes *product p*, then *user u* will be recommended *product p* and all such products preferred by *user v*.

2.1.3 Hybrid Filtering

The third approach is the hybrid approach. It is an amalgamation of content-based filtering and collaborative filtering. Here, first, the social network analysis of a user's neighborhood takes place. This step helps identify the neighbors' preferences, and items are kept in the account to be recommended to the user. In the second step, the user's history is studied, and depending on the items he bought previously and items having similar content or attributes as the past items; item recommendations are taken into account [3]. Finally, analyzing the results obtained in both the steps, the user is recommended the items. Authors in [4] differentiated user's preferences in generated recommendations by using a deep hybrid recommender system. In order to explain the underlying literature of hybrid filtering, let us take the example used in the previous Sect. 2.1.2. Let us say that *product p* has a similar product, *product q,* along with one or several features. The system will apply content-based filtering to identify products similar to *product p* and *product q*. Upon selecting such products, say *list l*, the system will use collaborative filtering, and *user b* will now be recommended *product p*, *product q*, and other similar products, i.e., products from *list l.*

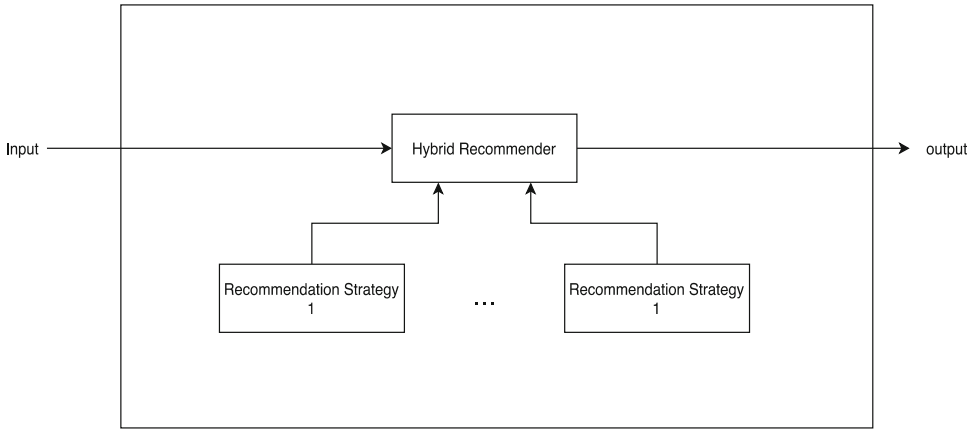


Fig. 1 Monolithic hybrid design

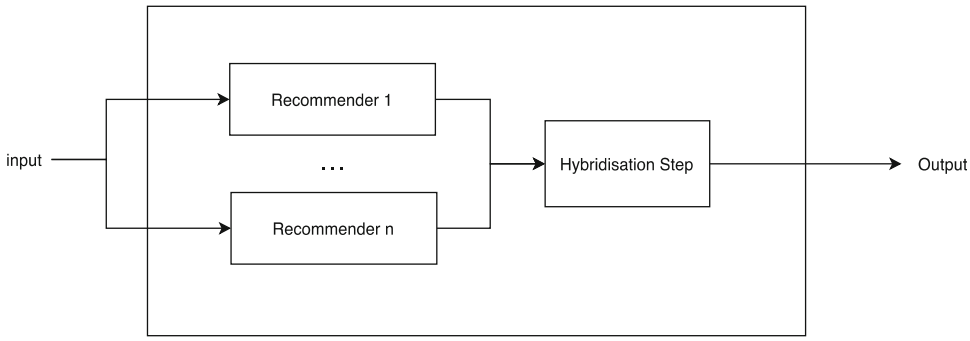


Fig. 2 Architecture of parallelized hybrid design

Any of the three techniques can be used to recommend items to users. Depending upon the chosen technique, the recommendations are made to the user. Hybrid recommender systems can be categorized into the following types:

a) Monolithic Hybrid Design

In this type, there exists a single recommendation component that aggregates multiple recommendation methods by preprocessing and combining numerous knowledge sources. The architecture of monolithic hybrid design is represented in Fig. 1.

b) Parallelized Hybrid Design

In this type, several recommender systems are run in parallel, and the output of each is combined at the later stages using an aggregation mechanism. These are further of three types, i.e., *mixed, weighted,* and *switching*. The architecture of parallelized hybrid design is shown in Fig. 2.

c) Pipelined Hybrid Design

In this type, every recommender system processes the input pertaining to its recommendation mechanism. The output, hence produced, is forwarded as input to subsequent recommender systems mimicking a staged process. The architecture of the pipelined hybrid design is given in Fig. 3.

2.2 Machine Learning

Machine learning is a section in computer science that deals with computing and solving problems intelligently and analyzing the results. The system is presented with a dataset, and various machine learning algorithms can be applied to that dataset to obtain results. Due to the availability of such extensive data and the need to get predictive results, machine learning algorithms are used widely. There exist three types of machine learning techniques, which can be seen in Fig. 4.

2.2.1 Supervised Learning

In supervised machine learning algorithms, the system is presented with the dataset, and the outcome is pre-defined. For example, it is known that the outcome has to be whether a given transaction is fraudulent or genuine for identifying fraudulent bank transactions. Since the result is known beforehand, the algorithm works to guide the expected result [5]. Hence, the entire working is supervised. A few supervised learning algorithms are Logistic Regression, Neural Network, Naïve Bayes, Decision Trees, Nearest Neighbors, and others.

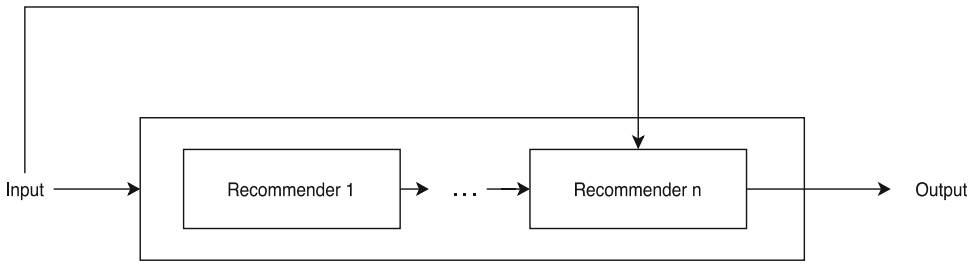


Fig. 3 Architecture of pipelined hybrid design

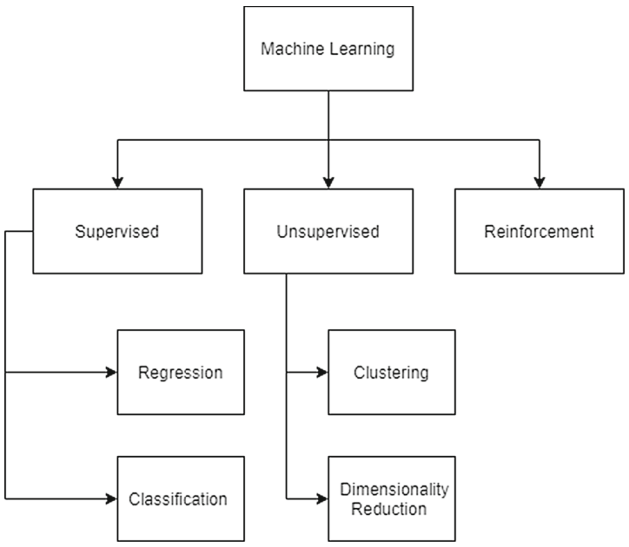


Fig. 4 Machine learning architecture

Semi-supervised learning is another type of technique in machine learning [6]. This technique is used when the expected result is known only for a few data points, i.e., when in a given dataset, not every data point is labeled. It helps in identifying and learning the structure of the input variables.

2.2.2 Unsupervised Learning

The second type of technique is unsupervised learning. In this type of technique, the data is not classified. The expected results are not known before its implementation. The system is presented with the dataset, and any of the unsupervised machine learning algorithms are applied to it, and the results are obtained. Since there is no supervision on the desired results, this technique is called unsupervised learning [7]. The main motive behind using this technique is to identify hidden clusters and patterns in the data. Examples of such algorithms are K-means clustering, self-organizing maps (SOM), hierarchical clustering, and others.

2.2.3 Reinforcement Learning

In reinforcement learning, the algorithm works on the concept of rewards and penalty. None of the data points in the dataset are labeled, i.e., the expected output is unknown, but the learning takes place in such a way that the system learns the environment on-the-go. The system is given an environment, and a set of actions are defined. Depending on the type of action the system takes, it is rewarded [8]. If the system's action takes it toward the goal state, it is rewarded. However, if the system's action takes it away from the goal state, it is charged with a penalty. To carry out this task, an objective function is determined, including all the possible action and state spaces. The main aim is to maximize this objective function, and policy is defined, which is to be followed. The most widely accepted and real-life example of reinforcement learning is how a toddler learns about the positive and negative outcomes of his actions by experiencing the rewards and penalties of those actions.

Machine learning algorithms can be employed in implementing recommender systems. Based on the type of recommender system being used, the data is collected. In collaborative filtering, the data is collected by performing social network analysis and identifying the items bought or liked by user's neighbors. These data points or choices serve as the dataset for implementing the algorithm. If content-based filtering is used, the user's past choices are identified as the data points. For example, for a movie recommender system, the movies user has watched in the past, the genre, director, actors, and others are taken into account, and these bits of information constitute the dataset. Further, algorithms are implemented on this dataset to recommend items. Lastly, hybrid filtering can be used for collating data using both supervised and unsupervised learning and presenting the final data points to the system.

2.3 Deep Learning

Deep learning is a discipline of machine learning algorithms that are layered. Here each layer does nonlinear processing producing different abstractions of the data. Each layer takes the output of previous layers as input, hence producing hierarchical abstractions. For example, in computer vision, the original image matrix is the data that is processed. The first layer takes this data, and it performs feature extraction and transformation in such a way that it identifies the edges. It then gives this extracted feature as an input to the next layer, which may create a new layer of abstraction by identifying

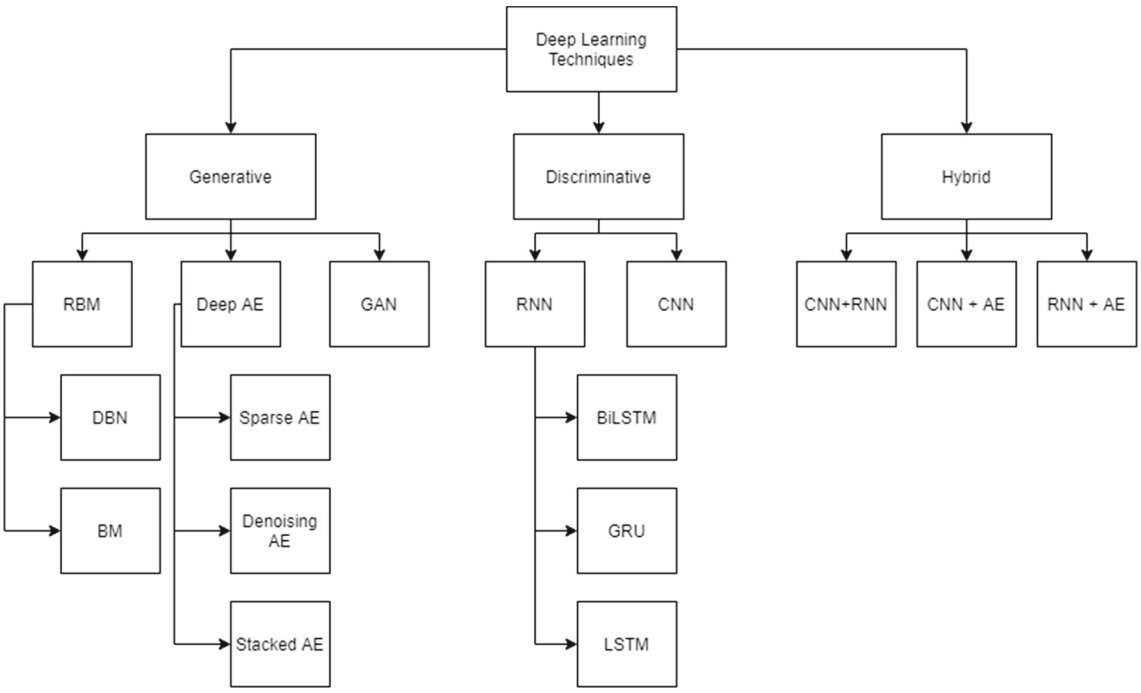


Fig. 5 Deep learning techniques

the orientation of the edges detected by the previous layer. This is repeated numerous times with multiple layers.

In this paper, as depicted in Fig. 5, research work done using several deep learning techniques has been studied.

2.3.1 Why Use Deep Learning?

The web is a vast pool of data; the dimensionality and modality of data present online are very high. To work with such multimodal and complex data with extensive features requires extensive machine learning support. If we use traditional design and technology, the recommendations generated will not be of any use because the system will not exploit the information at hand, and the complex pattern analysis of data would not be done. To solve this problem, we have used deep learning, which can work on highly complex data with hidden features and a high number of training instances. Deep learning's capability provides breakthrough results by extracting the data on a granular level and analyzing the patterns. Hence, the representation of all the details of data in a joint unified framework is made possible by using deep learning [9].

Conventional machine learning models like Matrix Factorization and others are linear models that can work efficiently on linear data, but when we have to deal with nonlinear data, the computations require complex functions like sigmoid, tanh, etc. To catch such intricate user–item interaction patterns, there arises a need for deep learning models.

The web is full of illustrative data; even for making recommendations, there is a lot of description and content attached to data. To understand this information and use it to our advantage, we have employed recommender systems. Another significant advantage of using deep learning is simplifying the process of feature engineering.

Every set of data has sophisticated features attached to it, and the task of deep learning is to understand the raw features and process them using supervised or unsupervised machine learning algorithms. Similarly, even for the content to every set of data, deep learning makes it possible for the system to use all the available data to generate expert recommendations.

Another important reason for using deep learning is sequential pattern mining. The task of sequence modeling that includes Natural Language Processing, speech recognition, etc. is usually performed by either recurrent neural network (RNN), which has internal memory to learn the next in sequence, or convolutional neural network (CNN) that performs the sequence modeling using temporal computations. We have explained this further in Sects. 4.3 and 4.6, respectively (Table 1).

3 Research Methodology

In this paper, various parameters were considered for identifying the relevant research papers to be included in the study. The broad area of consideration was deep learning in recommender systems. In this section, the research methodology adopted in this paper has been explored. Broadly, the research methods have been based upon ten attributes, i.e., publisher

Table 1 Application domains of studies

Application domain	Studies
Item recommendation	[10–19]
News recommendation	[20–25]
Movie recommendation	[26–31]
Cold-start problem	[32–36]
Session-based recommendation	[37–41]
Music recommendation	[42–45]
Text recommendation	[46–50]
Image recommendations	[51–53]
Social network-based recommendation	[54–56]
Hashtag recommendation	[57–59]
POI recommendation	[60–62]
Citation recommendation	[63–65]
Content-based recommendations	[66–68]
Video recommendation	[69, 70]
Quote recommendation	[71, 72]
E-learning recommendation	[73, 74]
Rating prediction	[47, 75]
Job recommendation	[33, 76]
Fashion recommendation	[77, 78]
Blog recommendation	[79]
Venue recommendation	[80]
Consumer preferences	[81]
Co-evolutionary latent feature processes	[82]
Artwork recommendation	[83]
Adaptive user-interfaces	[84]
Audience activity recommendation	[85]
Data sparsity	[86]
Healthcare service recommendation	[87]
Treatment recommendations	[88]

of the paper, year of publication, number of citations, and location of performing the study (Table 2).

3.1 Application Domain of Recommender Implementing Deep Learning

RQ 1 Which application domains have been included in the study?

There is a need for a dataset to apply machine learning or artificial intelligence algorithms to implement a recommender system. The dataset depends on the application domain of the recommender system. Table 1 presents all the application domains of the refereed papers and several studies performed in that domain.

It can be observed from the table that maximum work has been done in the domain of item recommendations, news recommendations, and movie recommendations.

Table 2 Dataset from external sources

Dataset	Studies
MovieLens	[13, 18, 23, 26–29, 31, 47, 48, 65, 74, 81, 85, 89–97]
Amazon	[10, 12, 15, 19, 36, 47, 48, 75, 86, 94, 95, 98–100]
Yelp	[12, 15, 16, 19, 60, 61, 75, 82, 86, 98]
IMDb	[29, 35, 41, 47, 96, 101]
NetFlix	[35, 69, 96, 101, 102]
CiteULike	[49, 89, 92, 102, 103]
Epinions	[52, 55, 87, 93]
Last.fm	[13, 41, 58, 96]
Delicious	[34, 58, 90, 96]
Million song dataset	[43–45, 100]
BookCrossing	[85, 97, 100]
FourSquare	[60, 61, 80]
The echo nest taste profile subset	[42, 44]
Dbbook	[46]
Wikiquote website	[71, 72]
Flixter	[54, 56]
Ciao	[55, 56]
YooChoose	[67, 100]
FilmTrust	[56, 85]
Twitter	[17]
Google news	[57]
Tumblr	[79]
Flickr	[80]
Picasa	[80]
Oxford concise dictionary of proverbs	[71]
Google Play	[104]
IPTV	[82]
YouTube	[70]
UGallery	[83]
Rossmann	[90]
Frappe	[91]
CLEF NewsREEL	[22]
Penn treebank	[100]
CADE web directory	[100]
RefSeer	[63]
Reddit	[41]
Economics thesaurus (STW)	[64]
EconBizRecSys evaluation dataset	[64]
Jester	[14]
PubMed	[65]
EqGraph	[84]
MSWeb	[84]
Assistments	[84]

Table 2 continued

Dataset	Studies
Beer	[15]
Douban	[93]
NAVER news	[24]
Instagram	[77]
Zalando	[77]
Trip.com	[17]
Facebook	[17]
Exact Street2Shop	[78]
Taobao advertising dataset	[18]
Meetup	[96]
DeepSurv	[88]

Table 3 Self-generated dataset

Self-generated dataset	Studies
1. Google chrome extension "Daum News Tracker" 2. Android application "KECI News.	[20]
1. Logs scraped from search engine Bing Web vertical 2 News article browsing history from Bing News vertical 3 App download logs from Windows AppStore 4 Movie/TV view logs from Xbox	[11]
User's news click history between 04/01/2014 and 09/30/2014	[21]
The images and users' information in this dataset were crawled from Flickr through its API	[51]
1. VIDXL—was collected over a 2-month period from a video site 2. CLASS—product view events of an online classified site.	[38]
"starc" platform which is based on Open edx for self-development which serves the fundamental education field	[73]
Collected from real-life users' browsing history of an online European department store over two weeks at the beginning of January 2016	[39]
Scraped Ukiyo-e images from the "the Ukiyo-e search service" and the Ukiyo-e pages of Ritsumeikan University Art Research Center	[52]
Wanted and missing persons' application of the Police of the Czech Republic	[53]
Collected 419,509 check-in records published by 49,823 users among 18,899 locations from August 2012 to July 2013 in Manhattan via the API of Foursquare	[62]
A large collection of quality article selection demonstration with an average length of 900 characters over six months, manually created by professional editors	[50]
1314 resumes which came in as a part of summer research intern application at IBM Research Labs	[76]
Sampled offline dataset collected from a commercial news recommendation application	[25]

Table 4 Type of recommender systems

Type of recommender system	Studies
Content based	[11, 13, 20, 21, 24, 25, 27, 30, 38, 45, 46, 50, 53, 65, 71–73, 78–80, 88, 99, 100]
Collaborative filtering	[10, 15, 17–19, 22, 23, 26, 31, 33–35, 39, 41, 48, 49, 52, 54–56, 58–61, 69, 70, 74, 75, 81, 84, 87, 89, 90, 92–94, 96, 98, 101, 103, 104]
Hybrid	[16, 28, 29, 42, 51, 68, 77, 97, 102]
Context-aware	[32, 47, 62, 63, 67, 85, 86, 95, 104]

3.2 Datasets Involved in the Referred Studies

RQ 2 What are the different pre-defined and self-generated datasets other authors have worked on in the past?

In this section, the datasets used in the research papers have been listed. In some papers, authors used publicly available datasets to perform the recommendation process using their proposed model. In other papers, authors scraped data from websites or applications using APIs or manually collected data to implement recommender systems. Table 2 enlists the external datasets used in previous studies, and Table 3 enlists self-generated datasets.

It is evident from Table 2 that the MovieLens dataset is extensively used for analyzing the effectiveness of implementing deep learning in recommender systems. Other widely used datasets are Amazon and Yelp.

3.3 Types of Recommender Systems Used in the Studied Papers

RQ 3 What types of recommender systems have been used in the past studies?

There are various types of recommender systems, depending on the recommendation technique used. In this section, the types of recommender systems used in the studied papers have been listed along with the studies they were used in.

It is evident from Table 4 that most of the studies have been done in content-based and collaborative filtering algorithms. This presents the readers an opportunity to work in other hybrids, context-aware, and other similar recommendation techniques.

3.4 Deep Learning Techniques Used in the papers

RQ 4 Which deep learning techniques have been used in the past research?

To efficiently implement the recommendation process, several deep learning techniques were used in the referred

Table 5 Classification on the basis of deep learning technique used

Deep learning technique	Studies
Recurrent neural network (RNN)	[10, 13, 21, 24, 37–41, 45–47, 49, 60, 62, 63, 65, 67, 71, 72, 74, 75], [41, 82, 84, 85, 99]
Deep neural network	[14, 16–18, 22, 27, 44, 53, 56, 59, 70, 83, 86, 90, 93, 95, 100, 104, 105]
Convolutional neural network (CNN)	[12, 15, 19, 24, 48, 50, 52, 60, 71, 76–80, 93, 94, 98, 102]
Deep belief network (DBN)	[20, 28, 36, 42, 61, 69, 73, 81]
Deep feed-forward neural network	[29, 31, 55, 57, 88, 104]
Stacked denoising autoencoder (SdA)	[23, 35, 94, 97, 101]
Collaborative deep learning (CDL)	[33, 51, 89, 102]
Sparse autoencoder	[58]
multi-view deep neural network (MV-DNN)	[11]
Supervised neural network	[26]
Deep learning matcher (DLM)	[32]
Variational autoencoder	[103]
Generative adversarial network (GAN)	[92]
Denoising autoencoder	[34]
Deep density networks (DDN)	[68]
Neural matrix factorization (NeuMF)	[96]

Table 6 Metrics used in the papers

	Metric
A	Time
B	Accuracy
C	Mean square error (MSE)
D	Root mean square error (RMSE)
E	Recall
F	Mean average precision (MAP)
G	Precision
H	Mean reciprocal rank (MRR)
I	Mean absolute error (MAE)
J	Rank score
K	Area under curve (AUC)
L	Normalized discounted cumulative gain (nDCG)
M	Pearson correlation
N	Hit ratio
O	F1-score
P	Mean average rank (MAR)
Q	Coverage

research work. Table 5 lists all the techniques used in the papers and hence, classifies the research work.

It can be seen from the table that the recurrent neural network (RNN) is the most used deep learning technique in recommender systems. Hence, users can try to implement other deep learning techniques and perform a comparative study.

3.5 Metrics Used for Analyzing the Results by Deploying Deep Learning in Recommender Systems

RQ 5 Which metrics have been used to analyze the results?

The result obtained by implementing the recommender systems is analyzed by using various metrics. Different papers deploy different metrics to get a multi-dimensional interpretation of the results. Table 6 enlists the metrics used in the papers. Figure 6 graphically analyzes the metrics used.

This analysis of metrics shows that the recall metric is used by maximum researchers to analyze their results. The graph also helps in analyzing the usage gaps between all the metrics (Fig. 7).

4 Related Work and Original Contribution of the Paper

Some of the significant works in deep learning in recommender Systems are covered in Table 7.

For performing the research papers' complete study, various aspects and attributes of these studies were divided into columns, and all such columns were collated together to form a spreadsheet. Such information clusters included the problem statement, the dataset, the proposed model, deep learning technique, type of recommender system used, and others. Later in the Conclusions section, the significant research gaps were identified and reported, extracted from the studied research papers. In this section, a holistic study of all the research papers has been presented.

In this subsection, the problems addressed in different papers have been categorized into eleven clusters of deep learning techniques as presented henceforth. Every subsection expounds a detailed research work done in recommender systems for each deep learning technique. At the end of every subsection, a detailed description of the findings and open issues have been presented.

4.1 Multilayer Perceptron (MLP)

Multilayer Perceptron (MLP) is a feed-forward network with one or several computation layers and has nonlinear activations. It has at least one layer that is connected in a

Table 7 Related work

S. No.	Title	Problem	Solution	Dataset	Deep learning technique	Metric used	RS technique	Application domain
1	Toward Twitter Hashtag Recommendation using Distributed Word Representations and a Deep Feed Forward Neural Network	The use of hashtags is restricted due to the additional user effort required in assigning them	1. Use of word2vec model to learn the representation of word 2. Use of deep feed-forward neural network	Google News dataset	Deep Feed-Forward Neural Network	MSE	Content based	Hashtag recommendation
2	Improving Content-based and Hybrid Music Recommendation using Deep Learning	Need to improve the music recommendations because the traditional features are unable to encapsulate all the pertinent information in the audio	Merged the two tech-niques—Deep Belief Network and probabilistic graphical model into an algorithm that learned audio features while making personalized recommendations	The Echo Nest Taste Pro le Subset	Deep Belief Network	Root Mean Square Error (RMSE)	Hybrid and content based	Music Recommendation
3	Deep Learning of Semantic Word Representations to Implement a Content-based Recommender for the RecSys Challenge'14	While reporting a document in the final feature space, there is no consideration of the similarity of the word semantics	Found the current and latent semantic features relating to the data of documents and fitted a linear regression method to estimate user preferences for documents.	Dbbook	Recurrent Neural Network (RNN)	Root Mean Squared Error (RMSE)	Content based	Text Recommendation

Table 7 continued

S. No.	Title	Problem	Solution	Dataset	Deep learning technique	Metric used	RS technique	Application domain
4	Collaborative Deep Learning for Recommender Systems	In traditional CF-based recommender systems, the ratings are generally very sparse in several applications, this results in CF-based methods to devalue remarkably in their recommendation performance	Implemented collaborative deep learning that carried out deep representation learning for the content information and collaborative filtering for the feedback matrix	CiteULike, Netflix	Collaborative Deep Learning (CDL)	Recall, mean average precision (mAP)	Content based and CF	General
5	The video recommendation system based on DBN	To improvise video recommendations	Expanded the traditional depth of the Deep Belief Network model together with CF-based algorithm to construct a video recommender system	Netflix movie score datasets published in 2005	Restricted Boltzmann Machine (RBM) is the base of DBN (Deep Belief Network)	mean square error (MSE), MAE, RMSE	CF	video recommendation
6	Tag-aware recommender systems based on deep neural networks	Data sparsity, ambiguity, and redundancy are some issues which occur in the user-defined tags	To make recommendations, extracted features were used which were more abstract, dense and representative	Last.fm, Del.icio.us,	Deep Neural Network Sparse Autoencoder	Precision, recall, rank score	User-based CF	tag information

Table 7 continued

S. No.	Title	Problem	Solution	Dataset	Deep learning technique	Metric used	RS technique	Application domain
7	Multi-Rate Deep Learning for Temporal Recommendation	To remediate the problem of tackling large-scale real-time data to model temporal behavior	Implemented a deep neural network-based model to amalgamate the preferences of static long-term as well as temporal short-term to enhance the recommendations	user's news click history between 04/01/2014 and 09/30/2014	Recurrent Neural Networks (RNN)	Precision, Area Under Curve (AUC), Mean Average Precision (MAP) and Mean Reciprocal Rank (MRR)	Content based	News recommendation system
8	Quote Recommendation in Dialogue using Deep Neural Network	To improvise the quote recommender system	Implemented a deep learning model to merge RNN and CNN to learn the semantic rendering for every dialog	Wikiquote website and Oxford Concise Dictionary of Proverbs	RNN and CNN	MRR, Recall, NDCG, Hit	Content based	quote recommendation
9	Solving Cold-Start Problem in Large-scale Recommendation Engines: A Deep Learning Approach	Remediating the cold-start problem in CF-based recommender systems	Implemented deep learning-based item–item matcher pertaining to the similarity in the documents using doc2vec model	Careerbuilder's CF-based recommendation engine	Deep Learning Matcher (DLM) was built utilizing doc2vec which is considered the current state-of-the-art deep learning algorithm for document embedding and matching	Pearson-r correlation, recall, precision, running time	Context based	Cold-Start Problem

Table 7 continued

S. No.	Title	Problem	Solution	Dataset	Deep learning technique	Metric used	RS technique	Application domain
10	Deep Learning with Consumer Preferences for Recommender System	To improve the problem of data sparsity	Realized a deep learning technique to anticipate the values of null ratings	MovieLens	Restricted Boltzmann Machines (RBMs) can be stacked and trained in a greedy manner to form so-called deep belief networks (DBNs)	Pearson Correlation Similarity(PCC), MAE	Collaborative filtering(CF)	Consumer Preferences
11	Comparative Deep Learning of Hybrid Representations for Image Recommendations	Certain tasks which are specific to the users, like image recommendations, require efficient representations of images and preferences of users over images.	Proposed a dual-deep net, wherein the two subnetworks map input images and liking of users into the same latent semantic space. Afterward, the distances between images and users in the latent space are deduced to make recommendations.	The images and users' information in this dataset are crawled from Flickr through its API	Collaborative Deep Learning (CDL)	Precision, Recall	Hybrid	Image Recommendations
12	"Tag-Aware Personalized Recommendation Using a Deep-Semantic Similarity Model with Negative Sampling"	To tackle the problem of redundancy, sparsity and ambiguity of social tags.	Implemented deep learning by including the tag-based user and item profiles to an abstract deep feature space, to maximize the resemblance between the deep-semantic users and their desired items		Deep Neural Network	recall, precision, F1-score, mean average precision (MAP) and mean reciprocal rank (MRR)	CF	Tag-Aware Personalized Recommendation

Table 7 continued

S. No.	Title	Problem	Solution	Dataset	Deep learning technique	Metric used	RS technique	Application domain
13	Wide & Deep Learning for Recommender Systems	To exploit and learn about the content and user interactions to make recommendations	Learned a high-dimensional semantic space independent of user interference, in which items were addressed based on their synonymity, and later learned a user-centric modification function to modify that space into a ranking pertaining to the previous preferences of the user	Google Play	Feed-Forward Neural Network	General	Context based	General
14	Parallel Recurrent Neural Network Architectures for Feature-rich Session-based Recommendations	To work on the drawback of real-time recommender systems which make recommendation based on the clicks of the user during a session	Implemented parallel recurrent neural network to realize sessions specific to the clicks and the images and textual features of the clicked items	1. VIDXL—was collected over a 2-month period from a YouTube—like video site 2. CLASS—product view events of an online classified site	Parallel RNN (p-RNN)	Recall, MRR	Content based	Session-based Recommendations
15	"Collaborative Filtering and Deep Learning Based Hybrid Recommendation For Cold Start Problem"	To remediate the data sparsity and cold-start problems	Implemented an recommender system that was a hybrid of item content features extracted from a deep neural network and its application to the timeSVD ++ CF model	Netflix, IMDB	SDAE	RMSE	CF	Cold-start problem

Table 7 continued

S. No.	Title	Problem	Solution	Dataset	Deep learning technique	Metric used	RS technique	Application domain
16	Application of deep belief nets for collaborative filtering	Working toward the problem of poor recommendation accuracy caused due to data sparsity	Extracted user features using Deep Belief Network and then applied K-nearest neighbor algorithm to make recommendations	MovieLens	Restricted Boltzmann Machines	RMSE	Hybrid	Movie recommendation
17	"DBNCF: Personalized Courses Recommendation System Based on DBN in MOOC Environment"	Low effectiveness of resource recommendation in MOOC environment due to data sparsity of online learners	Implemented Deep Belief Network by combining the online learner course feature vector for mining the learner's interest and using the course score as the class labels	"starc" platform which is based on Open Edx for self-development which serves fundamental education field	DBN	RMSE	Content based	MOOC Courses Recommendation
18	On Deep Learning for Trust-Aware Recommendations in Social Networks	The recommendations depend largely upon initializing the user and item latent feature vectors	Used deep learning to initialize in matrix factorization for trust-aware social recommendations and to study the result of the community in user's trusted friendships	Epinions and Flixster	Deep Autoencoder, which is an efficient approach to nonlinear dimensionality reduction	root mean squared error (RMSE), coverage, F-measure, precision	CF	social network-based recommendation
19	dTrust: a simple deep learning approach for social recommendation	To improve the privacy in social recommender systems	Built a system that depends on the topology of an anonymous trust-user item network which amalgamates user trust relations and rating scores	Epinions and Ciao	Deep Feed-Forward Neural Network	Root Mean Square Error (RMSE) and Mean Absolute Error (MAE)	CF	social recommendation

Table 7 continued

S. No.	Title	Problem	Solution	Dataset	Deep learning technique	Metric used	RS technique	Application domain
20	Application of Deep Learning to Sentiment Analysis for Recommender System on Cloud	The skewed contextual information result in inefficient sentiment analysis of short texts	Used deep learning to optimize the recommendations based on the sentiment analysis for various reviews retrieved from several online social networking websites.	Amazon	RNN	accuracy, F1-Measure, Response Time	Content based	Text Recommendation
21	A Recommendation Model Based on Deep Neural Network	To solve the problem of data sparsity	Implemented a system which took into account the tag and user information and retrieved the K-nearest neighbors	MovieLens, Epinions	Deep Neural Network (DNN).	Mean Absolute Error (MAE) method and the Root Mean Squared Error (RMSE)	CF	General
22	Neural Citation Network for Context-Aware Citation Recommendation	To solve the problem of appropriate citations in a fast-growing academic world	Proposed a context-aware citation recommender system	RefSeer dataset	Recurrent Neural Network	Recall, Mean Average Precision (MAP), Mean Reciprocal Rank (MRR) and Normalized discounted cumulative gain (NDCG)	Context based	Citation Recommendation
23	A Deep Architecture for Content-based Recommendations Exploiting Recurrent Neural Networks	To curb the problem of varied lengths of data to report the content of items to be recommended	Implemented a deep learning based Long Short-Term Memory (LSTM) network to collectively learn embeddings depicting the items to be recommended and user's preferences	MovieLens, Dbbook	RNN	F1-Measure	Content based	Content-based Recommendations

Table 7 continued

S. No.	Title	Problem	Solution	Dataset	Deep learning technique	Metric used	RS technique	Application domain
24	Recommendation with Social Relationships via Deep Learning	To improve social recommendations	Proposed a deep learning approach to improve social recommendations	Epinions, Flixster, Ciao and FilmTrust	Deep Neural Network	Recall, Ndcg	CF	Recommendation with Social Relationships
25	Recurrent Latent Variable Networks for Session-Based Recommendation	To curb the limitation of the sequential nature of the addressed predictive setup, and data sparsity	Proposed an algorithm wherein, the network recurrent units were considered as stochastic latent variables with a previous distribution assigned to them		Recurrent Neural Network	Recall, Mean Reciprocal Rank (MRR)	Content based	Session-Based Recommendation
26	A Deep Multimodal Approach for Cold-start Music Recommendation	To solve the problem of cold-start for new artists in music recommender systems	Applied deep learning on text and audio information alongside user feedback data	Million Song Dataset (MSD), Echo Nest Taste Profile Subset	Deep Neural Network	General	collaborative filtering(CF)	Music Recommendation
27	Re-ranking-based Recommender System with Deep Learning	To solve the problem of data overload of scientific papers for researchers and scholars	Proposed a deep learning technique which determined top-k recommendations and re-ranked the results	Economics thesaurus (STW), EconBizRecSys evaluation dataset	Deep Neural Network (DNN)	normalized discounted cumulative gain (nDCG)	Content based	scientific papers

Table 7 continued

S. No.	Title	Problem	Solution	Dataset	Deep learning technique	Metric used	RS technique	Application domain
28	Attention-based Recurrent Neural Network for Location Recommendation	To improve the shortcomings of the traditional POI recommender system	Proposed a deep neural network to analyze the sequential check-ins of users and capture their life patterns	collected 419,509 check-in records published by 49,823 users among 18,899 locations from August 2012 to July 2013 in Manhattan via the API of Foursquare	Recurrent Neural Network (RNN)	Precision, Recall	Context based	Point-of-Interest (POI) recommendation
29	Contextual Sequence Modeling for Recommendation with Recurrent Neural Networks	To work toward the inclusion of the side information in Recurrent Neural Networks	Implemented a Context-aware Recurrent Neural Network (CRNNs) that combined the contextual and the item embeddings	YooChoose dataset	Recurrent Neural Networks (RNNs)	Recall	context-aware recommendation	Recommendations involving contextual/side information
30	Product Recommendation: A Deep Learning Factorization Method Using Separate Learners	To improve the products recommendations	Implemented a deep neural network factorization method for separate learners	Jester	Deep Neural Network	Root mean squared error (RMSE)	Content based	product recommendation
31	Online News Recommender Based on Stacked Autoencoder	To solve the problems of data overloading, high-dimensionality, and data sparsity	Implemented a denoising stacked autoencoder to extricate the important low-dimension features from the sparse user–item matrices	MovieLens	Stacked Denoising Autoencoder (SDAE)	Mean Absolute Error (MAE), Precision, Recall	CF	News Recommendation

Table 7 continued

S. No.	Title	Problem	Solution	Dataset	Deep learning technique	Metric used	RS technique	Application domain
32	Deep Sequential Recommendation for Personalized Adaptive User Interfaces	To remediate the problem of data sparsity in Adaptive user-interfaces (AUIs).	Proposed an RNN based model that deployed Gated Recurrent Units to map user interaction histories to vectors in a Euclidean space exchanged among the user and desired item vectors	EqGraph, MSWeb and Assistments'15	RNN	mean average precision (MAP) and normalized discounted cumulative gain (NDCG)	CF	Adaptive user-interfaces (AUIs) based recommender
33	Spatial-Aware Hierarchical Collaborative Deep Learning for POI Recommendation	To improve data sparsity and cold-start problem	Implemented a Spatial-Aware Hierarchical Collaborative Deep Learning model (SH-CDL). The model carried out deep representation learning for Point of Interests from heterogeneous features and hierarchically additive representation learning for spatial-aware personal choices	Yelp, Foursquare	Deep Belief Network	Mean Absolute Error (MAE), Accuracy	CF	Point-of-interest (POI) recommendation

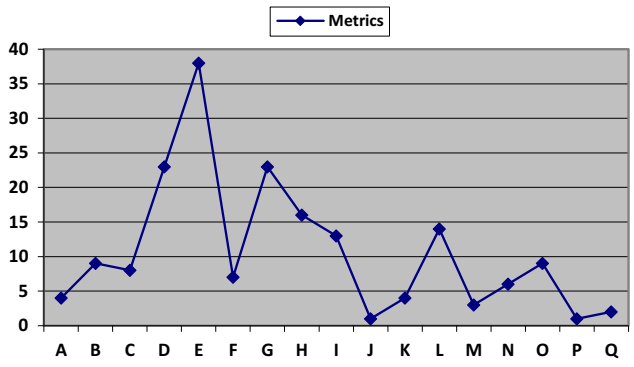


Fig. 6 Graphical analysis of metrics

feed-forward manner [106]. It can also be used to transform the linear methods of recommender systems into nonlinear models. Convolutional neural networks (CNN) are also a kind of feed-forward network.

In a deep forward neural network, the information flows in one direction across multiple layers, as shown in Fig. 7. The output from one layer becomes the input into the next layer. This architecture does not consist of any cycles, and hence it is called "feed-forward."

For a feed-forward neural network with D inputs, $x = [x_1,\ldots..,x_D]$, a layer with K hidden nodes $h = [h_1,\ldots..,h_K]$, then the output node y is given by Eq. 4:

$$y = v^T h = v^T f\left(w^T x\right) \tag{4}$$

where $v = [v_1,\ldots..,v_K] \in \mathbb{R}^K$, $W = [w_1, \ldots.., w_K] \in \mathbb{R}^{D\times K}$, f represents the nonlinear activation function.

The given Eq. 5 mathematically represents the value of every hidden node.

$$h_k = f\left(w_k^T x\right) = f\left(\sum_{d=1}^{D} w_{dk} x_d\right). \tag{5}$$

4.1.1 Wide & Deep Learning

The wide and deep learning model is capable of solving both regression and classification problems [107]. The wide learning element is a generalized linear model consisting of a single layer perceptron, whereas the deep learning element consists of the *multilayer perceptron*. The blend of the two above stated elements leads to the inclusion of both *memorization* and *generalization*. The ability of wide learning element to capture prominent features from historical data results in *memorization*. Whereas the ability of deep learning element to create general and abstract representations results in *generalization*. The amalgamation of the two results in diverse results and better performance of the recommender system.

The architecture of this model is given in Fig. 8.

This concept can be represented mathematically using Eq. 6, which represents the wide learning element, and Eq. 7, which represents the deep learning element, respectively. The final wide & deep learning model is represented by Eq. 8:

$$y = W_{\text{wide}}^T\{x, \phi(x)\} + b. \tag{6}$$

Here, W_{wide}^T and b represent the model parameters, x represents the raw input feature, and $\phi(x)$ represents the transformed feature.

$$\alpha^{l+1} = f\left(W_{\text{deep}}^l a^l + b^l\right). \tag{7}$$

Here, l represents the lth layer, $f(\cdot)$ represents the activation function, W_{deep}^l represents the weight term, and b^l represents the bias term

$$P\left(\hat{r}_{ui} = 1|x\right) = \sigma\left(W_{\text{wide}}^T\{x, \phi(x)\} + W_{\text{deep}}^T a^{l_f} + \text{bias}\right). \tag{8}$$

Here, $\sigma(\cdot)$ represents the sigmoid function, $\hat{r}_{ui}$ represents the binary rating label, a^{l_f} represents the final activation.

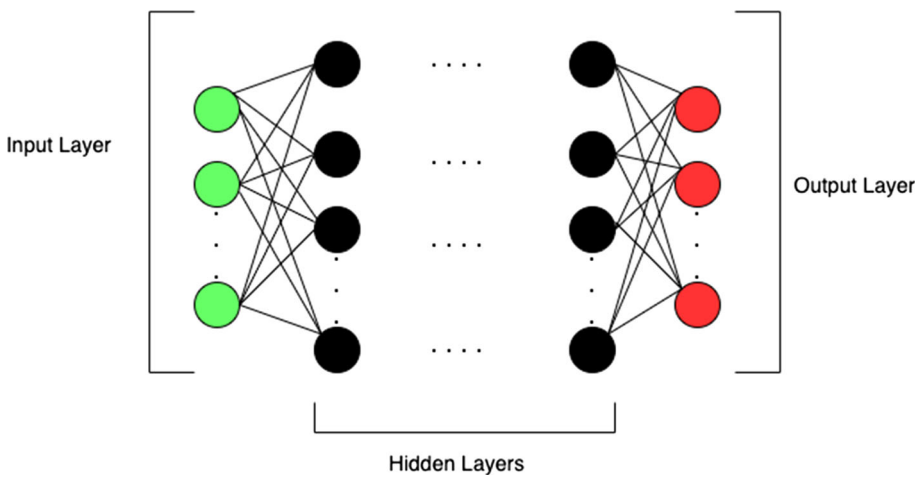


Fig. 7 Architecture of deep feed-forward neural network

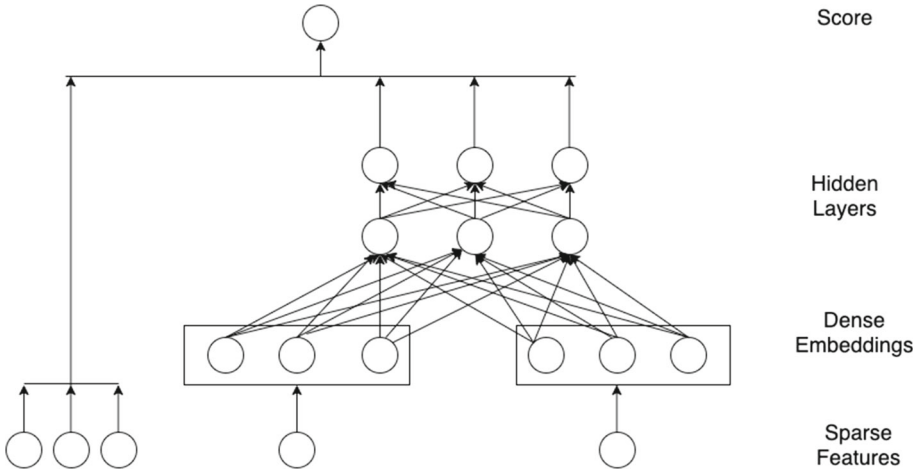


Fig. 8 Architecture of wide & deep network

Authors in [108] used a feed-forward multilayer neural network for collaborative filtering recommender systems. Once the model learned the hidden factors and features, the system feed-forwarded the latent features to identify <user,item> pairs having NULL ratings. The authors [57] realized that the usage of hashtags in social media required additional efforts by the user. Hence, they used the deep forward neural network to recommend relevant hashtags to the users. To carry out this task, first, they performed tweet collection and data preprocessing, followed by feature engineering by vector generation. In the end, they performed training and evaluation. In another work, the authors addressed the sparsity of data by using Wide & Deep learning by training comprehensive linear models with feed-forward deep neural networks [104].

It was found out that rating prediction techniques often rely on user's private information that is a threat to privacy. Hence, the authors came up with a feed-forward neural network-based dTrust model that used the topology of anonymous user–item interactions that combined user's trust relations with user rating scores [55]. In another implementation of the feed-forward deep neural network, the researchers addressed the 0/1 recommendation problem by combining the ratings given by users and Natural Language Processing (NLP) of texts [29]. For using hashtags, user-defined tags usually suffer from various problems like data sparsity, redundancy, ambiguity, and others. To solve this problem, the authors [58] abstracted features that were used to make recommendations instead of raw data. The authors also addressed this problem, and they solved it by using deep neural network [59].

In another work, the authors observed that since recommender systems influence both content and user interactions to create recommendations that adapt to the users' preferences, this can be used as a leverage to improve recommendations [27]. They proposed a Deep Space model that learned a user-independent high dimensional based on their substitutability, semantic space items were positioned, and according to the user's past preferences, it learned user-specific transformation function to convert this space into a ranking. In news recommender systems, modeling temporal behavior, the cost of estimating the parameters also increases, making the recommendations costly. The authors in [22] used the deep neural network to address this issue. It was observed by some researchers [56] that the preference of choosing friends in social media did not always match. Hence, it became challenging to recommend friends on online social networks. Thus, they introduced a deep learning approach to learn about both user preferences and the social influence of friends for an effective recommendation. In the paper [44], as a solution to the cold-start problem in music recommenders, the authors combined text and audio information with user feedback data using deep neural network frameworks.

Some authors in [53] observed that sometimes erroneous police photo lineups resulted in the conviction of innocent suspects. To avoid this, they came up with a two leveled approach. The first one was based on the visual descriptors of the deep neural network, and the other was based on the content-based features of persons. The authors [14] proposed a model based on the deep neural network for product recommendations, which required only ratings for making recommendations. In another work [18], the authors realized that the amount of calculation for the learned model to predict all user–item pairs' preferences was humungous. To solve this problem, they proposed a TDM attention-DNN (tree-based deep model using attention network). For an appropriate research paper recommendation to the scholars, a study [64] used the deep neural network with the paragraph vectors re-ranking method for adequate recommendations. It was observed by some researchers that side information written for business reviews were seldom taken into account for the recommendation [16]. Hence, they used an artificial neural network in a hybrid recommender with the inclusion

of side information. Some authors [17] addressed the issue of cross-domain in social recommendation. To solve this problem, they introduced the model Neural Social Collaborative Ranking (NSCR), which immaculately integrated user–item interactions of the information sector and user–user social relations of the social sector.

In a study, it was observed that the existing recommender systems did not take into account the effect of distrust among users [87]. They generate recommendations pertaining to the trust relations among users. As a solution, items were recommended for both the trust and distrust relations among active users. In this paper [109], the authors proposed a novel recommender system *RecDNNing* that combined the user and item embeddings using a deep neural network. First, the authors created deep embedding for users and items, and later the average and concatenated values of those embeddings are given as input to the system. The deep layers generated recommendations using the forward propagation method. The authors in [110] used the hybrid of content-based and collaborative filtering approaches to create a deep classification model for generating efficient music recommendations. To carry out this implementation, they came up with the Tunes Recommendation Systems (T-RecSys) algorithm. Authors in [111] proposed a deep learning-based novel collaborative filtering algorithm. The input of the system were the normalized values of user and item rating vectors. This resulted in decreasing the time complexity as the system need not learn the features of users and items. In paper [112], the researchers applied deep learning in the domain of agriculture. They used the Twitter platform to scrape agriculture tweets and applied sentiment analysis on those tweets. This helped the authors to predict the sentiment range of agriculture tweets.

4.2 Findings and Open Issues

Deep feed-forward networks Multilayer Perceptron is used to approximate any measurable function to any given degree of accuracy. It also acts as a basis of typical advanced approaches and is widely adopted in numerous domains. Implementing Multilayer Perceptron for feature representation is very elementary and remarkably efficient, although it might not be as demonstrative as autoencoders, CNNs, and RNNs.

However, one major limitation of deep feed-forward neural networks is that they do not have memories or loops for remembering preceding computations [113].

One of the primary reasons for employing Deep Neural Networks is that they are synthesized so that multiple neural networks can be merged into one big differentiable function and trained end-to-end. This application becomes a necessity because of the abundant availability of multimodal data. The DNN framework for recommender systems typically extracts user and item feature vectors or latent and explicit features. Another significant advantage of Deep Neural Network is that it can effectively extract essential features from raw data automatically.

The effect of upper-case letters in hashtags used for hashtag recommendations can be explored, and its effect on the recommendations can be analyzed in detail. Another scope in the future is to analyze the effect of language modeling on prediction performance. Another scope is to remediate the cold-start problem for user-based Collaborative Filtering by learning the similarities between user-to-user and user-to-job. This can be implemented on a multimodal document embedding.

4.3 Multi-view Deep Neural Network (MV-DNN)

Multi-view deep neural network (MV-DNN) is excellent in modeling domain recommendations [107]. It considers users as the main view and every other domain, say Z, as a secondary view. For every secondary or auxiliary user-domain pair, there exists a specific similarity score. The loss function of MV-DNN can be computed using Eq. 9. The architecture of MV-DNN is represented in Fig. 9.

$$\mathcal{L} = \text{argmin}_{\theta} \sum_{j=1}^{Z} \frac{\exp\left(\gamma \cdot \cos\left(Y_u, Y_{a,j}\right)\right)}{\sum_{X' \varepsilon R^{da}} \exp(\gamma \cdot cos(Y_u, f_a(X')))}. \quad (9)$$

Here, θ represents the model parameters, γ represents the smoothing factor, Y_u represents the user's output view, a represents active view's index, and R_{da} represents view a's input domain.

Figure 10 showcases the structure of the Deep Structured Semantic Model (DSSM) [11]. The crude textual features are fed as input in the form of a high-dimensional vector. The DSSM forwards these inputs to two neural networks separately and maps them into semantic vectors into a joint semantic space.

In Eqs. 10, 11, and 12, x represents the input vector, y represents the output vector, W_i represents the ith weight matrix, and l_i represents the hidden layers, such that $\forall i \in \{1, \ldots.., N-1\}$. q represents the query and d represents the document.

$$l_1 = W_1 x \quad (10)$$

$$l_i = f(W_i l_{i-1} + b_i), i \in \{2, \ldots\ldots, N-1\} \quad (11)$$

$$y = f(W_N l_{N-1} + b_N). \quad (12)$$

The activation function is given by Eq. 13:

$$f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}. \quad (13)$$

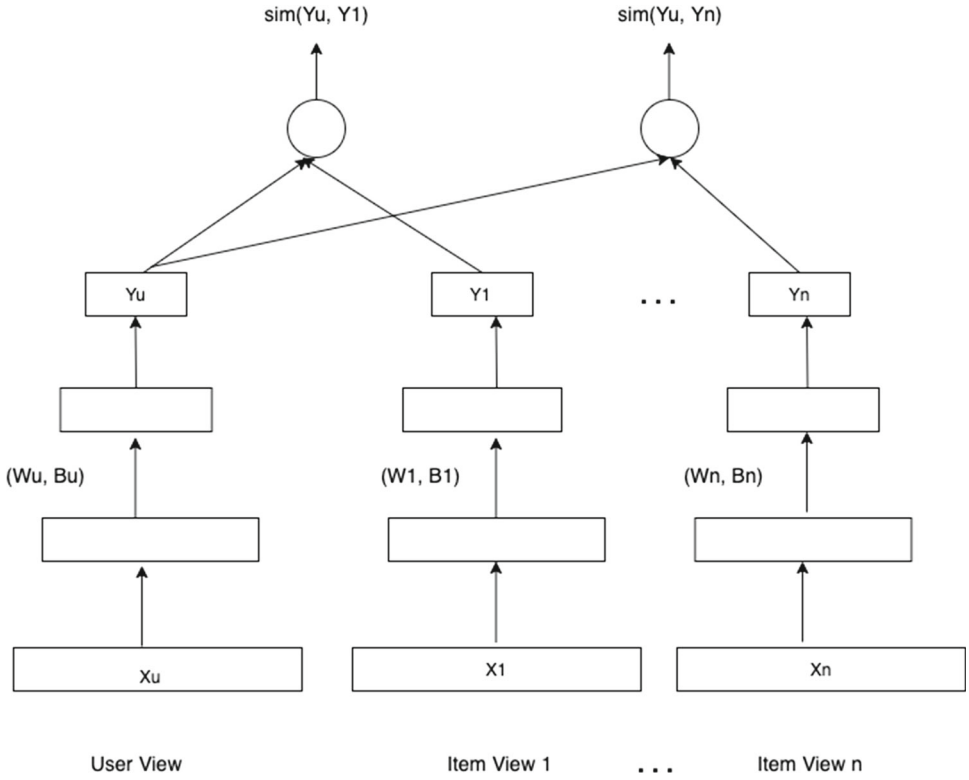


Fig. 9 Architecture of MV-DNN

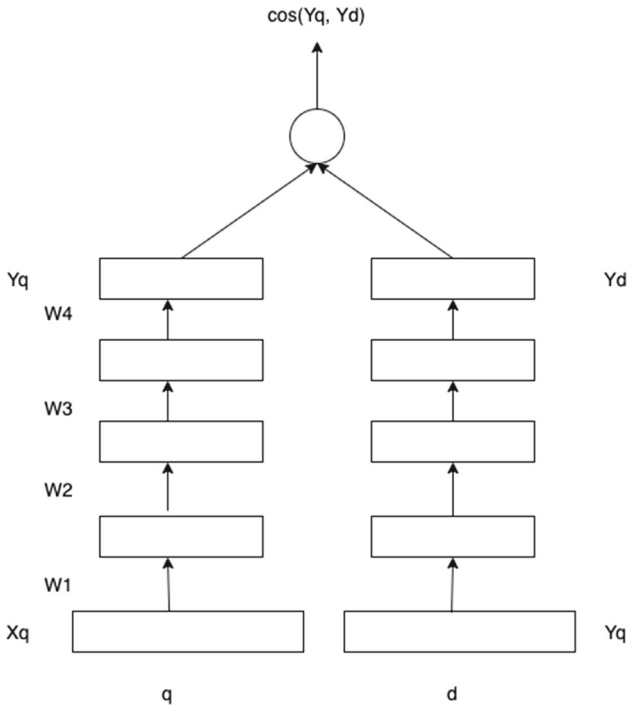


Fig. 10 Architecture of DSSM

The semantic relevance score between a query Q and a document D is given by Eq. 14:

$$R(q,d) = \cos\left(Y_q, Y_d\right) = \frac{Y_q^T Y_d}{\|Y_q\| \cdot \|Y_d\|}. \tag{14}$$

MV-DNN can be implemented in several domains. Conforming to the concepts of user-based collaborative filtering, as discussed in Sect. 2.1.2, users having similar preferences in one domain tend to have similar preferences in other domains as well. However, in many cases, this assumption may be rendered ineffective. Hence, the basic knowledge of the correlations between different domains is an essential aspect of MV-DNN. It is based on a Deep Structure-based Semantic Model (DSSM). In MV-DNN, the architecture of DSSM contains multiple views to map sparse features in high dimensions into the low-dimensional dense matrix. The authors observed that the online services present humungous content to the users, making this content user-centric [11]. Hence, they came up with the Deep Structured Semantic Model (DSSM), which used content-based filtering to improve the recommendations.

The authors on [30] observed that current CF-based techniques could only comprehend a single type of relations. RBM, for example, considers either user–user or item–item relations. The matrix factorization approach considers use-item relations only. To fix this problem, the authors propose a framework that first learns low-dimension user and item

vectors. This captures the user–user and item–item relations. This is then passed to a multi-view deep neural net, which models the user–item interaction.

4.4 Findings and Open Issues

Initially, Multi-view deep neural network (MV-DNN) was incepted for performing cross-domain recommendations.

However, the understanding of cross-domain recommendations may not be fruitful for MV-DNN. This is said essentially because the underlying literature of generating recommendations states that if a user likes an item a and there exists an item b similar to item a, then the user will like item b as well. However, this is not always true. For the times this hypothesis fails, this assumption becomes obstructive for the implementation of MV-DNN.

4.5 Convolution Neural Network (CNN)

For bio-inspired Multilayer Perceptrons and Computer Vision, CNN are the most used deep learning models. The essential elements of CNN are the convolutional layers and the subsampling layers. The convolutional layers work as a filter for the output from previous layers. These convolutional layers hence produce filtered outputs. The subsampling layers subsample the convolution output based on their activations [114]. To create a deep CNN model, the convolutional layers and the subsample layers are added alternatively. Hence, such a deep model of CNN can learn a hierarchy of complex features. Another massive advantage of CNN is that it has fewer parameters than the traditional feed-forward neural networks. This method is based on the feed-forward deep neural network. To reduce the preprocessing, multilayered perceptrons are used in this model. The architecture of CNN is shown in Fig. 11.

Cooperative neural network (CoNN) is a model where two neural networks work in tandem [15]. To discover novel user and item features, user and item reviews are given as input to the system. A common shared layer is added on the top of the two neural networks which couples them and the user and item features are mapped together into a common feature vector space to enable the interaction among them. Figure 12 represents the architecture of CoNN. Another latent layer is added to the architecture to enable the interaction of user and item latent features.

The two neural networks, i.e., neural network for items (NN_i) and neural network for users (NNu), run parallel. The item and user ratings are fed as inputs to the two neural networks, respectively. In the lookup layer, the user and item review texts are placed as matrices of word embeddings. The subsequent layers perform the functions of convolution, max pooling, and full connection, respectively. This layer also acts as a platform for computing the objective function for calculating the rating prediction error. One major drawback of Cooperative Neural Network is that it is incapable of addressing users and items which do not have ratings.

Many researchers addressed data sparsity, and they used CNN in their papers to solve this problem.

The authors in [15] observed that maximum recommender systems ignored the reviews leading to increased data sparsity problem. Thus, they proposed Deep Cooperative Neural Networks (DeepCoNN) model to learn item properties and user behaviors in the review text. Similarly, to solve the issue of the sparsity of data, authors integrated convolutional neural network (CNN) into probabilistic matrix factorization (PMF), resulting in convolutional matrix factorization (ConvMF) [48]. In the research work [92], the researchers solve the data sparsity problem by including tag and user information. First in nearest neighbors set, the most significant user impact is found using similarity metric, to process item information CNN is used. To get the results, the prediction matrix is decomposed by the probability matrix.

The profound use of CNN in the fashion industry was acknowledged in some research papers. The authors in [77] observed that not much work had been done on complicated recommendation scenarios involving knowledge transfer across multiple domains in fashion recommendations. Similarly, for online clothing shopping, current methods did not address the challenges in the cross-domain clothing retrieval scenario completely. The intra-domain and cross-domain data relations were considered together, and the numbers of matched and unmatched cross-domain pairs were imbalanced. Hence, the authors [78] proposed a deep cross-triplet embedding algorithm and a cross-triplet sampling technique to provide improved recommendations.

Some other applications of CNN have been included in the survey. For efficient venue recommendation, the authors proposed a City Melange framework that matched the interacting user to social media platforms with similar interests [80]. As a result, this approach could recommend both on- and off-the-beaten-track locations to the users. In another work, the authors proposed a hybrid music recommender system that used CNN to model real-world users' information and high-level rendering of audio data [102]. The authors observed in [52] that finding appropriate Ukiyo (a Japanese firm) e-prints that intrigue a novice is challenging. Thus, they proposed Ukiyo-e recommendation using a deep learning-based CNN model to present recommendations. The input to the recommender system is an image provided by the user. The CNN model is used to create a classifier that takes input images and other information related to the image and outputs whether the user will like or dislike it. The authors in [93] used tag information and user information and improved the recommendation by obtaining the nearest neighbor set that significantly impacted on the target user. They named their model convolution deep learning model on Label Weight

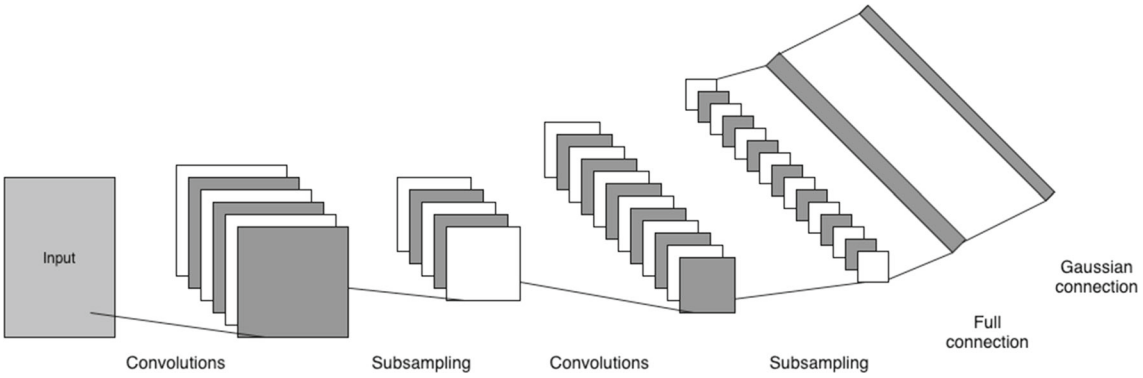


Fig. 11 Architecture of CNN

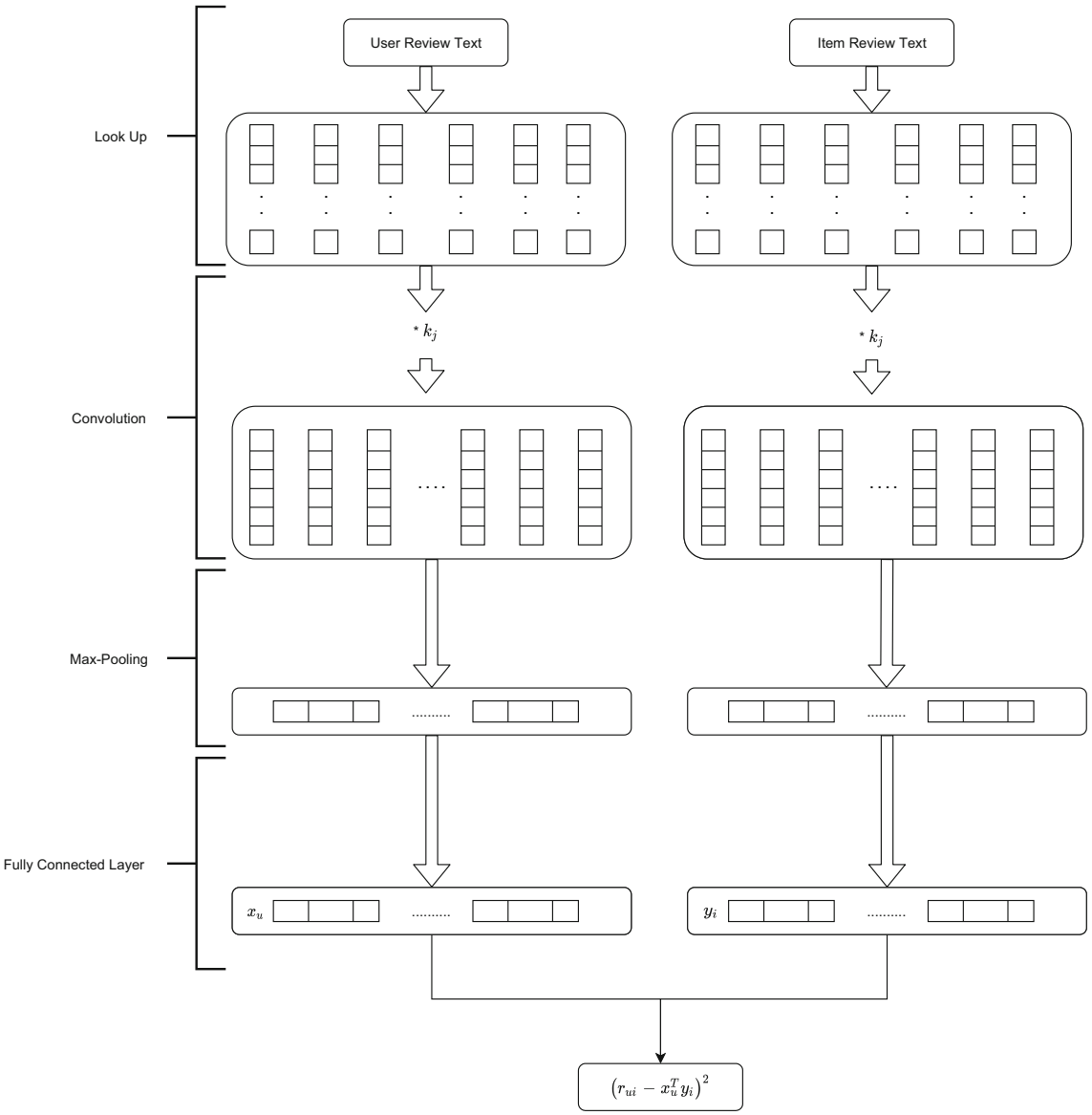


Fig. 12 Architecture of CoNN

Nearest neighbors (LWNCDL). Traditionally, candidate articles were handpicked from a vast pool of articles, and hence, its recommendation was manual and laborious. To solve this problem, the Dynamic Attention Deep Model (DADM) was proposed, which used multiple neural networks [50]. As a result, the authors used the convolutional neural network to study the connection between visual and textual inputs and examine the possibility of knowledge transfer from complex domains for expert recommendations.

The researchers in [76] addressed the problem of job recommendation by analyzing semi-structured resumes of candidates. To recommend a job, they used CNN to process the entire input at once, which is important since any part of the text in a resume can affect its semantics. The authors in [115] proposed an AODR model that considers user and item rating texts to infer item aspects and user opinions by applying deep learning. Such item aspects and user opinions were later merged using collaborative filtering to learn insightful information. In paper [106], the researchers introduced an algorithm, Aspect-Based Opinion Mining (ABOM), which analyzed the user reviews extensively to learn hidden features that further improve the recommendation accuracy. To carry out this task, they used multichannel deep CNN (MCNN) and Tensor Factorization (TF) machines. In paper [116], the authors used decentralized knowledge graphs in deep recommender systems to validate the efficiency of the system. The knowledge graph was constructed by crowdsourcing. The authors in [117] created a passenger hunting recommender system. The proposed system consists of two components, i.e., offline training and online implementation. In the former component, the authors applied deep CNN, and in the latter component, the authors proposed the DL-PHRec method. Hence, the authors could generate a personalized ranking list of destinations regions for each taxi driver.

4.6 Findings and Open Issues

CNNs are efficient in handling unstructured multimedia data with convolution and pooling functions. The majority of CNN based recommender models use CNNs for performing feature extraction. CNN based models are instrumental in learning deep features to model user and item latent factors. The significant advantage of using CNN is its implementation of pooling operation for reducing training data dimensions. CNN has been used for feature extraction to efficiently model user and item representations for Recommender Systems. It is also useful in image processing.

Another new scope is to embed the user's rating information in a matrix. A Convolution Neural Network can be trained on this matrix, treating it like an image where each data point represents a particular feature of the user. Exploiting Convolution Neural Networks for prediction of risk with medical imaging can be explored.

One major limitation is that it requires massive hyperparameter tuning to extract optimal features. In addition to this, it is also challenging to support intricate activity details. Although CNN uses feed-forwarding, it has fewer parameters than traditional deep feed-forward networks [114].

4.7 Autoencoders (AE)

Autoencoders are typically unsupervised neural networks that are trained to mirror its input as output. It is a tri-layered neural network, consisting of the input layer, hidden layer, and the output layer. The input layer is fed with the complex representations of the dataset, and in the hidden layer, such complex representations are transformed into low-dimensional representations. It essentially mirrors the working of an encoder, which encodes the complex, high-dimensional representations into low-dimensional representations. Mirroring the operations in a reverse manner, the low-dimensional representations are converted into high-dimensional representations as the data travels from hidden layer to output layer. This can also be termed as the working of a decoder. The architecture of denoising autoencoder is given in Fig. 13.

The feature extraction performed as the encoding is not robust enough. It was observed that the addition of Gaussian noise improved the problem described above [118]. Hence, to make the system more robust and training the hidden layers to identify hidden data features, denoising autoencoders were introduced. The structure of such autoencoders encodes the input while preserving the information about the input, and reduces the effect of the alteration process, applied stochastically to the input of the autoencoder. The authors in this work attempted to improve the automatic recommendations [26]. This was done by extracting the features of input data and reconstituting the input to perform recommendations. Authors in [119] extended generative models for the task of Collaborative filtering. To do this, they used Wasserstein autoencoders.

4.7.1 Stacked Denoising Autoencoders (SDAE)

A stacked autoencoder is a neural network having several layers of sparse autoencoders wherein the output of each layer is forwarded as the input of the successive layer. The hidden layers in a denoising autoencoders can be colluded to create a deep network by forwarding the output of the previous layer as the input to the next layer. Such a strong feature extraction capability makes stacked denoising autoencoder aptly suited for recommender systems. The architecture of a stacked denoising autoencoder is given in Fig. 14.

The method of generating top-N recommendations using *Stacked denoising Autoencoder* starts with selecting a dataset with user reviews [23]. Then, the similarity between the items

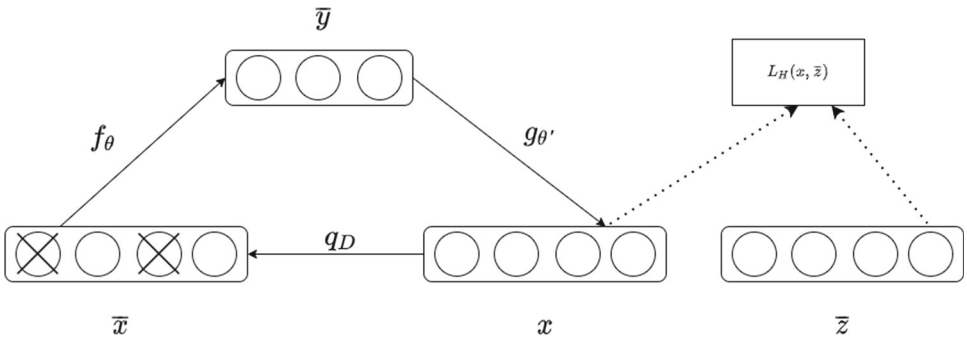


Fig. 13 Architecture of denoising autoencoder

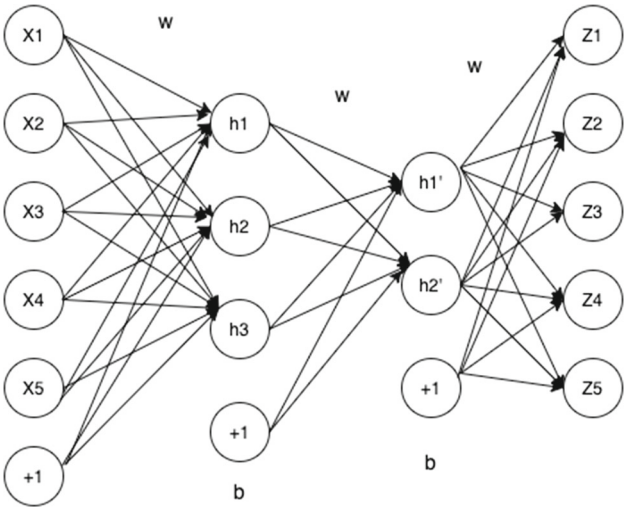


Fig. 14 Architecture of stacked denoising autoencoder

liked by the user is calculated for each user. Afterward, the nearest data points with a defined length, say K, are selected. Such K-nearest datasets are combined, and the user reviewed movies are excluded from such datasets to form another dataset. The similarity between the two datasets is calculated, and the top-N similar movies among the two datasets are selected and recommended to the user.

Upon stacking the denoising autoencoder, the deep model hence created increases the capability of feature extraction. To make this task robust, Gaussian noise is added to the system.

The authors proposed a Hybrid Recommendation system with CF and Deep learning (HRCD) [101]. It explored the content features of the items learned from a deep learning neural network and applied later to the timeSVD ++ CF model. The authors in [23] addressed the problem of information overloading and data sparsity. As a solution, stacked autoencoders were used with denoising to extract low-dimensional features from the sparse user–item matrix. It was observed that personalized recommendations often led to sparse observations of users' adoption of items. As a solution to the problem, the authors used Collaborative Topic Regression with Denoising Autoencoder or (CTR-DAE) [34]. Here, the user's community-based preference was bridged with his topic-based preference.

Similarly, some other researchers addressed data sparsity by proposing a hybrid model that executed deep learning of users and items' latent factors from side information and CF obtained from the rating matrix [97]. According to some authors, CF recommender systems suffer from Complete Cold-Start (CCS) problem where zero rating records are present and Incomplete Cold-Start (ICS) problems where significantly less rating records are available for new items or users in the system [35]. Hence, they used timeSVD++ models and used temporal dynamics of user preferences and item features for improvised recommendations.

4.7.2 Variational Autoencoders (VAE)

Due to the drawbacks of collaborative based filtering, the authors in [103] used variational autoencoders to include rating and content information for the recommendation. The novelty in the framework is that the latent distribution for content is learned in latent space rather than observational space. In another paper, the authors demonstrated the enhancement of modeling in CF with side information by using Variational Autoencoders-based Collaborative Filtering (VAE-CF) [92]. The authors [120] proposed a novel healthcare recommender system collaborative variational deep learning (CVDL) to eliminate the limitations of data sparsity and cold-start problem from recommender systems to generate useful recommendations. The paper [121] proposes a deep autoencoder model to learn low and high-dimensional features to remediate data sparsity and data imbalance problems in recommender systems.

To make the task of feature selection simpler and more efficient, this paper [122] proposed a fuzzy entropy-based deep learning recommender system to decrease data dimensionality and eliminating unnecessary noise from data. This paper [123] proposed an end-to-end recommender model by taking several content sources, for example, textual content, graphical content, and others. The authors used a stacked autoencoder to carry out this implementation. In this paper [124], the authors identified the limitations of side infor-

mation in RS. Apart from having extensive detail about the rating, side information also contains a tremendous amount of noise. Hence, the process of feature extraction from such side information becomes challenging. So, the authors proposed a Stacked Discriminative Denoising Autoencoder by merging deep learning with matrix factorization-based RS.

4.8 Findings and Open Issues

Autoencoder is an efficient feature representation learning method that learns feature representations from user–item content features. Autoencoder can be used in recommender systems by learning low-dimensional feature representations at the outer layer or by building the interaction matrix directly in the reconstruction layer. Autoencoder models have a high capability to work with noisy data to learn the complicated and hierarchical structure from the input data. Sparse autoencoders are very efficient for low-dimensional feature extraction from input data using a supervised learning technique. Denoising autoencoders are trained by initializing layers wherein each layer produces input data for the next layer. One of its significant advantages lies in implementing recommender systems as denoising autoencoders can be stacked to decrease the processing errors.

Autoencoders suffer from certain limitations. Autoencoder based Collaborative Filtering (ACF) is incapable of incorporating non-integer ratings, and partially observed features result in low recommendation accuracy. Another major disadvantage of using deep autoencoders is that they are incapable of searching for an optimal solution, and due to high parameter tuning, the time complexity of the training process increases manifold.

4.9 Neural Matrix Factorization (NMF)

Matrix factorization (MF) is one of the most widely and commonly used Collaborative Filtering approaches. Such MF models learn the low-dimensional embeddings of items and users in common latent factor space. The system generates a user–item matrix where the ratings provided for every item by the users are captured. However, due to the problem of data sparsity, the user–item matrix is sparse. Such a sparse matrix is factorized into a dense matrix to make computations simple. This is done by taking the product of two low-rank matrices consisting of user–item embeddings. Upon learning the latent factors, the similarity between user and item is computed. In addition to this, the newly discovered preferences are considered by analyzing the user and item latent factor representations

To learn such latent user–item factors, loss function has to be calculated as given in Eq. 15:

$$min_{v_u v_i} \sum_{(u,i)\in S} (r_{ui} - v_u, v_i)^2 + \lambda\Big(\|v_u\|^2 + \|v_i\|^2\Big). \tag{15}$$

Here, v_i represents the latent factor of item i, v_u represents the latent factor of user u, $\mathrm{r_{ui}}$ represents the rating given to item i by user u, and λ represents the item bias to prevent overfitting.

Factorization Machines (FM) are supervised learning models that are used in various prediction problems like regression, classification, and others. These models map random real features into low-dimensional latent feature vector space. They can determine the parameters of the model precisely given a sparse user–item matrix and train with linear complexity. Such characteristic makes FMs ideal for implementing real-world recommender systems.

4.9.1 Deep Factorization Machines (DeepFM)

DeepFM is a model wherein the capabilities of Factorization Machines are combined with deep feature learning to form a novel neural architecture, wherein Factorization Machines and Deep Neural Network (DNN) are integrated to model both low-dimensional and high-dimensional feature vectors.

As the name suggests deepFM is a dual component model, i.e., the *factorization machine* component and *deep learning* component. These components share the same input. These two components are jointly trained for the prediction model computed using Eq. 16.

$$\hat{y} = \mathrm{sigmoid}(y_{\mathrm{FM}} + Y_{\mathrm{DNN}}). \tag{16}$$

Here, $\hat{y}\varepsilon(0, 1)$, y_{FM} represents the output of the FM component, and Y_{DNN} represents the output of the deep learning component.

The main advantages of using this model are no requirement for pre-training the model, its capability to learn both low-dimensional and high-dimensional features. The shared feature embedding of FM and deep components enables the system to avoid extensive feature engineering.

Rather than conventional push–pull methodology among the user and item pairs, the authors proposed LRML (Latent Relational Metric Learning) model to learn latent relations that described every user–item interaction [96]. The authors observed in [114] that although *Weighted Matrix Factorization (*WMF*)* can learn latent features from implicit feedback, these features are still not good enough to train a CNN model. The authors in [95] proposed a novel matrix factorization method with neural network architecture and novel use of binary cross-entropy loss function.

In another work, the authors identified that Factorization Machines (FM) are incapable of modeling the nonlinear aspect of the real-world data [104]. In [54], the authors observed that the recommendation systems primarily relied on initializing the user and item latent feature vectors. In this way, they used deep learning to estimate the initialization in MF for trust-aware social recommendations and to differentiate the neighborhood effect in the user's trusted social circle. This was achieved with the help of the deep learning-based matrix factorization (DLMF) model.

4.10 Findings and Open Issues

Neural matrix factorization provides exemplary results for latent feature models; however, it is influenced by methods involving local graph structure.

One limitation of neural matrix factorization is that to investigate the scope of system architectures, activation functions, regularization techniques, and cross-validation strategies. There is a risk of overfitting, which may lead to erroneous or insignificant insights.

4.11 Neural Collaborative Filtering (NCF)

In recommender systems, the users and items collate together to create a mapping that helps to under the similarity between the two and further recommend items to the user. Since the phenomenon of generating recommendations in a dual process, which involves both user latent factors and item latent factors. Hence, to model this two-way interaction between user and item, Neural Collaborative Filtering (NCF) is used. The scoring function of NCF can be computed using Eq. 17:

$$\hat{r}_{ui} = f\left(U^T \cdot s_u^{\text{user}}, V^T \cdot s_i^{\text{item}} | U, V, \theta\right). \tag{17}$$

Here, $f(\cdot)$ represents the multilayer perceptron, s_u^{user} represents the side information of the user profile, s_i^{item} represents the side information of the item features, and θ represents the parameters of the network.

It becomes easier to combine the neural representation of MF with Multilayer Perceptron to create a model that incorporates the linear characteristics of matrix factorization and nonlinear characteristic of MLP. This results in improved performance of the recommender engine. The cross-entropy loss for implicit and explicit feedback is given by Eq. 18:

$$\mathcal{L} = \; - \sum r_{ui} \log \hat{r}_{ui} + (1 - r_{ui}) \log\left(1 - \hat{r}_{ui}\right). \tag{18}$$

Here, u represents the user, i represents the item, and r_{ui} represents the rating of item i given by user u.

4.11.1 Collaborative Deep Learning (CDL)

This technique is an amalgamation of deep representation of content and collaborative filtering for the ratings. CDL is a deep learning recommender model that learns from the text of the reviews given by the users [114]. It combines the autoencoder and *Click Through Rate* (CTR) to model ratings.

To learn from such review texts, CDL implements *Bag-Of-Words* (BOW) technique. This model has certain limitations, such as it models only item review texts, limiting the capabilities of a recommender system, and due to its usage of the BOW method, it only considers the frequency of words in a text, and not the similarities between different texts. Another limitation of the CDL model is that it considers the order of words in a text which sometimes contradicts the semantics of the text. Figure 15 represents the architecture of SDAE, upon which the SDAE is built. Figure 16 represents the architecture of CDL.

The authors argued that the sole criteria for recommending items were user reviews [102]. However, this often led to data sparsity, hence degrading the recommendation quality. To avoid this, the authors came up with a collaborative deep learning (CDL) model, which simultaneously performed deep learning of content and collaborative filtering for the ratings. In another work, the authors used CDL-based Bayesian deep learning for Recommender Systems model to achieve integrated intelligence that involved both perception and inference in a principled probabilistic framework [89]. In [70], the researchers attempted to improve YouTube recommendations by using a deep collaborative filtering model that efficiently integrated various signals and added layers of depth to the interaction. Some authors in their paper [83] compared the Deep Neural Network (DNN) features with visual features for artwork recommendation and concluded that DNN features outperformed the other.

The authors in [90] presented a novel neural network model, Neighborhood-based Neural Collaborative Filtering (NNCF), which jointly characterized both user–item interactions and neighborhood information for the recommendation.

The authors in [125] indicated that the Collaborative Filtering algorithms primarily deal with low-dimensional and linear interactions between users and items. To address this issue, the authors proposed a novel deep CF recommender algorithm for service recommendations. The authors in [126] used MLP to learn the interaction function to implement the recommender system. The authors [127] proposed a deep multi-criteria collaborative filtering RS to generate accurate multi-criteria recommendations.

4.12 Findings and Open Issues

Collaborative deep learning (CDL) can perform deep learning of content and collaborative filtering for the ratings of

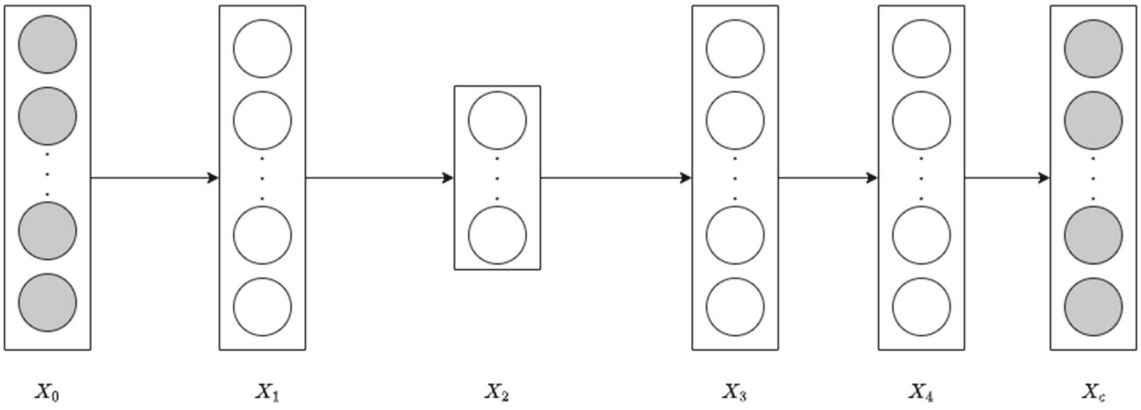


Fig. 15 Architecture of SDAE

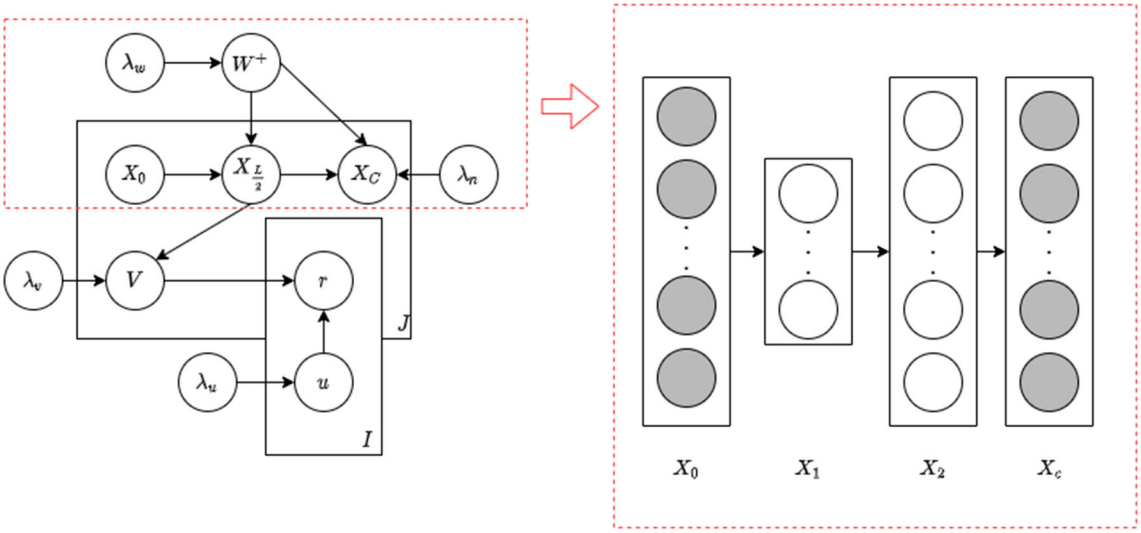


Fig. 16 Architecture of CDL

items. This allows the model to balance the effects of side information and interaction history.

One significant limitation of the CDL model is that it cannot apprehend contextual information with word embeddings and convolutional filters.

4.13 Deep Belief Network (DBN)

It is a class of deep learning where several layers of hidden units are stacked together to form a deep net. The hidden units present in each layer are not directly connected. They can be viewed as a class of unsupervised networks like *Restricted Boltzmann Machines (RBM)* [128]. Authors [129] observed that classic neural networks suffer from the problem of optimization. DBN was introduced to solve this problem [130]. This technique integrates supervised and unsupervised learning by implementing a local search to get an optimized result and learn the data distribution without prior knowledge [129]. DBN employs a stacked RBM structure for extracting deep features of data [69]. Thus, it can be viewed as a generative probabilistic model [36]. In DBN, the hidden layers are trained in a bottom-up manner. This pre-training of the layers is termed as the *pre*-train stage, and upon the addition of the objective layer, the training of the model becomes supervised [131]. Figure 17 represents the architecture of the Deep Belief Network.

The standard RBM is given by Eq. 19

$$E(\vec{u}, v) = -\sum_{i=1}^{p}\sum_{j=1}^{q} w_{i,j}u_i v_j - \sum_{i=1}^{p} x_i u_i - \sum_{j=1}^{q} y v_j, \quad (19)$$

where $\vec{u}$ is the visible unit vector and $\vec{v}$ is the hidden unit vector. w represents the weight matrix and $\vec{x}$ is the visible bias vector and $\vec{y}$ is the hidden bias vector.

In video recommender systems, generating recommendations for videos becomes complicated because extracting intricate features from graphics is difficult. The workflow of such a recommender system is presented in Fig. 18. To make

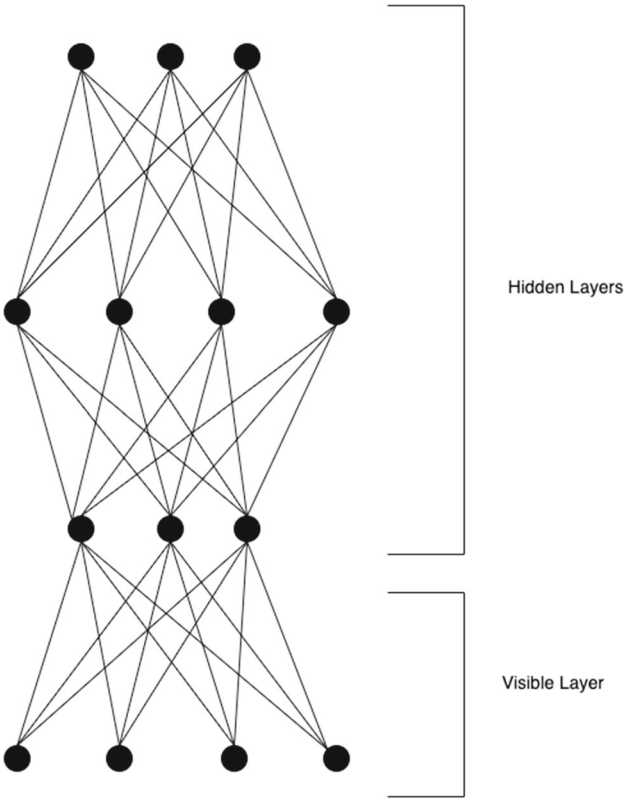


Fig. 17 Architecture of deep belief network (DBN)

this task simpler, authors in [69] divided the user–item matrix into 0–1 matrices and feed them as input to the DBN model. Hence, post-training, the difference between trained features and raw features becomes prominent. Afterward, UBCF calculates the similarity between neighboring users to generate recommendations.

4.14 Findings and Open Issues

One of the most significant advantages of DBN is that Deep Belief Network has been successfully employed to perform extensive feature engineering. It is useful in extracting hidden and useful features from audio data [106]. Another significant advantage of using DBN is that it is also used for dimensionality reduction. DBN has also proved to help solve the data sparsity problem, which leads to the cold-start problem and generation of poor-quality recommendations.

However, DBN comes with several limitations as well. Due to extensive parameter initialization, the DBN model is challenging to train. Another major disadvantage of DBN is that it cannot learn the representation of features from labeled and unlabeled data[132].

The training of the Deep Belief Network model can be further optimized to improve the recommendations.

4.15 Recurrent Neural Network (RNN)

Unlike feed-forward deep neural network, the layers present in a recurrent neural network form a graph, and hence the layers interact with each other. The graphs thus formed can either be cyclic or acyclic [107]. Essentially RNN captures the sequential data temporally, which is used for training the model [133]. The RNN takes the input and computes the weighted sums as given in Eq. 20. The architecture of the recurrent neural network is presented in Fig. 19.

$$z_{hj} = \sum_{i=1}^{n_x} x_i w_{ij}^{xh} + \sum_{i=1}^{n_h} h_i w_{ij}^{hh}, \tag{20}$$

where n_x are the input nodes and n_h are the hidden nodes.

The weights between input and hidden nodes are given by w^{xh}, and w^{hh} gives the weights between the hidden nodes. Using Eq. 21, the activation values can be calculated.

$$h_j = \tanh\left(z_{hj}\right) = \tanh\left(\sum_{i=1}^{n_x} x_i w_{ij}^{xh} + \sum_{i=1}^{n_h} h_i w_{ij}^{hh}\right). \tag{21}$$

The weighted sums of the output layer can be calculated as given in Eq. 22

$$z_{yj} = \sum_{i=1}^{n_h} h_i w_{ij}^{hy}. \tag{22}$$

4.15.1 Gated Recurrent Unit for Recommender System (GRU4REC)

GRU4REC is a session-based recommender system proposed by [134]. The input given to the model is an encoding of 1 of N, where N represents the number of items. The output of the system determines the possibility of an item to move to the subsequent session. If the item is active in the current session, the coordinate is assigned value 1, otherwise 0. The architecture of GRU4REC is given in Fig. 20.

The ranking loss function for this model is computed using Eq. 23:

$$\mathcal{L}_s = \frac{1}{S} \sum_{j=1}^{S} \sigma\left(\hat{r}_{sj} - \hat{r}_{si}\right) + \sigma\left(\hat{r}_{sj}^2\right). \tag{23}$$

Here, S represents the sample size, $\hat{r}_{si}$ represents the scores on negative item i at session s, $\hat{r}_{sj}$ represents the scores on positive item j at session s, and σ represents the logistic sigmoid function.

In GRU-based recurrent neural networks (RNN), the input given to the system depicts the current state of the system, and the output of the system depicts the subsequent event of the system. Unlike in GRU4REC, where 1-of-N encoding is

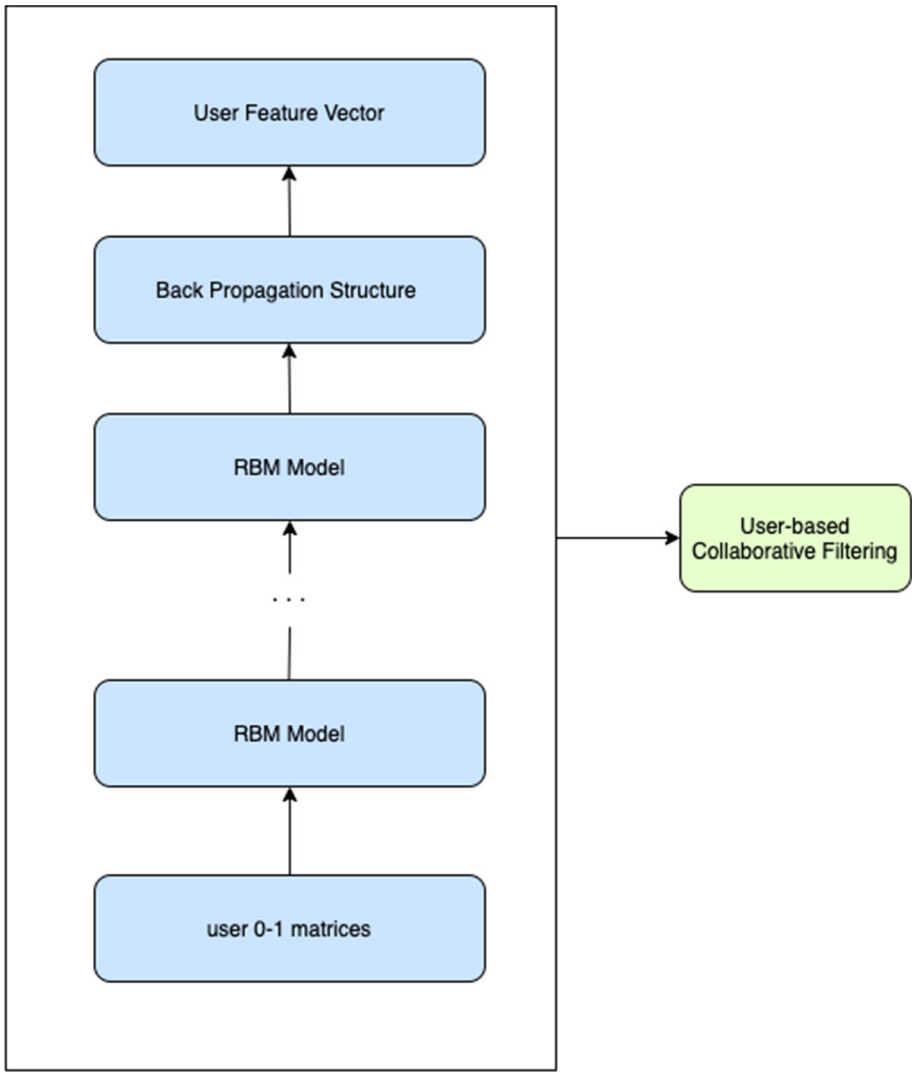


Fig. 18 Video recommender system implementing DBN and CF

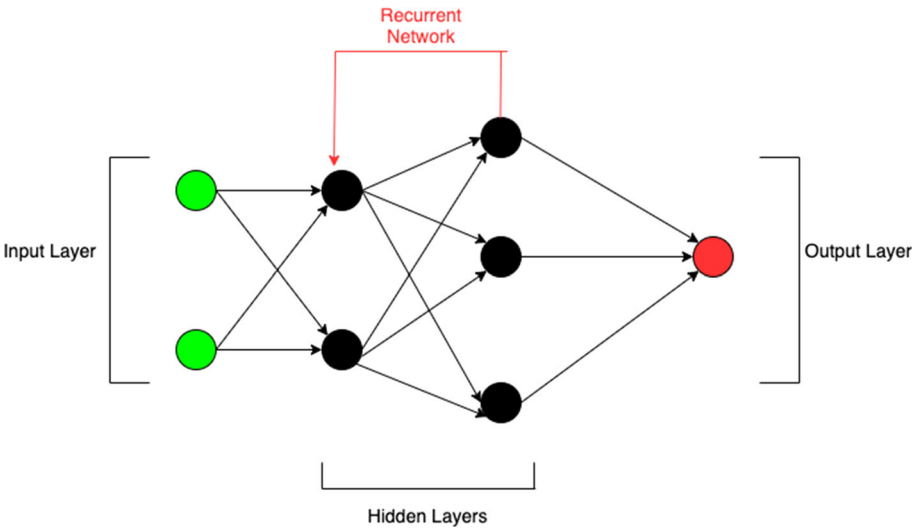


Fig. 19 Architecture of recurrent neural network

used, in GRU based RNN for session-based recommender systems, a weighted sum of the representations is used. In such cases, the events which occurred in the past are given as discounted when fed as an input. Hence, to make the system stable, the input vector is normalized.

When multiple feed-forward layers are added to the system, the output predicts whether the item will be included in the next session or not. The addition of multiple GRU layers uses the output of the previous layer as the input to the next layer. This deep connection of layers improves the performance of the system.

In some research studies, the side information was included to make the recommendations efficient. It was observed in [10] that recommender systems based on collaborative filtering could be enhanced by including side information like natural language reviews. To carry out this task, the authors used the bag-of-words technique along with RNN. In [67], according to the authors, current recommendation techniques using RNN only took into account the user's past activities and did not consider the essential side information. To cater to this problem, they used contextual recurrent neural networks (CRNN) to include the side information to give recommendations. Some authors considered the temporal while making the recommendations and studied its effects. To model the temporal behavior in recommendation systems, a group of researchers [21] introduced the Temporal Deep Semantic Structured Model (TDSSM) based on RNN, where they used long-term static and short-term temporal user choices to improvise the performance of the recommender system.

Similarly, researchers in [62] observed that the conventional POI recommender systems did not consider the temporal aspect while generating the recommendation. Hence, they came up with the idea to capture sequential check-in data of users and used RNN based deep neural network to provide recommendations. Catering to the same issue, the authors in [85] observed that a notable problem with recommender systems is that they do not encompass the context of time. So, they came up with the idea of using RNN and including the temporal shift while performing recommendations.

In a few research papers, the RNN technique was applied for session recommendations. In [37], the authors used the recurrent neural network for session-based recommendations. They incorporated two procedures for improving the model. The first one was data augmentation, and the second was to observe shifts in the input data distribution. Similarly, it was observed in [38] that in the real-life recommender systems, the session-based recommendations were based solely on the clicks of a user session. To solve it, the authors introduced several parallel RNN (p-RNN) frameworks to model sessions depending on the clicks and the features of the clicked items. The researchers in [41] proposed a model for session-based recommendations, where the recommendations were made at the starting of the sessions, which avoided the cold-start problem. In [72], the authors used Long Short-Term Memory (LSTM), a neural network-based RNN technique for quotes recommendation. Similarly, to make recommendations using the data content describing the items, the authors proposed Ask Me Any Rating (AMAR), which used Long Short-Term Memory (LSTM) networks to simultaneously learn two embeddings representing the items to be recommended and user preferences [65]. The concept of LSTM was also used in [41], where the authors proposed Recurrent Recommender Networks (RRN) to incorporate dynamic recommendations using Long Short-Term Memory (LSTM) autoregressive model.

In some papers, the problem of data sparsity was addressed and rectified using RNN. The authors in [40] introduced a ReLaVaR model that addressed data sparsity by considering the network recurrent units as stochastic latent variables with a former distribution performed over them. Similarly, it was observed that personalizing Adaptive User-Interfaces (AUIs) became difficult with several interaction modules as the item-user matrix is sparse. The authors in [84] used architecture consisting of RNN that performed sequential recommendations of content and control elements to solve this problem. Several other independent research works were performed using RNN. The authors in [46] analyzed that in documents, the semantic similarities between texts were not considered for making recommendations. They were able to effectively model the content of the abstracts of the documents using linear regression based on RNN. In another work, it was observed that the evolution and drifting of features might take place over time as users interact with different items. Hence, the authors utilized the recurrent neural network to learn a representation of impact from drift, evolution, and co-evolution of user and item attributes to make recommendations [82]. To improve the performance of multi-task learning, the authors used deep RNN to encode the text sequence into a latent vector trained on the collaborative filtering procedure [49].

Performing recommendations using sentiment analysis of short texts like sentences did not yield correct results due to less contextual information. Hence, the authors used the deep learning technique recurrent neural networks to solve this problem [99]. To find relevant citations or related work, the authors came up with Neural Citation Network (NCN) model to provide a curated list of high-quality candidates from a short passage of text [63]. For generating sequential recommendations, the authors in [13] came up with user-based RNN, which allowed generating personalized suggestions. In another work, to find relevant research articles, researchers usually rely only on the keyword-based search or by following citations. In the paper, the authors analyzed user activities to provide recommendations [65]. For song recommendations, it was observed by the authors that the combination

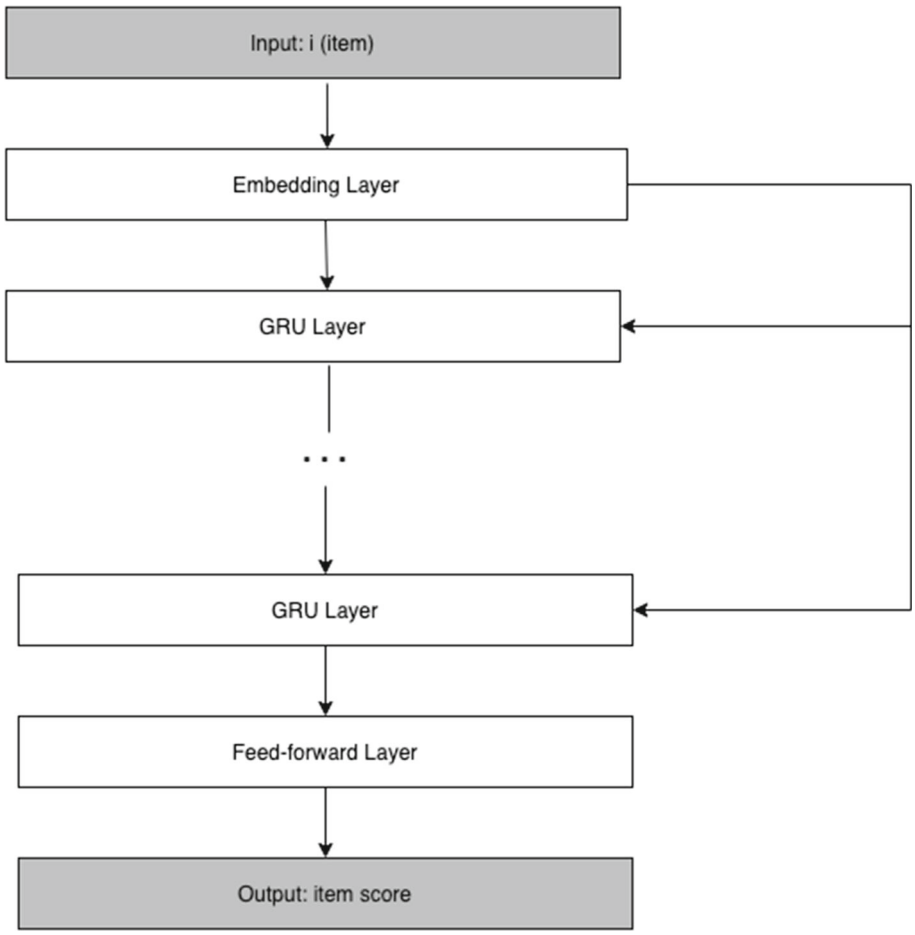


Fig. 20 Architecture of GRU4REC

factor of lyrics and genre was not included for the recommendation [45]. Hence, they used RNN to predict the user's next song of interest-based on the similarity factor. Some authors analyzed that the recommendation quality of rating predictions in traditional systems had scope for improvement [75]. Hence, they proposed a Neural Rating and Tips generation (NRT) framework, which could predict accurate ratings and create conceptual tips and strong linguistic quality. In another work, the authors addressed the limitation of efficient methods in the e-learning area [74].

To improve the recommendations, the authors used Gated Recurrent Unit (GRU) neural networks, a type of RNN. In another work, the authors deployed TARMF (Topical Attention Regularized Matrix Factorization) model that co-learned user and item details from ratings and customer reviews by optimizing matrix factorization and an attention-based GRU network [19].

To fully exploit and understand the user sentiments present in the review texts, the authors in this [135] paper implemented a deep learning-based neural network model, SDRA. Gated Recurrent Unit (GRU) is a type of Long Short-Term Memory (LSTM) model, which is an extension of an RNN model. GRU is a much simpler version of the LSTM model; however, the parameter setting in GRU for better accuracy is much easier as compared to LSTM [131].

4.16 Findings and Open Issues

Recurrent neural network is efficient for using sequential data. Also, for subsuming side information like time, logs, etc., RNNs are beneficial.

An important open issue is for music recommendation systems, the recommendations can be improved by performing temporal analysis of the features of music using recurrent neural network, and fascinating characteristics of music can be uncovered by interpreting automatically learned features. Another scope is to explore recurrent neural networks (RNNs) with bi interaction pooling to model sequential data. To further improve the performance of LSTM models, attention-based memory networks can be explored. The possibility of using a sequence-to-sequence (seq 2seq) learning framework can be explored, aiming to create a convincing recommendation description to aid customers in making a better purchase.

An important limitation of RNN that can also be considered as an open issue is that although applying the

neighborhood approach in RNN can lead to good results, but it is also suggested that some baselines in recent research studies are not well-justified and correctly evaluated.

Gated Recurrent Unit primarily addresses the problem of the vanishing gradient. GRUs often combine several memories and gates to record sequential activities. Another advantage of this model is that it can perform rating prediction and create general tips with linguistic quality. This model efficiently stores contextual information for rendering user–item latent features into a brief sentence.

4.17 Hybrid Networks

4.17.1 word2vec and Convolutional Neural Networks

The authors in [79] worked on a blog recommendation. They stated that due to a large number of blogs surfacing each day, it was crucial to recommend the right blogs to the right users. For this, they came up with the Boosted Inductive Matrix Completion method. Hence, they used the side information of users and blogs for adequate recommendations.

4.17.2 RNN and CNN

The authors worked toward improving quote recommendation. For this, first, they used RNN to model the tweet sequences, and then they used CNN for mapping tweets to intermediate vectors [71]. The authors in [60] observed that for POI recommendations, modeling multi-source information was one-dimensional. Hence, they proposed a Deep Context-aware POI Recommendation (DCPR) model, which consisted of different layers. One layer performed feature mining using convolutional neural network; the second layer was based upon recurrent neural network, and the third layer modeled these together. In another work [24], it was observed that the main challenge with news recommendation was to recommend the latest news articles to the users. To solve this issue, the authors presented an improved session-based recurrent neural network (RNN) model, which studied the users' history of reading news articles and thus made recommendations. The authors in [86] used the dual-regularized matrix factorization technique, which consisted of a multilayered neural network model by simultaneously implementing a convolutional neural network and gated recurrent neural network, to create an independently distributed rendering of contents of users and items.

4.17.3 CNN and Stacked Denoising Autoencoder

To address data sparsity, the authors used the Probabilistic model of the Hybrid Deep collaborative filtering model (PHD) and combined a stacked denoising autoencoder and a CNN alongside the auxiliary side information to extract users and items' latent factors [94].

4.18 Findings and Open Issues

Several deep learning-based recommendation methods employ more than one deep learning techniques. Deep neural networks give the system an edge over every other method by combining numerous neural computations and complementing each other to form a more efficient hybrid model. Each combination of a neural network technique is very specific and is tailored to each use case, as every problem statement requires a unique set of neural network operations.

An open issue is exploring all the possible combinations of deep learning techniques since many combinations have not been exploited yet and may provide useful insight into the increasing efficiency of recommender systems.

5 Conclusion and Future Work

In this paper, a literature survey has been performed on deep learning in recommender systems. It was observed that deep learning provides a considerable advantage in performance, especially when data is available in abundance. It was also observed that by using deep learning, we could extract feature representations that are much more comprehensive and better performing than the features extracted using traditional feature engineering. Deep learning can also incorporate side information like time using models that can incorporate temporal data like LSTM.

However, deep learning only works well when data is available in abundance. In situations where data sparsity exists, deep learning is not the best fit. Also, deep learning requires extensive hyperparameter tuning to achieve the desired accuracy. Quintessentially, hyperparameter tuning is a problem that plagues all machine learning algorithms, and deep learning models have a lot of additional hyperparameters making the problem even worse. Poor tuning can lead to overfitting or underfitting of data points. As future work, these limitations of deep learning can be worked upon using the deep learning models specified in Sect. 4.

Deep learning is a great tool to exploit the ever-explosive data available online and has shown to improve performance in all domains of recommender systems. However, creating and training deep learning models is a sensitive process, and proper hyperparameter tuning is required to get the best results out of it. Deep learning also lacks interpretability and often functions as a black box. Despite these limitations, deep learning has become an integral part of most state-of-the-art recommender systems and is increasingly becoming more relevant with the advent of big data.

References

1. Kim, K.; Ahn, H.: A recommender system using GA K-means clustering in an online shopping market. Expert Syst. Appl. **34**, 1200–1209 (2008). https://doi.org/10.1016/j.eswa.2006.12.025
2. Meteren, R. Van; Someren, M. Van: Using content-based filtering for recommendation. ECML/MLNET Work. Mach. Learn. New Inf. Age. (2000). https://doi.org/10.1125/5743
3. Vekariya, V.; Kulkarni, G.R.: Hybrid recommender systems: survey and experiments. In: 2012 2nd International Conference on Digital Information and Communications Technology its APPL. DICTAP 2012, pp. 469–473 (2012). https://doi.org/10.1109/DICTAP.2012.6215409
4. Zhang, X.; Liu, H.; Chen, X.; Zhong, J.; Wang, D.: A novel hybrid deep recommendation system to differentiate user's preference and item's attractiveness. Inf. Sci. (Ny) **519**, 306–316 (2020). https://doi.org/10.1016/j.ins.2020.01.044
5. Peterson J.J.; Yahyah M.; Lief K.; H.N.: Predictive distributions for constructing the ICH Q8 design space. In: Comprehensive Quality by Design for Pharmaceutical Product Development and Manufacture, pp. 55–70 (2017). https://doi.org/10.1145/1143844.1143865
6. Zhu, X.: Semi-supervised learning literature survey contents. Sci. York. **10**, 10 (2008)
7. Grant, P.: Assessment and selection. Bus. Giv. (2014). https://doi.org/10.1057/9780230355033.0018
8. Liu, Y.H.; Wang, X.K.; Liu, P.X.; Zheng, J.P.; Shu, C.Y.; Pan, G.S.; Luo, J.: Bin: Modification on the tribological properties of ceramics lubricated by water using fullerenol as a lubricating additive. Sci. China Technol. Sci. **55**, 2656–2661 (2012). https://doi.org/10.1007/s11431-012-4938-y
9. Zhang, Y.; Ai, Q.; Chen, X.; Croft, W.B.: Joint representation learning for top-N recommendation with heterogeneous information sources. In: International Conference on Information and Knowledge Management. Proceedings of Part F1318, 1449–1458 (2017). https://doi.org/10.1145/3132847.3132892
10. Almahairi, A.; Kastner, K.; Cho, K.; Courville, A.: Learning distributed representations from reviews for collaborative filtering. In: RecSys 2015 – Proceedings of 9th ACM Conference on Recommendation Systems, pp. 147–154 (2015). https://doi.org/10.1145/2792838.2800192
11. Elkahky, A.M.; Song, Y.; He, X.: A multi-view deep learning approach for cross domain user modeling in recommendation systems. In: Proceedings of 24th International Conference on World Wide Web - WWW'15, pp. 278–288 (2015). https://doi.org/10.1145/2736277.2741667
12. Seo, S.; Huang, J.; Yang, H.; Liu, Y.: Interpretable convolutional neural networks with dual local and global attention for review rating prediction. In: Proceedings of Elsevier ACM Conference on Recommendation Systems - RecSys'17, pp. 297–305 (2017). https://doi.org/10.1145/3109859.3109890
13. Donkers, T.; Loepp, B.; Ziegler, J.: Sequential user-based recurrent neural network recommendations. In: Proceedings of Elsevier ACM Conference on Recommendation Systems - RecSys'17, pp. 152–160 (2017). https://doi.org/10.1145/3109859.3109877
14. Purkaystha, B.; Datta, T.; Islam, M.S.; Marium-E-Jannat: Product recommendation: a deep learning factorization method using separate learners. In: 20th The International Conference on Information and Computer Technologies ICCIT 2017. 2018-Janua, pp. 1–5 (2018). https://doi.org/10.1109/ICCITECHN.2017.8281852
15. Zheng, L.; Noroozi, V.; Yu, P.S.: Joint Deep Modeling of Users and Items Using Reviews for Recommendation, pp. 425–433 (2017). https://doi.org/10.1145/3018661.3018665
16. Paradarami, T.K.; Bastian, N.D.; Wightman, J.L.: A hybrid recommender system using artificial neural networks. Expert Syst. Appl. **83**, 300–313 (2017). https://doi.org/10.1016/j.eswa.2017.04.046
17. Wang, X.; He, X.; Nie, L.; Chua, T.-S.: Item Silk Road: Recommending Items from Information Domains to Social Users, pp. 185–194 (2017). https://doi.org/10.1145/3077136.3080771
18. Zhu, H.; Li, X.; Zhang, P.; Li, G.; He, J.; Li, H.; Gai, K.: Learning Tree-based Deep Model for Recommender Systems. (2018). https://doi.org/10.1145/3219819.3219826
19. Lu, Y.; Dong, R.; Smyth, B.: Coevolutionary Recommendation Model: Mutual Learning between Ratings and Reviews. Www, pp. 773–782 (2018). https://doi.org/10.1145/3178876.3186158
20. Oh, K.J.; Lee, W.J.; Lim, C.G.; Choi, H.J.: Personalized news recommendation using classified keywords to capture user preference. In: The International Conference on Advance and Communcations Technologies ICACT, pp. 1283–1287 (2014). https://doi.org/10.1109/ICACT.2014.6779166
21. Song, Y.; Elkahky, A.M.; He, X.: Multi-rate deep learning for temporal recommendation. In: SIGIR 2016 - Proceedings of 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 909–912 (2016). https://doi.org/10.1145/2911451.2914726
22. Kumar, V.; Khattar, D.; Gupta, S.; Gupta, M.; Varma, V.: User profiling based deep neural network for temporal news recommendation. IEEE International Conference on Data Mining Work. ICDMW. 2017-Novem, pp. 765–772 (2017). https://doi.org/10.1109/ICDMW.2017.106
23. Cao, S.; Yang, N.; Liu, Z.: Online news recommender based on stacked auto-encoder. In:Proceedings of - 16th IEEE/ACIS International Conference on Computational Science ICIS 2017, pp. 721–726 (2017). https://doi.org/10.1109/ICIS.2017.7960088
24. Park, K.; Lee, J.; Choi, J.: Deep Neural Networks for News Recommendations. In: Proceedings of 2017 ACM Conference on Information and Knowledge Management - CIKM'17, pp. 2255–2258 (2017). https://doi.org/10.1145/3132847.3133154
25. Shani, G.; Heckerman, D.; Brafman, R.I.; Liebman, E.; Saar-Tsechansky, M.; Stone, P.; Zhao, X.; Zhang, L.; Ding, Z.; Yin, D.; Zhao, Y.; Tang, J.; Feng, J.; Li, H.; Huang, M.; Liu, S.; Ou, W.; Wang, Z.; Zhu, X.; Cai, Q.; Filos-Ratsikas, A.; Tang, P.; Zhang, Y.; Zheng, G.; Zhang, F., Zheng, Z.; Xiang, Y.; Yuan, N.J.; Xie, X.; Li, Z.; Mahmood, T.; Ricci, F.; Taghipour, N.; Kardan, A.; Ghidary, S.S.; Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J.; Chapelle, O.; Wu, Q.; Wang, H.; Hong, L.; Shi, Y.; Zhou, M.; Ding, Z.; Tang, J.; Yin, D.; Du, N.; Wang, Y.; He, N.; Sun, J.; Song, L.; Kapoor, K.; Subbian, K.; Srivastava, J.; Schrater, P.; Zhao, X.; Xia, L.; Zhang, L.; Ding, Z.; Yin, D.; Tang, J.; Xia, L.; Tang, J.; Yin, D.; Chen, S.-Y.; Yu, Y.; Da, Q.; Tan, J.; Huang, H.-K.; Tang, H.-H.; Shi, J.-C.; Yu, Y.; Da, Q.; Chen, S.-Y.; Zeng, A.-X.: DRN: A Deep Reinforcement Learning Framework for News Recommendation. In: Proceedings of 2018 World Wide Web Conf. World Wide Web. 6, 113–120 (2018). https://doi.org/10.1145/3178876.3185994
26. Yi, B.; Shen, X.; Zhang, Z.; Shu, J.; Liu, H.: Expanded autoencoder recommendation framework and its application in movie recommendation. In: Ski. 2016 - 2016 10th International Conference on Software, Knowledge, Information Managements Applcations, pp. 298–303 (2017). https://doi.org/10.1109/SKIMA.2016.7916236
27. Vuurens, J.B.P.; Larson, M.; De Vries, A.P.: Exploring deep space: Learning personalized ranking in a semantic space. In: ACM International Conference on Proceeding Series, 15-Septemb, pp. 23–28 (2016). https://doi.org/10.1145/2988450.2988457
28. Zhao, C.; Shi, J.; Jiang, T.; Zhao, J.; Chen, J.: Application of deep belief nets for collaborative filtering. 2016 16th Int. Symp.

Commun. Inf. Technol. Isc. 2016. 201–205 (2016). https://doi.org/10.1109/ISCIT.2016.7751621
29. Sottocornola, G.; Stella, F.; Zanker, M., Canonaco, F.: Towards a deep learning model for hybrid recommendation. In:Proceedings of Int. Conf. Web Intell. - WI'17. 1260–1264 (2017). https://doi.org/10.1145/3106426.3110321
30. Taheri, S.M.; Irajian, I.: DeepMovRS: A unified framework for deep learning-based movie recommender systems. 2018 6th Iran. Jt. Congr. Fuzzy Intell. Syst. 200–204 (2018). https://doi.org/10.1109/CFIS.2018.8336633
31. Fu, M.; Qu, H.; Yi, Z.; Lu, L.; Liu, Y.: A Novel Deep Learning-Based Collaborative Filtering Model for Recommendation System. IEEE Trans. Cybern. 1–13 (2018). https://doi.org/10.1109/TCYB.2018.2795041
32. Yuan, J.; Shalaby, W.; Korayem, M.; Lin, D.; AlJadda, K.; Luo, J.: Solving Cold-Start Problem in Large-scale Recommendation Engines: {A} Deep Learning Approach. CoRR. abs/1611.0, 1901–1910 (2016)
33. Chen, W.; Zhang, X.; Wang, H.; Xu, H.: Hybrid deep collaborative filtering for job recommendation. 2017 2nd IEEE Int. Conf. Comput. Intell. Appl. ICCIA 2017. 2017-Janua, 275–280 (2017). https://doi.org/10.1109/CIAPP.2017.8167222
34. Nguyen, T.T.; Lauw, H.W.: Collaborative Topic Regression with Denoising AutoEncoder for Content and Community Co-Representation. In:Proceedings of 2017 ACM Conference on Information and Knowledge Management - CIKM'17. 2231–2234 (2017). https://doi.org/10.1145/3132847.3133128
35. Wei, J.; He, J.; Chen, K.; Zhou, Y.; Tang, Z.: Collaborative filtering and deep learning based recommendation system for cold start items. Expert Syst. Appl. **69**, 1339–1351 (2017). https://doi.org/10.1016/j.eswa.2016.09.040
36. Zhang, Y.; Yin, H.; Huang, Z.; Du, X.; Yang, G.; Lian, D.: Discrete Deep Learning for Fast Content-Aware Recommendation. In:Proceedings of Elev. ACM Int. Conf. Web Search Data Min. - WSDM'18. 717–726 (2018). https://doi.org/10.1145/3159652.3159688
37. Tan, Y.K.; Xu, X.; Liu, Y.: Improved Recurrent Neural Networks for Session-based Recommendations (2016). https://doi.org/10.1145/2988450.2988452
38. Hidasi, B.; Quadrana, M.; Karatzoglou, A.; Tikk, D.: Parallel Recurrent Neural Network Architectures for Feature-rich Session-based Recommendations. In:Proceedings of 10th ACM Conference Recommendations Systems - RecSys'16, pp. 241–248 (2016). https://doi.org/10.1145/2959100.2959167
39. Greenstein-Messica, A.; Rokach, L.; Friedman, M.: Session-Based Recommendations Using Item Embedding. In:Proceedings of 22nd Int. Conf. Intell. User Interfaces - IUI'17. 629–633 (2017). https://doi.org/10.1145/3025171.3025197
40. Chatzis, S.P.; Christodoulou, P.; Andreou, A.S.: Recurrent Latent Variable Networks for Session-Based Recommendation. In:Proceedings of 2nd Workshop on Deep Learning based Recommender System - DLRS 2017. 38–45 (2017). https://doi.org/10.1145/3125486.3125493
41. Ruocco, M.; Skrede, O.S.L.; Langseth, H.: Inter-Session Modeling for Session-Based Recommendation (2017). https://doi.org/10.1145/3125486.3125491
42. Wang, X.; Wang, Y.: Improving Content-based and Hybrid Music Recommendation using Deep Learning. In: Proceedings of ACM Int. Conf. Multimed. - MM'14. 627–636 (2014). https://doi.org/10.1145/2647868.2654940
43. Chiliguano, P.; Fazekas, G.: Hybrid music recommender using content-based and social information. ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc. 2016-May, 2618–2622 (2016). https://doi.org/10.1109/ICASSP.2016.7472151
44. Oramas, S.; Nieto, O.; Sordo, M.; Serra, X.: A Deep Multimodal Approach for Cold-start Music Recommendation. In:Proceedings of 2nd Workshop on Deep Learning based Recommender System - DLRS 2017. 32–37 (2017). https://doi.org/10.1145/3125486.3125492
45. Jiang, M.; Yang, Z.; Zhao, C.: What to play next? A RNN-based music recommendation system. Conf. Rec. 51st Asilomar Conf. Signals, Syst. Comput. ACSSC 2017. 2017-Octob, 356–358 (2018). https://doi.org/10.1109/ACSSC.2017.8335200
46. Florez, O.U.; Nachman, L.: Deep Learning of Semantic Word Representations to Implement a Content-based Recommender for the RecSys Challenge' 14. 1–5
47. Kim, D.; Park, C.; Oh, J.; Yu, H.: Deep hybrid recommender systems via exploiting document context and statistics of items. Inf. Sci. (Ny) **417**, 72–87 (2017). https://doi.org/10.1016/j.ins.2017.06.026
48. Kim, D.; Park, C.; Oh, J.; Lee, S.; Yu, H.: Convolutional Matrix Factorization for Document Context-Aware Recommendation. In:Proceedings of 10th ACM Conference Recommendations Systems. - RecSys'16. 233–240 (2016). https://doi.org/10.1145/2959100.2959165
49. Bansal, T.; Belanger, D.; McCallum, A.: *Ask the GRU*. In:Proceedings of 10th ACM Conference Recommendations Systems. - RecSys'16. 107–114 (2016). https://doi.org/10.1145/2959100.2959180
50. Wang, X.; Yu, L.; Ren, K.; Tao, G.; Zhang, W.; Yu, Y.; Wang, J.: Dynamic Attention Deep Model for Article Recommendation by Learning Human Editors' Demonstration. In:Proceedings of 23rd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. - KDD'17, pp. 2051–2059 (2017). https://doi.org/10.1145/3097983.3098096
51. Lei, C.; Liu, D.; Li, W.; Zha, Z.-J.; Li, H.: Comparative Deep Learning of Hybrid Representations for Image Recommendations, pp. 2545–2553 (2016). https://doi.org/10.1109/CVPR.2016.279
52. Wang, J.; Kawagoe, K.: Ukiyo-e Recommendation based on Deep Learning For Learning Japanese Art and Culture. In:Proceedings of 2017 International Conference on Information Syst. Data Min. - ICISDM'17. 119–123 (2017). https://doi.org/10.1145/3077584.3077612
53. Peska, L.; Trojanova, H.: Towards Recommender Systems for Police Photo Lineup. In:Proceedings of 2nd Workshop on Deep Learning based Recommender System - DLRS 2017. 19–23 (2017). https://doi.org/10.1145/3125486.3125490
54. Deng, S.; Huang, L.; Xu, G.; Wu, X.; Wu, Z.: On Deep Learning for Trust-Aware Recommendations in Social Networks. IEEE Trans. Neural Networks Learn. Syst. **28**, 1164–1177 (2017). https://doi.org/10.1109/TNNLS.2016.2514368
55. Dang, Q.V.; Ignat, C.L.: DTrust: A Simple Deep Learning Approach for Social Recommendation. In:Proceedings of - 2017 IEEE 3rd Int. Conf. Collab. Internet Comput. CIC 2017. 2017-Janua, 209–218 (2017). https://doi.org/10.1109/CIC.2017.00036
56. Rafailidis, D.; Crestani, F.: Recommendation with Social Relationships via Deep Learning. In:Proceedings of ACM SIGIR Int. Conf. Theory Inf. Retr. - ICTIR'17. 151–158 (2017). https://doi.org/10.1145/3121050.3121057
57. Tomar, A.; Godin, F.; Vandersmissen, B.; De Neve, W.; Van De Walle, R.: Towards Twitter hashtag recommendation using distributed word representations and a deep feed forward neural network. In:Proceedings of 2014 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2014. 362–368 (2014). https://doi.org/10.1109/ICACCI.2014.6968557
58. Zuo, Y.; Zeng, J.; Gong, M.; Jiao, L.: Tag-aware recommender systems based on deep neural networks. Neurocomputing. **204**, 51–60 (2016). https://doi.org/10.1016/j.neucom.2015.10.134
59. Xu, Z.; Chen, C.; Lukasiewicz, T.; Miao, Y.; Meng, X.: Tag-Aware Personalized Recommendation Using a Deep-Semantic Similarity Model with Negative Sampling. In:Proceedings of 25th ACM

International Conference on Information and Knowledge Management - CIKM'16. 1921–1924 (2016). https://doi.org/10.1145/2983323.2983874
60. Wang, F.; Qu, Y.; Zheng, L.; Lu, C.T.; Yu, P.S.: Deep and Broad Learning on Content-Aware POI Recommendation. In:Proceedings of - 2017 IEEE 3rd Int. Conf. Collab. Internet Comput. CIC 2017. 2017-Janua, 369–378 (2017). https://doi.org/10.1109/CIC.2017.00054
61. Yin, H.; Wang, W.; Wang, H.; Chen, L.; Zhou, X.: Spatial-Aware Hierarchical Collaborative Deep Learning for POI Recommendation. IEEE Trans. Knowl. Data Eng. **29**, 2537–2551 (2017). https://doi.org/10.1109/TKDE.2017.2741484
62. Xia, B.; Li, Y.; Li, Q.; Li, T.: Attention-based recurrent neural network for location recommendation. In:Proceedings of 2017 12th Int. Conf. Intell. Syst. Knowl. Eng. ISKE 2017. 2018-Janua, 1–6 (2018). https://doi.org/10.1109/ISKE.2017.8258747
63. Ebesu, T.; Fang, Y.: Neural Citation Network for Context-Aware Citation Recommendation. In:Proceedings of 40th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR'17. 1093–1096 (2017). https://doi.org/10.1145/3077136.3080730
64. Huck-Fries, V.; Wiegand, F.; Klinker, K.; Wiesche, M.; Krcmar, H.: Reranking-based Recommender System with Deep Learning. Inform. 2017. 585–596 (2017). https://doi.org/10.18420/in2017
65. Hassan, H.A.M.: Personalized Research Paper Recommendation using Deep Learning. In:Proceedings of 25th Conf. User Model. Adapt. Pers. - UMAP'17. 327–330 (2017). https://doi.org/10.1145/3079628.3079708
66. Suglia, A.; Greco, C.; Musto, C.; de Gemmis, M.; Lops, P.; Semeraro, G.: A Deep Architecture for Content-based Recommendations Exploiting Recurrent Neural Networks. In:Proceedings of 25th Conf. User Model. Adapt. Pers. - UMAP'17. 202–211 (2017). https://doi.org/10.1145/3079628.3079684
67. Smirnova, E.; Vasile, F.: Contextual Sequence Modeling for Recommendation with Recurrent Neural Networks. (2017). https://doi.org/10.1145/3125486.3125488
68. Verma, M.; Ganguly, D.: LiRME: Locally interpretable ranking model explanation. SIGIR 2019 - Proc. 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval 1281–1284 (2019). https://doi.org/10.1145/nnnnnnn.nnnnnnn
69. Hongliang, C.; Xiaona, Q.: The video recommendation system based on DBN. In:Proceedings of - 15th IEEE The International Conference on Information and Computer Technologies CIT 2015, 14th IEEE Int. Conf. Ubiquitous Comput. Commun. IUCC 2015, 13th IEEE Int. Conf. Dependable, Auton. Se. 1016–1021 (2015). https://doi.org/10.1109/CIT/IUCC/DASC/PICOM.2015.154
70. Covington, P.; Adams, J.; Sargin, E.: Deep Neural Networks for YouTube Recommendations. ACM Conference Recommendations Systems. 191–198 (2016). https://doi.org/10.1145/2959100.2959190
71. Lee, H.; Ahn, Y.; Lee, H.; Ha, S.; Lee, S.: Quote Recommendation in Dialogue using Deep Neural Network. In:Proceedings of 39th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR'16. 957–960 (2016). https://doi.org/10.1145/2911451.2914734
72. Tan, J.; Wan, X.; Xiao, J.: A Neural Network Approach to Quote Recommendation in Writings. In:Proceedings of 25th ACM International Conference on Information and Knowledge Management - CIKM'16. 65–74 (2016). https://doi.org/10.1145/2983323.2983788
73. Zhang, H.; Yang, H.; Huang, T.; Zhan, G.: DBNCF: Personalized courses recommendation system based on DBN in MOOC environment. In:Proceedings of - 2017 Int. Symp. Educ. Technol. ISET 2017. 106–108 (2017). https://doi.org/10.1109/ISET.2017.33
74. Wang, X.; Zhang, Y.; Yu, S.; Liu, X.; Yuan, Y.; Wang, F.Y.: E-learning recommendation framework based on deep learning. 2017 IEEE Int. Conf. Syst. Man, Cybern. SMC 2017. 2017-Janua, 455–460 (2017). https://doi.org/10.1109/SMC.2017.8122647
75. Li, P.; Wang, Z.; Ren, Z.; Bing, L.; Lam, W.: Neural Rating Regression with Abstractive Tips Generation for Recommendation. 345–354 (2017). https://doi.org/10.1145/3077136.3080822
76. Maheshwary, S.; Misra, H.: Matching Resumes to Jobs via Deep Siamese Network. Companion Proc. Web Conf. **2018**, 87–88 (2018). https://doi.org/10.1145/3184558.3186942
77. Jaradat, S.: Deep Cross-Domain Fashion Recommendation. In:Proceedings of Elev. ACM Conference Recommendations Systems. - RecSys'17. 407–410 (2017). https://doi.org/10.1145/3109859.3109861
78. Jiang, S.; Wu, Y.; Fu, Y.: 5 Deep Bidirectional Cross-Triplet Embedding for Online Clothing Shopping. ACM Trans. Multimed. Comput. Commun. Appl. Artic. 14, 1–22 (2018). https://doi.org/10.1145/3152114
79. Webb, T.; Harnden, D.G.: The transformation by simian virus 40 of cells from patients with mucopolysaccharidosis and from normal controls. Cancer Res. **36**, 203–212 (1976)
80. Zahalka, J.; Rudinac, S.; Worring, M.: Interactive multimodal learning for venue recommendation. IEEE Trans. Multimed. **17**, 2235–2244 (2015). https://doi.org/10.1109/TMM.2015.2480007
81. Gao, T.; Li, X.; Chai, Y.; Tang, Y.: Deep learning with consumer preferences for recommender system. 2016 IEEE International Conference on Information Autom. IEEE ICIA 2016. 1556–1561 (2017). https://doi.org/10.1109/ICInfA.2016.7832066
82. Dai, H.; Wang, Y.; Trivedi, R.; Song, L.: Recurrent coevolutionary latent feature processes for continuous-time recommendation. ACM Int. Conf. Proceeding Ser. 15-Septemb, 29–34 (2016). https://doi.org/10.1145/2988450.2988451
83. Dominguez, V.; Messina, P.; Parra, D.; Mery, D.; Trattner, C.; Soto, A.: Comparing Neural and Attractiveness-based Visual Features for Artwork Recommendation. In:Proceedings of 2nd Workshop on Deep Learning based Recommender System - DLRS 2017. 55–59 (2017). https://doi.org/10.1145/3125486.3125495
84. Soh, H.; Sanner, S.; White, M.; Jamieson, G.: Deep Sequential Recommendation for Personalized Adaptive User Interfaces. In:Proceedings of 22nd Int. Conf. Intell. User Interfaces - IUI'17. 589–593 (2017). https://doi.org/10.1145/3025171.3025207
85. Jishan, S.T.; Wang, Y.: Audience Activity Recommendation Using Stacked-LSTM Based Sequence Learning. In:Proceedings of 9th Int. Conf. Mach. Learn. Comput. - ICMLC 2017. 98–106 (2017). https://doi.org/10.1145/3055635.3056606
86. Wu, H.; Zhang, Z.; Yue, K.; Zhang, B.; He, J.; Sun, L.: Dual-regularized matrix factorization with deep neural networks for recommender systems. Knowledge-Based Syst. **145**, 1–14 (2018). https://doi.org/10.1016/j.knosys.2018.01.003
87. Yuan, W.; Li, C.; Guan, D.; Han, G.; Khattak, A.M.: Socialized healthcare service recommendation using deep learning. Neural Comput. Appl. **30**, 2071–2082 (2018). https://doi.org/10.1007/s00521-018-3394-4
88. Katzman, J.L.; Shaham, U.; Cloninger, A.; Bates, J.; Jiang, T.; Kluger, Y.: DeepSurv: Personalized Treatment Recommender System Using A Cox Proportional Hazards Deep Neural Network. 1–12 (2018). https://doi.org/10.1186/s12874-018-0482-1
89. Wang, H.; Yeung, D.: IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING Towards Bayesian Deep Learning: A Framework and Some Existing Methods. 1–14
90. Bai, T.; Wen, J.-R.; Zhang, J.; Zhao, W.X.: A Neural Collaborative Filtering Model with Interaction-based Neighborhood. In:Proceedings of 2017 ACM Conference on Information and

Knowledge Management - CIKM'17. 1979–1982 (2017). https://doi.org/10.1145/3132847.3133083
91. He, X.; Chua, T.-S.: Neural Factorization Machines for Sparse Predictive Analytics. 355–364 (2017). https://doi.org/10.1145/3077136.3080777
92. Lee, W.; Song, K.; Moon, I.-C.: Augmented Variational Autoencoders for Collaborative Filtering with Auxiliary Information. In:Proceedings of 2017 ACM Conference on Information and Knowledge Management - CIKM'17. 1139–1148 (2017). https://doi.org/10.1145/3132847.3132972
93. Zhang, W.; Liu, F.; Jiang, L.; Xu, D.: Recommendation based on collaborative filtering by convolution deep learning model based on label weight nearest neighbor. In:Proceedings of - 2017 10th Int. Symp. Comput. Intell. Des. Isc. 2017. 2, 504–507 (2018). https://doi.org/10.1109/ISCID.2017.235
94. Liu, J.; Wang, D.: PHD : A Probabilistic Model of Hybrid Deep Collaborative Filtering for Recommender Systems. Acml. 1–16 (2017)
95. Xue, H.J.; Dai, X.Y.; Zhang, J.; Huang, S.; Chen, J.: Deep matrix factorization models for recommender systems. IJCAI Int. Jt. Conf. Artif. Intell. 3203–3209 (2017). https://doi.org/10.24963/ijcai.2017/447
96. Tay, Y.; Luu, A.T.; Hui, S.C.: Latent Relational Metric Learning via Memory-based Attention for Collaborative Ranking. 729–739 (2017). https://doi.org/10.1145/3178876.3186154
97. Dong, X.; Yu, L.; Wu, Z.; Sun, Y.; Yuan, L.; Zhang, F.: A Hybrid Collaborative Filtering Model with Deep Structure for Recommender Systems. In:Proceedings of 31st AAAI Conf. Artif. Intell. 1309–1315 (2017). https://doi.org/10.1103/PhysRevLett.93.077207
98. Catherine, R.; Cohen, W.: TransNets: Learning to Transform for Recommendation. 288–296 (2017). https://doi.org/10.1145/3109859.3109878
99. Preethi, G.; Krishna, P.V.; Obaidat, M.S.; Saritha, V.; Yenduri, S.: Application of deep learning to sentiment analysis for recommender system on cloud. IEEE CITS 2017 - 2017 Int. Conf. Comput. Inf. Telecommun. Syst. 93–97 (2017). https://doi.org/10.1109/CITS.2017.8035341
100. Serrà, J.; Karatzoglou, A.: Getting deep recommenders fit: Bloom embeddings for sparse binary input/output networks. 279–287 (2017). https://doi.org/10.1145/3109859.3109876
101. Wei, J.; He, J.; Chen, K.; Zhou, Y.; Tang, Z.: Collaborative Filtering and Deep Learning Based Hybrid Recommendation for Cold Start Problem. In:Proceedings of - 2016 IEEE 14th Int. Conf. Dependable, Auton. Secur. Comput. DASC 2016, 2016 IEEE 14th Int. Conf. Pervasive Intell. Comput. PICom 2016, 2016 IEEE 2nd Int. Conf. Big Data. 874–877 (2016). https://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2016.149
102. Wang, H.; Wang, N.; Yeung, D.-Y.: Collaborative Deep Learning for Recommender Systems, pp. 1235–1244 (2014). https://doi.org/10.1145/2783258.2783273
103. Li, Q.; Zheng, X.; Wu, X.: Collaborative Autoencoder for Recommender Systems, pp. 305–314 (2017). https://doi.org/10.1145/3097983.3098077
104. Cheng, H.-T.; Koc, L.; Harmsen, J.; Shaked, T.; Chandra, T.; Aradhye, H.; Anderson, G.; Corrado, G.; Chai, W.; Ispir, M.; Anil, R.; Haque, Z.; Hong, L.; Jain, V.; Liu, X.; Shah, H.: Wide & Deep Learning for Recommender Systems. (2016). https://doi.org/10.1145/2988450.2988454
105. Chen, C.; Zhao, P.; Li, L.; Zhou, J.; Li, X.; Qiu, M.: Locally Connected Deep Learning Framework for Industrial-scale Recommender Systems. In:Proceedings of 26th International Conference on World Wide Web Companion - WWW'17 Companion, pp. 769–770 (2017). https://doi.org/10.1145/3041021.3054227
106. Da'u, A.; Salim, N.; Rabiu, I.; Osman, A.: Recommendation system exploiting aspect-based opinion mining with deep learning method. Inf. Sci. (Ny). 512, 1279–1292 (2020). https://doi.org/10.1016/j.ins.2019.10.038
107. Zhang, S.; Yao, L.; Sun, A.; Tay, Y.: Deep learning based recommender system: a survey and new perspectives. ACM Comput. Surv. **52**, 1–35 (2019). https://doi.org/10.1145/3285029
108. Bobadilla, J.; Alonso, S.; Hernando, A.: Deep learning architecture for collaborative filtering recommender systems. Appl. Sci. (2020). https://doi.org/10.3390/app10072441
109. Zarzour, H.; Al-Sharif, Z.A.; Jararweh, Y.: RecDNNing: A recommender system using deep neural network with user and item embeddings. 2019 10th International Conference on Information Commun. Syst. ICICS 2019. 99–103 (2019). https://doi.org/10.1109/IACS.2019.8809156
110. Fessahaye, F.; Perez, L.; Zhan, T.; Zhang, R.; Fossier, C.; Markarian, R.; Chiu, C.; Zhan, J.; Gewali, L.; Oh, P.: T-RECSYS: a novel music recommendation system using deep learning. In: 2019 IEEE International Conference on Consumer Electronics. ICCE 2019. (2019). https://doi.org/10.1109/ICCE.2019.8662028
111. Lee, H.; Lee, J.: Scalable deep learning-based recommendation systems. ICT Express. **5**, 84–88 (2019). https://doi.org/10.1016/j.icte.2018.05.003
112. Nimirthi, P.; Venkata Krishna, P.; Obaidat, M.S.; Saritha, V.: A framework for sentiment analysis based recommender system for agriculture using deep learning approach. SpringerBriefs Appl. Sci. Technol. (2019). https://doi.org/10.1007/978-981-13-1456-8_5
113. Nweke, H.F.; Teh, Y.W.; Al-garadi, M.A.; Alo, U.R.: Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the art and research challenges. Expert Syst. Appl. **105**, 233–261 (2018). https://doi.org/10.1016/j.eswa.2018.03.056
114. Zheng, L.: A survey and critique of deep learning on recommender systems. (2016)
115. Da'u, A.; Salim, N.; Rabiu, I.; Osman, A.: Weighted aspect-based opinion mining using deep learning for recommender system. Expert Syst. Appl. (2020). https://doi.org/10.1016/j.eswa.2019.112871
116. Wang, S.; Huang, C.; Li, J.; Yuan, Y.; Wang, F.Y.: Decentralized construction of knowledge graphs for deep recommender systems based on blockchain-powered smart contracts. IEEE Access. **7**, 136951–136961 (2019). https://doi.org/10.1109/ACCESS.2019.2942338
117. Huang, Z.; Tang, J.; Shan, G.; Ni, J.; Chen, Y.; Wang, C.: An efficient passenger-hunting recommendation framework with multitask deep learning. IEEE Internet Things J. **6**, 7713–7721 (2019). https://doi.org/10.1109/JIOT.2019.2901759
118. Vincent, P.; Larochelle, H.: Extracting and Composing Robust Features with Denoising.pdf. 1096–1103 (2008)
119. Zhang, X.; Zhong, J.; Liu, K.: Wasserstein autoencoders for collaborative filtering. Neural Comput. Appl. (2020). https://doi.org/10.1007/s00521-020-05117-w
120. Deng, X.; Huangfu, F.: Collaborative variational deep learning for healthcare recommendation. IEEE Access. **7**, 55679–55688 (2019). https://doi.org/10.1109/ACCESS.2019.2913468
121. Pan, Y.; He, F.; Yu, H.: Learning social representations with deep autoencoder for recommender system. World Wide Web **23**, 2259–2279 (2020). https://doi.org/10.1007/s11280-020-00793-z
122. Saravanan, B.; Mohanraj, V.; Senthilkumar, J.: A fuzzy entropy technique for dimensionality reduction in recommender systems using deep learning. Soft. Comput. **23**, 2575–2583 (2019). https://doi.org/10.1007/s00500-019-03807-9
123. Guan, Y.; Wei, Q.; Chen, G.: Deep learning based personalized recommendation with multi-view information integration. Decis. Support Syst. **118**, 58–69 (2019). https://doi.org/10.1016/j.dss.2019.01.003

124. Wang, K.; Xu, L.; Huang, L.; Wang, C.D.; Lai, J.H.: SDDRS: stacked discriminative denoising auto-encoder based recommender system. Cogn. Syst. Res. **55**, 164–174 (2019). https://doi.org/10.1016/j.cogsys.2019.01.011
125. Zhang, Y.; Yin, C.; Wu, Q.; He, Q.; Zhu, H.: Location-aware deep collaborative filtering for service recommendation. IEEE Trans. Syst. Man, Cybern. Syst. (2019). https://doi.org/10.1109/tsmc.2019.2931723
126. Ahamed, M.T.; Afroge, S.: A Recommender System Based on Deep Neural Network and Matrix Factorization for Collaborative Filtering. 2nd Int. Conf. Electr. Comput. Commun. Eng. ECCE 2019. 1–5 (2019). https://doi.org/10.1109/ECACE.2019.8679125
127. Nassar, N.; Jafar, A.; Rahhal, Y.: A novel deep multi-criteria collaborative filtering model for recommendation system. Knowl.-Based Syst. **187**, 104811 (2020). https://doi.org/10.1016/j.knosys.2019.06.019
128. Feinman, R.: A Deep Belief Network Approach to Learning Depth From Optical Flow, pp. 1–14
129. Pacheco, A.G.C.; Krohling, R.A.; da Silva, C.A.S.: Restricted Boltzmann machine to determine the input weights for extreme learning machines. Expert Syst. Appl. **96**, 77–85 (2018). https://doi.org/10.1016/j.eswa.2017.11.054
130. Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Netw. **61**, 85–117 (2015). https://doi.org/10.1016/j.neunet.2014.09.003
131. Hinton, G.E.; Osindero, S.; Teh, Y.-W.: A fast learning algorithm for deep belief nets. Neural Comput. **18**, 1527–1554 (2006). https://doi.org/10.1162/neco.2006.18.7.1527
132. Luo, L.; Zhang, S.; Wang, Y.; Peng, H.: An alternate method between generative objective and discriminative objective in training classification Restricted Boltzmann Machine. Knowl.-Based Syst. **144**, 144–152 (2018). https://doi.org/10.1016/j.knosys.2017.12.032
133. Liu, W.; Wang, Z.; Liu, X.; Zeng, N.; Liu, Y.; Alsaadi, F.E.: A survey of deep neural network architectures and their applications. Neurocomputing **234**, 11–26 (2017). https://doi.org/10.1016/j.neucom.2016.12.038
134. Hidasi, B.; Karatzoglou, A.: Recurrent Neural Networks with Top-k Gains for Session-Based Recommendations, pp. 370–371 (2017). https://doi.org/10.1145/3269206.3271761
135. Da'U, A.; Salim, N.: Sentiment-aware deep recommender system with neural attention networks. IEEE Access. **7**, 45472–45484 (2019). https://doi.org/10.1109/ACCESS.2019.2907729

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-020-09837-y

SketchFormer: transformer-based approach for sketch recognition using vector images

Anil Singh Parihar[1] · **Gaurav Jain**[1] · **Shivang Chopra**[1] · **Suransh Chopra**[1]

Received: 26 February 2020 / Revised: 10 August 2020 / Accepted: 9 September 2020 /
Published online: 07 November 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Sketches have been employed since the ancient era of cave paintings for simple illustrations to represent real-world entities and communication. The abstract nature and varied artistic styling make automatic recognition of these drawings more challenging than other areas of image classification. Moreover, the representation of sketches as a sequence of strokes instead of raster images introduces them at the correct abstract level. However, dealing with images as a sequence of small information makes it challenging. In this paper, we propose a Transformer-based network, dubbed as AttentiveNet, for sketch recognition. This architecture incorporates ordinal information to perform the classification task in real-time through vector images. We employ the proposed model to isolate the discriminating strokes of each doodle using the attention mechanism of Transformers and perform an in-depth qualitative analysis of the isolated strokes for classification of the sketch. Experimental evaluation validates that the proposed network performs favorably against state-of-the-art techniques.

Keywords Sketch recognition · Transformers · Vector images · Deep learning

1 Introduction

The ability to draw and comprehend eclectic notions through sketches reflects the intellectual capabilities of human beings. Drawing has long been a part of human behavior,

✉ Anil Singh Parihar
parihar.anil@gmail.com

Gaurav Jain
gauravjain13298@gmail.com

Shivang Chopra
shivangchopra11@gmail.com

Suransh Chopra
suransh2008@gmail.com

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

mutating its utility from rock carvings in the ancient era, to blueprints of drafts in the modern age. Additionally, these sketch drawings assist humans in communication and creative design, such as art. Hence, intelligent machines must efficiently fathom the rapid proliferation of interaction with these human activities to be pervasive in all environments. Sketch analysis has been efficiently utilised in diverse domains like mathematics [18], chemistry [24], and electronics [23]. Further, sketch analysis spans a wide spectrum of applications, such as sketch segmentation [16, 19, 20, 35], image retrieval [12, 41, 43], and sketch recognition [11, 46, 48].

In this work, we focus on *sketch recognition*, which aims to identify human drawings and classify them into their respective categories. Although existing methods provide satisfactory accuracy performance for *image recognition* [17], only a few approaches have addressed the issue of identifying human drawn illustrations [27, 28, 45]. Sketch recognition was first introduced by Sutherland et al. [36]. Since then, several approaches have been introduced for better sketch comprehension in this active area of research. Conventional studies [28, 45] consider sketches as nothing but images, converting them to binary representations and extracting features for training classifiers like Support Vector Machines (SVM). With the success of deep learning-based approaches in recent years for image classification tasks [17], these techniques have gained importance in sketch analysis. In particular, convolutional neural network-based approaches [25, 46] have outperformed the conventional techniques in sketch recognition with significantly dominant accuracy performances. However, most of the techniques often overlook the characteristic properties of sketches, and hence leave room for improvement.

Perceiving sketch is a challenging task due to two main reasons, (1) heterogeneous representations and (2) level of abstraction. Representation of the same object is dependent upon the interpretation of that object, which consequently leads to high intraclass variation between samples of the same class. Figure 1 illustrates this diversity among different samples of the same class. For example, distinct views of a cat in Fig. 1b include different representations which render the task of finding patterns relatively difficult for the

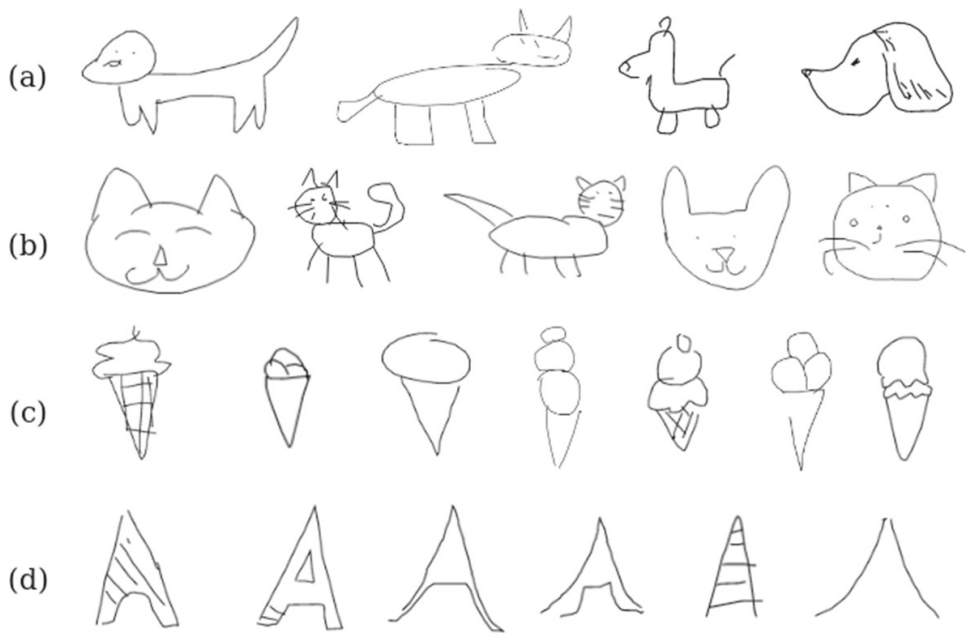


Fig. 1 Visualization of few classes in the QuickDraw [14] dataset, **a** Dog, **b** Cat, **c** Ice-cream, **d** Eiffel Tower showing the inter-class and intra-class homophilic behaviour and the highly abstract nature of sketches which poses significant challenges to the models

models. Alternatively, sketches of one class could be interpreted as belonging to another class. For instance, the second sketch of the dog class in Fig. 1a resembles a cat shown in Fig. 1b. Hence, learning such overlapping representations which may exhibit a significantly anomalous, yet acceptable, view proves to be an arduous task. Further, pixel-level images of sketches misinterpret how humans actually comprehend these drawings on an abstract level. Sketches are rendered as a series of strokes that build up the desired illustration over time. Hence, a more plausible expression of these sketches is the *vector image format*, which represents sketch drawings as a sequence of pen strokes. Besides, sketches generated using these vector image formats in Sketch-RNN [14] have corroborated their aptness for sketches. Therefore, an approach that deals with these complex correlations among sketches, interpreting sketches as a sequential abstraction is required. Moreover, applications like Handwriting Recognition, Optical Character Recognition (OCR) can hugely benefit from the advent of a robust classifier that can handle data with such huge variations.

In order to address the aforementioned issues, a transformer-based approach is proposed for attentive sketch recognition. The main contributions of the proposed approach are as follows:

- **Sketch Recognition using Transformers:** This is the first approach to the best of our knowledge that employs transformers for sketch recognition. We leverage the attention mechanism of Transformers to identify objects. For this, sketches are interpreted as a sequence of strokes, like humans actually comprehend drawings in real-time.
- **Attention-based analysis of sketches:** Through extensive qualitative analysis of the sketches using the attention mechanism, we isolate parts of the sketches necessary for object classification.

The rest of the paper is organized as follows: In Section 2, a literature review has been conducted to discuss related approaches. Section 3 discusses the preliminary concepts used in the proposed architecture. The proposed approach is presented in Section 4, which is followed by an exhaustive qualitative and quantitative experimental evaluation of the proposed technique is in Sections 5 and 6. Finally, Section 7 concludes the paper along with future directions.

2 Related work

In this section, existing frameworks for sketch classification have been discussed in depth. These approaches can be broadly classified as, (1) Conventional Classifiers, (2) Convolutional Neural Network (CNN) based approaches, and (3) Recurrent Neural Network (RNN) based approaches.

2.1 Conventional classifiers

Most of the traditional methods employ raster images as input. These pre-processed pixel-level images are used to extract local features from these sketches. These features are then utilized to classify sketches, using either classifier or guiding different learning strategies. For instance, Eitz et al. [11] proposed to represent sketches as bag-of-features, which trains a multi-class Support Vector Machine (SVM) to classify sketches. On the other hand, Schneider et al. [28] chose a Fisher Vector representation of raster images. In this, a Gaussian Mixture Model (GMM) was first generated and then used to encode the image using its deviation from the GMM, in terms of the Fisher Vector representation. Finally, an SVM

is trained to classify images. In order to overcome the high structural complexity and absence of visual cues, Li et al. [21] proposed a star graph representation and employed a multi-kernel framework that fuses features extracted using different approaches.

Further, Yanki et al. [45] focused on reducing the amount of annotated data required, while maintaining recognition accuracy. For this, active learning was introduced, which measures the informativeness of unlabeled samples and selects batches of informative samples. Alternative approaches have been introduced by exploiting domain-specific knowledge [5, 7, 8, 32, 48, 51]. Although these approaches made progress in sketch recognition, most of these methods failed to address the complex and subtle style variations. Additionally, freehand sketches are casually drawn and hence pose the issue of inaccurate inputs, which further aggravates the complexity involved in sketch recognition.

2.2 Convolutional neural network-based

Recently, deep learning-based models have been found to perform favorably against traditional methods for classification and detection in diverse fields [22, 34, 38, 39] and even for sketch recognition [25, 42, 47, 50]. Several approaches on optimization algorithms have also been proposed [1–4, 6]. Among these, CNN based methods that use the same raster images as input have considerably improved recognition accuracy. In [26], authors propose a residual network (ResNet) based CNN architecture for freehand sketch recognition, validating that deeper networks perform favorably in comparison to shallow networks. Although the performance of different architectures were compared, this work lacks an exhaustive hyper-parameter tuning and discusses networks only as deep as 25 layers. Similarly, Yang et al. [44] proposed a deep CNN model, in order to overcome the need for feature engineering, usually required in traditional approaches. The authors further extended recognition capabilities to design an image retrieval method.

In [29], the authors proposed to extract features using deep CNN, and employed K-nearest neighbors for similarity search, in order to retrieve images. Additionally, the authors proposed to extract medium and high-level features that are extracted from deeper and shallow layers of the network, respectively. Seddati et al. [30] proposed DeepSketch 3, which utilizes the capabilities of deep convolutional neural networks for diverse applications like, partial sketch recognition, and sketch retrieval using query-by-example. In [31], authors propose to explore transfer learning for sketch recognition. Recently, Zhang et al. [49] proposed to exploit both the appearance and shape representations in a sketch. For this, a hybrid CNN was introduced, which extracted features using two different CNNs, one for shape, and the other for appearance. These features were then used to train an SVM and classify sketches. Several efforts have been focused on exploiting the powerful CNN architectures like Alexnet [17], and Google net [37].

Although a CNN-based approach [46] was the first to surpass human performance on the TU-Berlin dataset [11], these approaches employ a flawed representation, which proves to be the bottleneck in achieving even superior accuracy. Using current state-of-the-art deep CNN -based recognition models for this task proves unavailing due to the fact that the sketches are far less visually complex than photographs. The absence of color channels and the highly abstract nature of the sketches can be observed in Fig. 1d, where it is very hard for most of the sketches to qualify as the respective class. Furthermore, while people can agree on what an object looks like, how they ultimately render the object can vary significantly, thereby leading to lots of artistic variation. This lack of homophily in the sketches and their abstract nature renders most of the existing works done for sketch recognition at a serious disadvantage.

2.3 Recurrent neural network-based

In order to address the issue of faulty representation, Sketch-RNN [14] is the first approach to represent sketches in the vector image format. In this, recurrent neural networks (RNNs) process these vectors that constitute a sequence of strokes. The RNN encoder and RNN decoder, collectively compute the temporal correlation between strokes in order to classify sketches in their respective category. Vector representations have been proved to perform favorably for tasks such as sketch generation [13, 33] as well. Although RNN based approaches have been able to overcome the representation barriers, these approaches fail to model the underlying distribution when the number of object categories increase. Most of these methods use Long Short Term Memory (LSTM) units to encode the temporal information. While LSTMs have been exhaustively employed in both natural language processing applications and time series analysis, they have been observed to provide lower performance for large and complex sequences.

In order to model such sequences, *attention mechanism* [9] has been brought into action. This allows models to direct their resources to part of the information that most effectively represents that object. Consequently, facilitating models to achieve superior accuracy with lower data requirements. Our approach aligns with this ideology, and hence adapts the state-of-the-art RNN model, Tranformers [40] to these requirements. RNNs and Long Short Term Memory (LSTM) cells fail to encode long term dependencies and adapt to variable-length inputs. To address these issues, we propose to exploit the self-attention mechanism in Transformers, along with its capability to encode larger information, to identify multiple classes with one model. The following section discusses the proposed approach in detail.

3 Preliminaries

LSTMs perform recurrent operations in a sequential manner. On the contrary, Transformers [40] replace these complex recurrent computations with the attention mechanism. This allows parallel computation, completely relying on self-attention, thereby reducing computation complexity.

The transformer forms the building block in our model, and we introduce some background on the attention mechanism it uses. Attention can be described as a mapping between the query vector (Q) and key-value pairs (K, V) to obtain an output vector. The output vector is the weighted sum of values, $V = \{v_i | i \in \{1, 2, ..., d\}\}$, where Attention ($\mathcal{A}$) is computed using the *scalar-dot product* between a query and the keys, followed by its softmax, to obtain the weights of values:

$$\mathcal{A}(Q, K, V) = \text{SOFTMAX}\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{1}$$

where $K \in \mathbb{R}^d$. With a single attention, averaging disarms the model to combine information from different representations. In order to facilitate this, *multi-head self-attention* is performed (Fig. 2). In this, queries, keys, and values are linearly projected to obtain varying set of projections to incorporate representations at different positions. Attention is computed in parallel for m attention heads. These are then concatenated to generate the final output. Equation (2) computes the output:

$$\mathcal{MA}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, ..., \text{head}_m)\mathbf{W}^O \tag{2}$$

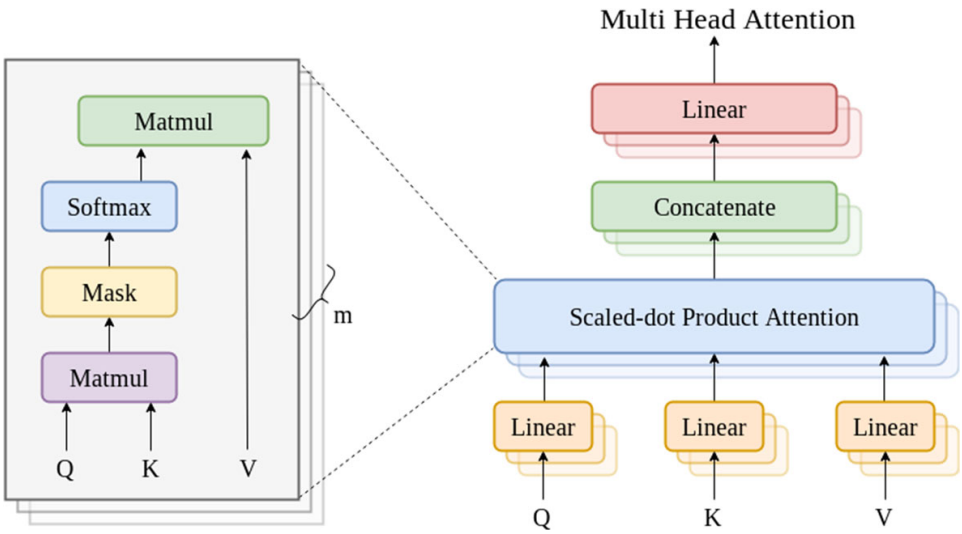


Fig. 2 Internal structure of the Multi-Headed Self-Attention Mechanism in a Transformer block explaining the various operations involved in obtaining the multi-head attention maps used for the ablation and attention-based analysis done in Section 5

where $\text{head}_j = \mathcal{A}(Q\mathbf{W}_j^Q, K\mathbf{W}_j^K, V\mathbf{W}_j^V)$, such that the project matrices $\mathbf{W}_j^Q, \mathbf{W}_j^K, \mathbf{W}_j^V \in \mathbb{R}^{\frac{d}{m} \times d}$, $\forall j \in \{1, 2, ..., m\}$, and $\mathbf{W}^O \in \mathbb{R}^{d \times d}$. The proposed network uses multi-head attention with $m = 8$ attention heads, along with $d = 64$. The parallel computation of these heads is analogous to performing a single attention. Hence, allowing the model to capture positional and temporal information without increasing time complexity. The following section discusses the role of these multi-head attention in the proposed approach in detail.

4 Proposed methodology

In this section, we establish the motivation and then describe the technical details of the proposed approach. Figure 3 shows the proposed architecture. Firstly, the data is pre-processed to transform it into a scale and rotation invariant format. Next, the Auto-Encoder module extracts a latent vector from the input to increase the input dimension, followed by processing of this latent vector by the Transformer module. Finally, the class probabilities are used to obtain sketch category.

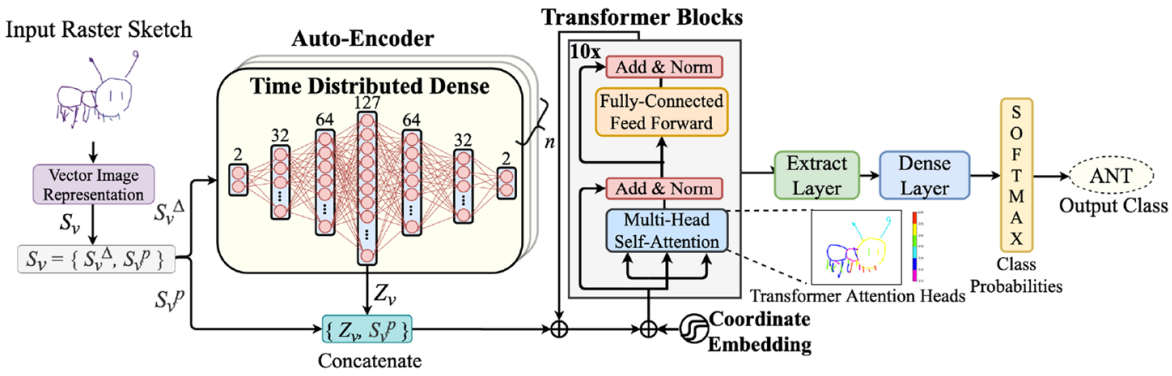


Fig. 3 Overview of the proposed Network depicting the pre-processing of raster sketch to form the vector image representation. Projecting the input to a higher latent dimension using the Auto-Encoder and the final classification using the attention maps from the Transformer blocks connected via skip connections

4.1 Motivation

The proposed approach is based on the premise of choosing a vector-image representation to facilitate real-time detection. This format enables the data to be proposed across the *time* dimension, instead of the *space* dimension as in case of a CNN-based approach. The major advantage of the proposed approach is to be able to detect sketches in real-time, i.e., even partially rendered sketches can be processed. Hence, we employ a Transformer-based architecture to process the sequential data.

4.2 Input representation

The input data is preprocessed to transform the sketches into the vector-image format. Our representation is an extension of the format used by Graves [13]. Sketches are represented using a multi-state pen stroke event. In this, $\mathcal{S}$ is a sketch with sequence of strokes s_i, i.e. $\mathcal{S} = \{s_1, s_2, ..., s_n\}$, where n is the sequence length. Each stroke, s_i, is defined using a 3-point format:

$$s_i = \{\Delta x_i, \Delta y_i, p_i\}, \quad \forall i \in \{1, 2, ..., n\} \tag{3}$$

where $(\Delta x_i, \Delta y_i)$ is the offset distance in the x and y direction. For each sketch, $(\Delta x_1, \Delta y_1)$ begin with origin as the initial starting point. Pen-state, p_i is a binary variable, with $p_i = 1$ indicating that the pen is in contact with surface and a stroke is drawn over the offsets $(\Delta x_i, \Delta y_i)$, while $p_i = 0$ represents that the pen is lifted off the surface and moved from the previous point in the direction of offsets. The following subsection discusses the processing of these inputs, $\mathcal{S}$, which is fed into the proposed network to obtain respective object categories.

4.3 AttentiveNet

Figure 3 illustrates the proposed architecture, which consists of two modules, (1) *auto-encoder*, and (2) *transformer-encoder*. The auto-encoder module extracts features from the input, while simultaneously facilitating a larger set of features to be fed into the transformer module. Further, the transformer module processes this input to attentively capture the characteristic information from the strokes at each time step. We now discuss these modules in detail.

4.3.1 Auto-encoder

The input representation in the form of 3-point format (3) works for RNNs [14]. However, the number of input features to the transformer must be divisible by the number of attention heads in each transformer block. Hence, a mechanism was required to project the input features to a higher latent dimension. Initial attempts of using a *convolution layer* and a *dense layer* were successful only to a limited extent and failed to capture the exact representation of the input features. To mitigate this issue, an auto-encoder module is proposed. The auto-encoder performs two major functions, (1) extract latent vector ($\mathcal{Z}_v$) from the input and (2) reshape the input dimension to feed a larger context vector to the transformer module for effective processing of the input data.

The input $\mathcal{S}_v$ is not directly fed into the auto-encoder. Instead, it is decomposed as $\mathcal{S}_v = \{\mathcal{S}_v^{\Delta}, \mathcal{S}_v^{p}\}$, where $\mathcal{S}_v^{\Delta} = (\Delta x_i, \Delta y_i)$ and $\mathcal{S}_v^{p} = p_i$. We only feed the offsets, $\mathcal{S}_v^{\Delta} \in \mathbb{R}^{2\times n}$, to obtain latent vector, $\mathcal{Z}_v \in \mathbb{R}^{127\times n}$, with a sequence length of n. Since $\mathcal{S}_v^{p}$ is a binary variable, it is not processed by the auto-encoder. This ensures that the pen-state information

is not processed with the offsets, because these two sets of information are independent of each other. Hence, we concatenate the $\mathcal{S}_v^p$ with the latent vector obtained, as discussed below.

Figure 3 shows the auto-encoder module, which consists of *time-distributed dense layers* with dimensions (2, 32, 64, 127, 64, 32, 2). The middle layer, dubbed as the bottleneck layer is extracted as the latent vector $\mathcal{Z}_v$. Each layer uses ReLU activation. Further, latent vector is concatenated with the pen-state information to form the input to the tranformer module, $\mathcal{S}_v^T \in \mathbb{R}^{128\times n}$, defined as:

$$\mathcal{S}_v^T = \{\mathcal{Z}_v, \mathcal{S}_v^p\} \tag{4}$$

The encoder is trained independently to learn the input and output mapping between the offsets, $\mathcal{S}_v^\Delta$, and its extended representation in the form latent vector $\mathcal{Z}_v$. The processing of extended input, $\mathcal{S}_v^T$, by the transformer follows.

4.3.2 Transformer

Transformers [40] have been employed in the proposed approach to utilize the parallel processing capabilities while enabling real-time sketch recognition. In the past, Transformers have been primarily used for language modeling. We proposed significant changes to the original transformer architecture to adapt it to the sketch classification task. First, the embedding layer is removed since semantic information is not required to be preserved by our model as opposed to the domain of Natural Language Processing.

Further, to preserve the temporal nature of our data, we use *coordinate embedding* [10]. In this embedding, a tuple with (position, time) is computed, instead of calculating just the position. $\mathcal{P}^t \in \mathbb{R}^{x\times d}$ is obtained by generating a sinusoidal position embedding with $1 \leq i \leq h$ positions, and $1 \leq t \leq T$ time steps for each vector-dimension $1 \leq j \leq d$:

$$\mathcal{P}^t_{i,2j} = sin\left(\frac{i}{10000^{2j/d}}\right) + sin\left(\frac{t}{10000^{2j/d}}\right) \tag{5}$$

$$\mathcal{P}^t_{i,2j+1} = cos\left(\frac{i}{10000^{2j/d}}\right) + cos\left(\frac{t}{10000^{2j/d}}\right) \tag{6}$$

In the proposed network, we use 10 identical transformer blocks connected via skip-connections (for better information retention), as shown in Fig. 3. Each block consists of multi-head self-attention with $m = 8$ heads, which is followed by addition of the residual and normalization. The attention mechanism allows the model to learn characteristic information from each sketch category on an abstract level. This ability mitigates the inherent problem of high intra-class variation and low inter-class variations present in the human-drawn sketches. In addition, identical feed-forward fully-connected layer is applied at each position separately, which use varying set of parameters at each layer. The feed-forward ($\mathcal{F}$) layer is defined as:

$$\mathcal{F}(x) = ReLU(xW_1 + b_1)W_2 + b_2 \tag{7}$$

The encoded input $\mathcal{P}$ is fed into the Transformer blocks to obtain the point-wise attention-maps $\mathcal{A}_k \forall k \in \{1, 2, ..., m\}$. The custom *Extract layer* is used to pick the features (ext_{out}) corresponding to the first timestamp of the transformer output. Further, ext_{out} is passed through a dense layer with a softmax activation function to obtain class probabilities $\Psi = \{\psi_1, \psi_2, ..., \psi_c\}$, where c represents the number of classes. Finally, dense layers with softmax activation facilitate the model to return class probabilities. The following subsection specifies the model training configurations.

4.4 Implementation details

For training the proposed architecture, the two modules are trained separately. Firstly, we train the auto-encoder with a batch size of 1024, for 50 epochs and an initial learning rate of 0.0001. Next, we freeze the weights of the auto-encoder and train the complete network to learn the transformer parameters with a batch size of 32. Both modules share the same training configurations. Early stopping configured with $\delta_e = 0.00001$ and patience$= 10$ is used. Also, Learning Rate Scheduler defines the learning rate (lr_i) for i^{th} epoch as:

$$lr_i = ((lr_{beg} - lr_{min}) * \delta_{lr}^{i}) + lr_{min} \tag{8}$$

where $\delta_{lr} = 0.9$ is the decay rate, $lr_{beg} = 0.0001$ represents the initial learning rate, and $lr_{min} = 0.00001$ is the minimum learning rate. Finally we use Adam Optimizer to reduce the cross-entropy loss defined as:

$$\mathcal{L} = \sum_{j=1}^{c} \psi_j \times log(\psi_j) \tag{9}$$

All computations were carried out on a Linux workstation with Intel Xeon Silver 4110 CPU, having 128 GB RAM and a TITAN RTX GPU with 24 GB memory. A stratified 5-fold grid search was carried out on the hyper-parameters. The following section analyses the proposed method, both quantitatively and qualitatively.

5 Experiments

5.1 Dataset and evaluation metrics

The performance of the proposed Transformer-based network is evaluated using the benchmark Quick Draw dataset [14], which consists of over 50 million sketches scattered across 345 categories. This dataset was generated by collecting real-time sketches by players of the online game, *The Quick, Draw!*. While acquisition, the players were given 20 seconds to draw a sketch. Hence, generating abstract level sketches.

To validate the performance of the proposed and state-of-the-art methods, recognition accuracy is reported on the test-set. A train-test split of 80-20% is performed with 50,000 samples per class for training, and 10,000 samples per class during testing. For exhaustive comparison, we formulate three subsets of the Quick Draw dataset, having 5, 20, and 50 classes, respectively. For each subset, the respective number of classes are picked up randomly, with 10-fold cross-validation. Further, we compute the top-k accuracy for $k = \{1, 5, 10\}$, i.e., top-1, top-5, and top-10 recognition rates as well. Hence, we report the average accuracy for each subset. Trends in recognition accuracy on these datasets reveal the dependence of accuracy on the number of classes. The following subsection compares the accuracy performance with baselines. [1]

5.2 Comparative evaluation

We compared the proposed AttentiveNet with three types of approaches, (1) traditional classifiers, (2) CNN-based approaches, and (3) RNN-based approaches. Table 1 reports

[1] Official TensorFlow implementation of RNN for QuickDraw

Table 1 Comparative evaluation of recognition accuracy on the Quick Draw dataset, with (a) 5 classes, (b) 20 classes, (c) 50 classes

Method	5	20	50
HOG-SVM [11]	75.21%	66.79%	63.22%
Fisher-Vectors [28]	79.53%	75.80%	72.90%
AlexNet [17]	77.18%	75.22%	73.06%
Sketch-a-Net v2 [46]	94.78%	88.64%	85.19%
Resnet50-CNN [15]	**96.47%**	90.06%	86.20%
RNN-Tensorflow[1]	95.58%	87.33%	83.98%
Ours	96.21%	**90.31%**	**88.72%**

Values in **bold** depict the best accuracy for each subset

the recognition accuracy. We implemented and trained various approaches to compare the performance of the proposed approach. To ensure fairness, training is performed on the same subsets for all techniques. AttentiveNet outperforms the state-of-the-art approaches for 20 and 50 classes. However, Resnet50-CNN [46] performs marginally better than AttentiveNet with a gain of 0.26% for dataset with 5 classes.

As the number of classes increases from 5 to 20, most of the approaches show a significant decrease inaccuracy. The drop in accuracy is observed to be 11.20% for HOG-SVM [11], 4.69% for Fisher-Vectors [46], 2.53% for AlexNet [17], 6.48% for Sketch-a-Net v2 [46], 6.22% for Fisher-Vectors [46], 8.63% for RNN-Tensorflow[1], 6.13% for AttentiveNet. Further increasing the number of classes from 20 to 50, drop-in accuracy was observed to be 5.34% for HOG-SVM [11], 3.8% for Fisher-Vectors [46], 2.88% for AlexNet [17], 3.45% for Sketch-a-Net v2 [46], 4.22% for Fisher-Vectors [46], 3.83% for RNN-Tensorflow[1], 1.17% for AttentiveNet. For the proposed approach, increasing the number of classes observes a lower drop in accuracy, compared to other approaches. Due to limited resource availability, analysis over the complete dataset could not be performed. To account for this, we randomly selected c classes, where $c \in \{5, 20, 50\}$, and report the averages over 10-fold cross-validation.

In addition, we compare the proposed approach with the state of the art approaches over all 345 categories in the dataset. For this, we have compared the top-1, top-5, and top-10

Table 2 Comparative evaluation of recognition accuracy on the *Quick, Draw!* dataset, with (a) Top-1, (b) Top-5, (c) Top-10 accuracy

	Method	Top-1	Top-5	Top-10
Traditional Classifiers	HOG-SVM [11]	50.25 ± 2.05%	62.51 ± 0.64%	64.22 ± 1.64%
	Fisher-Vectors [28]	56.11 ± 1.34%	74.28 ± 1.00%	79.19 ± 0.33%
CNN-based Classifiers	AlexNet [17]	63.25 ± 0.95%	80.83 ± 2.07%	85.02 ± 1.36%
	ResNet50-CNN [15]	78.86 ± 2.18%	91.10 ± 0.86%	94.14 ± 0.97%
	Sketch-a-Net v2 [46]	74.84 ± 1.01%	87.19 ± 1.13%	91.25 ± 2.00%
RNN-based Classifiers	RNN-Tensorflow[1]	75.59 ± 2.04%	90.14 ± 1.27%	**94.20 ± 2.08%**
	Ours	**79.35 ± 0.96%**	**91.30 ± 0.84%**	93.95 ± 1.38%

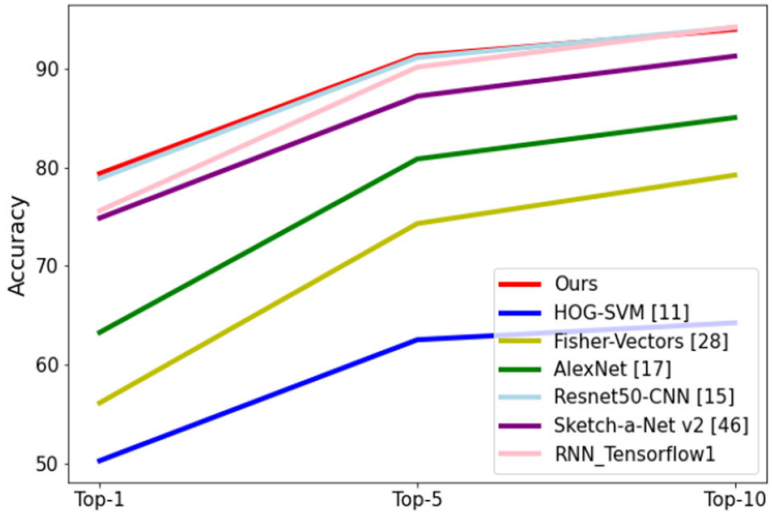


Fig. 4 Relative classification accuracy of the baselines and their trends with respect to Top-k accuracies

recognition accuracy. Table 2 reports the accuracy performance, and Fig. 4 visualizes these results. Quantitative evaluation confirms the effectiveness of representing sketches as vector images. Moreover, this representation facilitates real-time recognition of sketches. Ablation studies follow in the next subsection.

6 Discussions

6.1 Ablation analysis

6.1.1 Model configurations

The proposed approach majorly constitutes two modules, the auto-encoder, and the transformer blocks. In this section, we evaluate the effectiveness of these propositions. For this, we compare the recognition accuracy of different models by tweaking four model configurations. First, is the choice between Transformers [40] and Universal Transformers [10]. Since we adopt these sequential models given their success in language modeling, universal transformers provide a lucrative alternative against transformers. Second, we vary the number of transformer blocks. Third, we add skip-connections between each transformer block to exploit the power of residual learning. Finally, experiments that corroborate the effectiveness of the proposed auto-encoder is performed. Table 3 reports the observed performance.

For most of the language modeling tasks, Universal transformers perform favorably against the original transformer. However, the proposed architecture outperforms universal transformers. This is because universal transformer blocks share weights. Consequently, reducing the number of parameters. This leads to its inability to capture complex intraclass and interclass variations in sketches. Transformers with 5 blocks perform significantly better than its counterpart with universal transformers for all three subsets, with a gain of 0.1%, 3.07%, and 4.84%, respectively. Further, using the Adaptive computation time (ACT) mechanism with universal transformers still performs inferior to transformers, due to the early halting.

Varying the number of transformer blocks alters the number of parameters. The proposed network with 10 blocks performs significantly better than the 5-block variant. This enables the model to address the complexity of sketch recognition. Increasing the number

Table 3 Ablation Analysis and comparative results for different models on the QuickDraw [14] Dataset in terms of Accuracy reported on three subsets with 5, 20 and 50 classes respectively

Model Configurations				Recognition Accuracy		
T/UT	No. of Blocks	Skip-connections	Auto-encoder	5	20	50
T	5	✗	✗	93.66%	84.44%	78.90%
T	5	✓	✗	94.56%	86.38%	81.60%
T	10	✗	✗	92.91%	85.32%	75.03%
T	10	✓	✗	94.60%	86.55%	77.44%
UT	5	✗	✓	93.76%	87.11%	82.87%
UT + ACT	5	✓	✓	92.73%	83.14%	79.55%
T	10	✗	✓	95.71%	89.13%	87.21%
T	10	✓	✓	**96.21%**	**90.31%**	**88.72%**

T: Transformer UT: Universal Transformer ACT: Adaptive Computation Time. Values in **bold** depict the highest accuracy achieved in each subset

of blocks further provided an insignificant increase in accuracy with a relatively large training overhead, such as time and space complexity. Further, inclusion of residual learning observes improvement in model performances for CNNs [15]. In our network, we add skip-connections between the transformer blocks. This is because gradients often get lost when passed through multiple layers. In order to address this issue, skip-connections ensure that lower layers receive information, even if the network gets deeper.

The proposed architecture with auto-encoder displays significantly superior recognition accuracy, as compared to other ablations. This is due to the fact that the offsets $(\Delta x_i, \Delta y_i)$, contain fairly low amount of information for computationally intensive architectures like transformers. Hence, mapping the offsets to a larger latent vector enables the transformer to tweak the attention heads with a lot more available input data.

6.1.2 Attention heatmaps

The proposed method introduces transformers to exploit the sequence of strokes to draw attention to parts of sketches. Figure 5 illustrates the attention maps for various classes. The attention softmax scores corresponding to the first input of the first head is used for the attention-based ablation study. This particular choice is made taking into consideration the fact that the sketches are being recognised by extracting the logits from output corresponding to the first timestamp of the input. For the *bat* class, attention on both the wings are equally focused. This is analogous to how humans recognize bats through their peculiar wing shape. For the *cello* class, the model directs its attention towards the typical round shape. This trend is also observed for the *fan* class, where central spikes and pedestal are aptly recognized by the proposed model for recognition. It is interesting to note that for the *star* class, equally high attention is given to the complete structure. Perhaps, structures like stars that are symmetric in nature and constitute a distinct shape are identified based on the overall view of the sketches. Similarly, certain parts of the sketch are more important in recognizing an object; this is supported by the attention heatmaps presented.

The role of auto-encoders is discussed in the previous sub-subsection. In Fig 6, we compare the attention maps generated with (a) network containing the auto-encoder module, and (b) network without an auto-encoder module. The auto-encoder enhances the attention

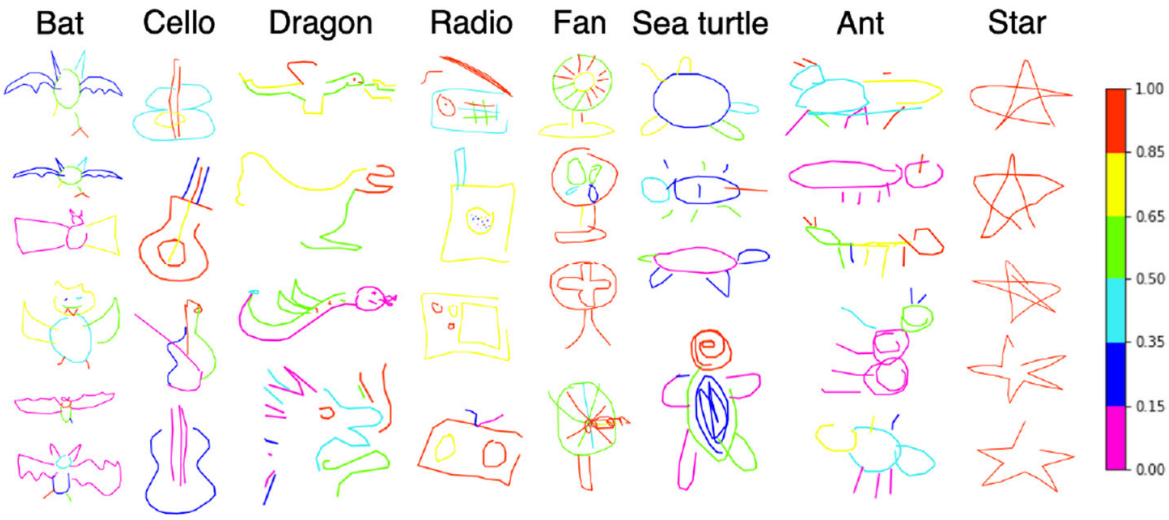


Fig. 5 Visualization of the attention maps of a few classes in the QuickDraw [14] dataset depicting the relative importance of the points in the strokes as observed by the model while inference

magnitude over the right areas for most of the classes. While the absence of auto-encoder provides fewer patches with high attention. This supports the hypothesis that the autoencoder module effectively extracts information and hence plays a pivotal role in the proposed AttentiveNet architecture. Figure 7 visualizes a sample User-Interface performance using the proposed approach, with inference in real-time.

6.2 Computational complexity analysis

In this subsection, we discuss the computational cost of the proposed approach in comparison to existing techniques. Conventional and CNN-based techniques can only perform classification, once the sketch has been rendered completely. On the other hand,

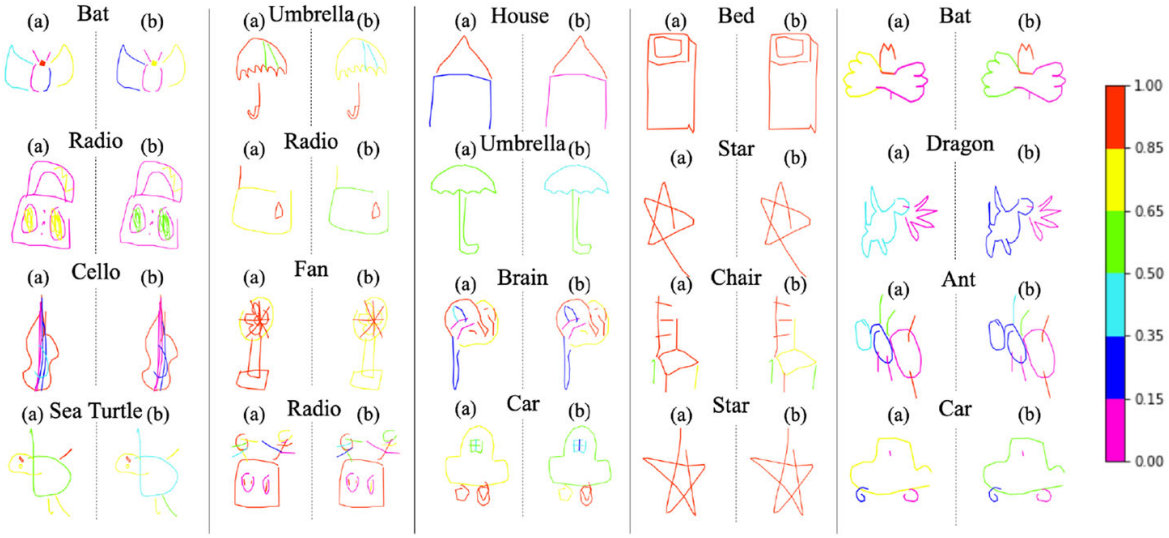


Fig. 6 Attention-map based ablation study using a Transformer (10 Blocks) + Auto Encoder b Transformer (10 Blocks)

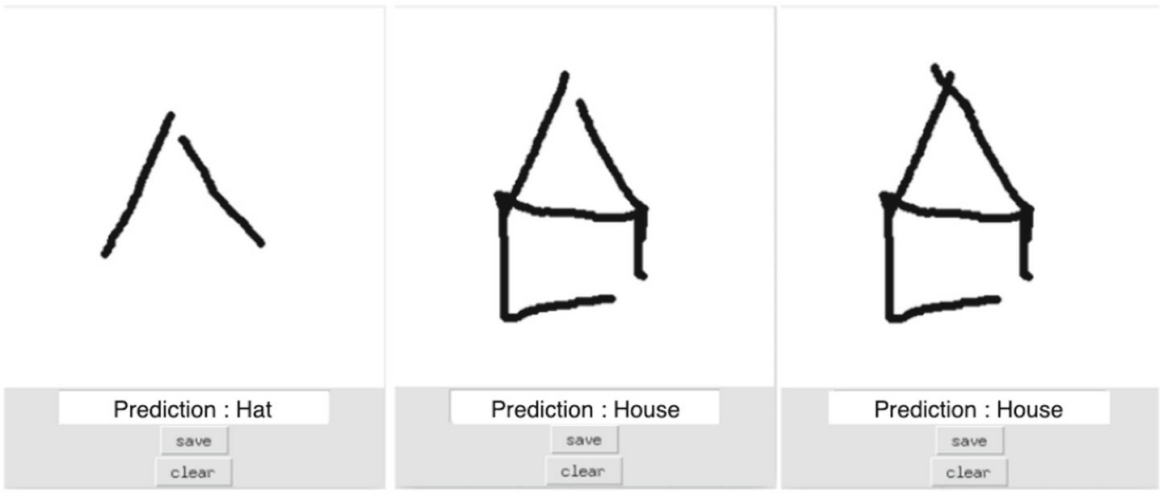


Fig. 7 Custom user interface designed to facilitate real-time inference. Progressive inference is performed on the sketch to accurately predict the object being drawn on the interface

sequence-based models have the ability to classify partially rendered drawings as well – facilitating real-time inference. Therefore, the use of sequence-based classifiers over the conventional and CNN-based sketch classifiers provides a significant advantage in terms of the inference times. This extended functionality of sequence-based classifiers establishes dominance over these conventional and CNN-based approaches in terms of computational efficiency.

Amongst the sequence-based classifiers, the Transformer-based architectures outperform the RNN-based techniques in terms of computational costs involved. This is mainly due to ability of Transformers to utilize parallel processing. To process an input with n time-steps, the Transformers require O(1) sequential operations, while RNN-based architectures require O(n) sequential operations [40]. We compare the average inference times per sample, averaged across 10,000 samples, for the difference sequence-based architectures, reported in Table 4. A noticeable improvement in inference times can clearly be observed for the proposed Transformer-based architecture as compared to existing RNN-based implementation i.e. RNN-Tensorflow[1].

7 Conclusion and future directions

In this work, we proposed a novel Transformer-based architecture, dubbed as AttentiveNet, for recognizing sketches using vector images. The autoencoder extracts information from the input data in the form of a latent vector. Moreover, this renders the input dimension compatibility between the input data and transformer-blocks. The proposed approach enables real-time inference capabilities, which is an essential functionality in any practical setup. Moreover, this functionality is incorporated at reduced computational cost, when

Table 4 Average per-sample inference times for the various architectures for the sequence-based models

Model	Inference Time
RNN-Tensorflow[1]	112.29 ± 10 ms
Transformer	80.38 ± 6 ms

compared with existing approaches. Further, transformer-blocks enable the model to focus on characteristic information from each sketch. The proposed approach achieves favorable recognition accuracy when compared with state-of-the-art approaches. Ablation analysis, along with attention heatmaps reveal that the proposed approach makes prediction based on certain parts of sketches, that are intuitive in nature.

In future, transformer-based architectures can be adapted to address problems in the domain of computer vision applications. Potential applications may involve tasks such as sketch retrieval or sketch synthesis, which require descriptive line-drawing features, could be exciting to explore. The proposed approach opens avenues to improve the cognition abilities of computers to process data in sequential nature. Alternate realms of computer vision that utilize such sequential information include human action recognition for complex motion sequences like, dance, martial arts etc.

References

1. Abualigah L (2020) Multi-verse optimizer algorithm: a comprehensive survey of its results, variants, and applications. Neural Comput Applic pp 1–21
2. Abualigah L, Diabat A (2020) A novel hybrid antlion optimization algorithm for multi-objective task scheduling problems in cloud computing environments. Clust Comput pp 1–19
3. Abualigah LM, Khader AT (2017) Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering. J Supercomput 73(11):4773–4795
4. Abualigah LM, Khader AT, Hanandeh ES (2018) A combination of objective functions and hybrid krill herd algorithm for text document clustering analysis. Eng Appl Artif Intell 73:111–125
5. Abualigah LM, Khader AT, Hanandeh ES (2018) Hybrid clustering analysis using improved krill herd algorithm. Appl Intell 48(11):4047–4071. https://doi.org/10.1007/s10489-018-1190-6
6. Abualigah LM, Khader AT, Hanandeh ES (2018) A new feature selection method to improve the document clustering using particle swarm optimization algorithm. J Comput Sci 25:456–466
7. Abualigah LMQ (2018) Feature selection and enhanced krill herd algorithm for text document clustering, 1st edn, Springer Publishing Company, Incorporated
8. Arandjelović R, Sezgin TM (2011) Sketch recognition by fusion of temporal and image-based features. Pattern Recogn 44(6):1225–1234
9. Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. arXiv:1409.0473
10. Dehghani M, Gouws S, Vinyals O, Uszkoreit J, Kaiser Ł (2018) Universal transformers. arXiv:1807.03819
11. Eitz M, Hays J, Alexa M (2012) How do humans sketch objects? ACM Trans Graph 31(4):44–1
12. Eitz M, Richter R, Boubekeur T, Hildebrand K, Alexa M (2012) Sketch-based shape retrieval. ACM Trans Graph (TOG) 31(4):31
13. Graves A (2013) Generating sequences with recurrent neural networks. arXiv:1308.0850
14. Ha D, Eck D (2017) A neural representation of sketch drawings. arXiv:1704.03477
15. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
16. Huang Z, Fu H, Lau RW (2014) Data-driven segmentation and labeling of freehand sketches. ACM Trans Graph (TOG) 33(6):175
17. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
18. LaViola JJ Jr, Zeleznik RC (2004) Mathpad 2: a system for the creation and exploration of mathematical sketches. ACM Trans Graph (TOG) 23(3):432–440
19. Li K, Pang K, Song J, Song YZ, Xiang T, Hospedales TM, Zhang H (2018) Universal sketch perceptual grouping. In: Proceedings of the european conference on computer vision (ECCV), pp 582–597
20. Li L, Fu H, Tai CL (2018) Fast sketch segmentation and labeling with deep learning. IEEE Comput Graph Appl 39(2):38–51
21. Li Y, Hospedales TM, Song YZ, Gong S (2015) Free-hand sketch recognition by multi-kernel feature learning. Comput Vis Image Underst 137:1–11

22. Liu P, Yu H, Cang S (2019) Adaptive neural network tracking control for underactuated systems with matched and mismatched disturbances. Nonlinear Dynam 98(2):1447–1464
23. Lu T, Tai CL, Su F, Cai S (2005) A new recognition model for electronic architectural drawings. Comput Aided Des 37(10):1053–1069
24. Ouyang TY, Davis R (2011) Chemink: a natural real-time recognition system for chemical drawings. In: Proceedings of the 16th international conference on Intelligent user interfaces, ACM, pp 267–276
25. Sangkloy P, Burnell N, Ham C, Hays J (2016) The sketchy database: learning to retrieve badly drawn bunnies. ACM Trans Graph (TOG) 35(4):119
26. Sarvadevabhatla RK, Babu RV (2015) Freehand sketch recognition using deep features. arXiv:1502.00254
27. Sarvadevabhatla RK, Surya S, Mittal T, Babu RV (2018) Game of sketches: deep recurrent models of pictionary-style word guessing. In: Thirty-second AAAI conference on artificial intelligence
28. Schneider RG, Tuytelaars T (2014) Sketch classification and classification-driven analysis using fisher vectors. ACM Trans Graph (TOG) 33(6):174
29. Seddati O, Dupont S, Mahmoudi S (2015) Deepsketch: deep convolutional neural networks for sketch recognition and similarity search. In: 2015 13th international workshop on content-based multimedia indexing (CBMI), IEEE, pp 1–6
30. Seddati O, Dupont S, Mahmoudi S (2017) Deepsketch 3. Multimed Tools Appl 76(21):22,333–22,359
31. Sert M, Boyacı E (2019) Sketch recognition using transfer learning. Multimed Tools Appl 78(12):17,095–17,112
32. Sezgin TM, Davis R (2008) Sketch recognition in interspersed drawings using time-based graphical models. Comput Graph 32(5):500–510
33. Song J, Pang K, Song YZ, Xiang T, Hospedales TM (2018) Learning to sketch with shortcut cycle consistency. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 801–810
34. Sun L, Zhao C, Yan Z, Liu P, Duckett T, Stolkin R (2018) A novel weakly-supervised approach for rgb-d-based nuclear waste object detection. IEEE Sensors J 19(9):3487–3500
35. Sun Z, Wang C, Zhang L, Zhang L (2012) Free hand-drawn sketch segmentation. In: European conference on computer vision. Springer, New York, pp 626–639
36. Sutherland IE (1964) Sketchpad a man-machine graphical communication system. Simulation 2(5) R–3
37. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9
38. Tang Z, Yu H, Lu C, Liu P, Jin X (2019) Single-trial classification of different movements on one arm based on erd/ers and corticomuscular coherence. IEEE Access 7:128,185–128,197
39. Tang ZC, Li C, Wu JF, Liu PC, Cheng SW (2019) Classification of eeg-based single-trial motor imagery tasks using a b-csp method for bci. Front Inform Technol Electron Eng 20(8):1087–1098
40. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: Advances in neural information processing systems, pp 5998–6008
41. Wang F, Kang L, Li Y (2015) Sketch-based 3d shape retrieval using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1875–1883
42. Wang X, Chen X, Zha Z (2018) Sketchpointnet: a compact network for robust sketch recognition. In: 2018 25th IEEE international conference on image processing (ICIP), IEEE, pp 2994–2998
43. Xu P, Huang Y, Yuan T, Pang K, Song YZ, Xiang T, Hospedales TM, Ma Z, Guo J (2018) Sketchmate: deep hashing for million-scale human sketch retrieval. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 8090–8098
44. Yang Y, Hospedales TM (2015) Deep neural networks for sketch recognition. arXiv:1501.07873 1(2), 3
45. Yanık E, Sezgin TM (2015) Active learning for sketch recognition. Comput Graph 52:93–105
46. Yu Q, Yang Y, Liu F, Song YZ, Xiang T, Hospedales TM (2017) Sketch-a-net: a deep neural network that beats humans. Int J Comput Vision 122(3):411–425
47. Zhang H, Liu S, Zhang C, Ren W, Wang R, Cao X (2016) Sketchnet: sketch classification with web images. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1105–1113
48. Zhang J, Chen Y, Li L, Fu H, Tai CL (2018) Context-based sketch classification. In: Proceedings of the joint symposium on computational aesthetics and sketch-based interfaces and modeling and non-photorealistic animation and rendering, ACM, p 3
49. Zhang X, Huang Y, Zou Q, Pei Y, Zhang R, Wang S (2019) A hybrid convolutional neural network for sketch recognition. Pattern Recognition Letters

50. Zhao P, Liu Y, Lu Y, Xu B (2019) A sketch recognition method based on transfer deep learning with the fusion of multi-granular sketches. Multimed Tools Appl 78(24):35,179–35,193
51. Zou C, Yu Q, Du R, Mo H, Song YZ, Xiang T, Gao C, Chen B, Zhang H (2018) Sketchyscene: richly-annotated scene sketches. In: Proceedings of the european conference on computer vision (ECCV), pp 421–436

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

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/TIA.2020.3030873, IEEE Transactions on Industry Applications

Small-Signal Modeling of Mutually Coupled Switched Reluctance Motor

Siddharth Mehta, *Student Member, IEEE,* Prerit Pramod, *Senior Member, IEEE,* Iqbal Husain, *Fellow, IEEE,* and Md Ashfanoor Kabir, *Member, IEEE*

***Abstract*—A small-signal model based on the Net Flux Method is developed for mutually coupled switched reluctance machines (MCSRMs) to facilitate the design and analysis of the current controller for the drive. The underlying methodology of using net flux allows MCSRM modeling on a per-phase basis which reduces the time involved in characterizing the motor and the number of machine model parameters. A comprehensive theory of modeling the machine and extracting the machine parameters with analytical derivations is provided. Furthermore, a complete design procedure and analysis for a proportional-integrator controller is presented. The derived MCSRM small-signal model is useful to evaluate the controller's current tracking performance, disturbance rejection capability, and response features over a wide speed range. Simulation analysis and experimental results for a 3 phase, 150W, full-pitched MCSRM drive system are provided both for small-signal model validation and current controller performance evaluation.**

***Index Terms*—Mutually coupled SRM, mutual flux, small-signal model, linearization, current controller**

I. Introduction

MUTUALLY coupled switched reluctance motor (MCSRM) is an emerging alternative to conventional switched reluctance motor (SRM) that holds the same durability, robustness, and simplicity in the design and manufacturing as that of an SRM [1], [2]. With improved utilization of the machine from an electromagnetic point of view, MCSRMs can have higher torque density as compared to SRMs [3]. Depending upon the needs of the application, these machines can also be operated using a conventional two-level voltage source inverter (VSI) similar to any other ac machine [4]–[6].

For high-performance applications, SRM drives have difficulty in making its case because of the unconventional power-converter coupled with high torque ripple and structural vibrations. With the balanced current excitation, MCSRMs have the advantages of using conventional VSI along with minimal changes in the winding connections and layout. MCSRMs have both self and mutual flux varying with respect to current and rotor position. Both self and mutual flux linkages in MCSRM can contribute to torque production depending upon their winding configuration and rotor and stator designs [2], [3], [7]. Therefore, simple but accurate modeling of MCSRMs is a challenging problem.

In literature, a significant amount of research focusing on modeling dedicated to MCSRMs exists. Modeling based on a lookup table (LUT) predicting flux and torque has been presented in [7]–[9]; these models are developed with the objective of predicting the torque of MCSRMs with accuracy close to the finite element (FE) models. A methodology to calculate the phase currents in MCSRMs based on the torque commands has been developed in [10] based on the modeling approach developed by Mecrow et al. [8]. A LUT based model based on the synchronous reference frame theory has been developed and presented by Dong et al. [11], which also focuses on developing machine models to predict the torque performance of MCSRMs. Furthermore, Uddin et al. and Kokernak et al. have developed analytical models of the MCSRM based on the geometrical parameters of the motor [12], [13]; these models can be used for designing or sizing the motor.

There are several linearized models available for SRMs that can be used to model MCSRMs and develop controllers for MCSRMs [14], [15]. In most of these models, the mutual coupling is ignored while modeling SRMs; in some cases where the models are extended to include mutual coupling (for MCSRMs), it becomes complicated as mutual flux in itself is a non-linear parameter which depends upon rotor position and phase currents. In addition, estimation of mutual flux or mutual inductance in real time is time consuming and not accurate. Since the MCSRM plant is time-variant, the drive controller design becomes complicated. To the authors' best knowledge, the literature lacks a simplified and linearized MCSRM model that can be utilized to design a current controller relating to the machine parameters.

In this work, a small-signal model of MCSRM that is linearized around an operating point is developed based on the Net Flux Method (NFM) [7], [16]. The design and analysis of a current-controlled MCSRM drive based on the developed linearized model are presented next. The MCSRM used in this work has full-pitched winding and is excited using bipolar balanced trapezoidal currents. Unlike SRMs, the currents are balanced because the winding of the motor is connected in Y, hence, conventional two-level VSI has been used to drive this machine. However, this model is applicable to both unipolar and bipolar MCSRMs. The paper has been configured in the following way; the conventional modeling technique and its limitations are covered in Section II and the non-linear

This work was accomplished through the support of the ABB Distinguished Professorship endowment and Nexteer Automotive Corporation for research on electric machines and drives at the FREEDM Center.

S. Mehta and I. Husain are with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27606 USA (e-mail:smehta5@ncsu.edu; ihusain2@ncsu.edu).

P. Pramod is with the Advanced Systems and Software Engineering group of Nexteer Automotive Corporation, Saginaw, Michigan, USA (email: prerit.pramod@nexteer.com).

M. A. Kabir is with Turntide Technologies (e-mail: mkabir@ncsu.edu).

0093-9994 (c) 2020 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 November 06,2020 at 04:57:52 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/TIA.2020.3030873, IEEE Transactions on Industry Applications

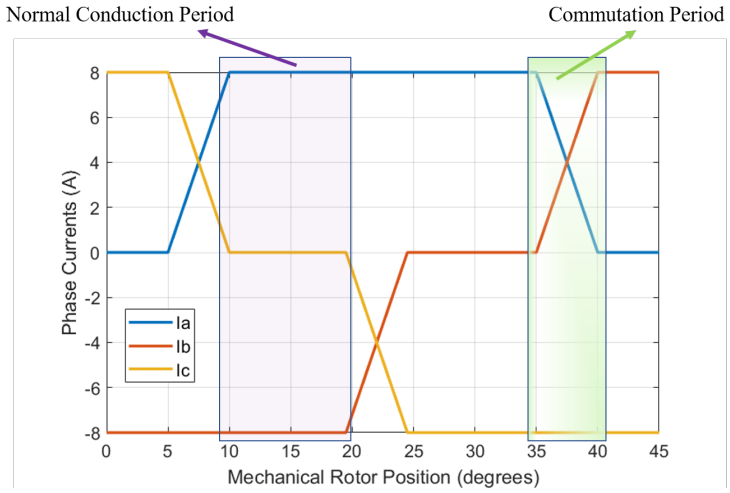


Fig. 1. Phase currents of MCSRM illustrating different conduction periods.

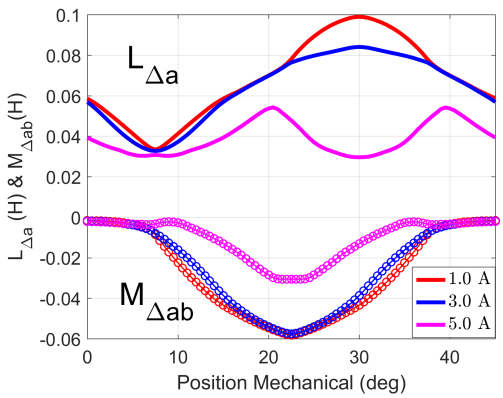


Fig. 2. Incremental self and mutual inductance for a 12/8 fractional slot MCSRM [7].

machine model based on NFM is illustrated in Section III. The small-signal model is presented in Section IV. Then the model is utilized to develop a current controller in Section V. A case study is presented for a 150 W full-pitched MCSRM (FP-MCSRM) and results are validated with experiments in Section VI.

II. Conventional Modeling of Mutually Coupled Switched Reluctance Motor

Unlike SRMs, mutual coupling between phases plays a significant role in torque production for MCSRMs. Hence, mutual flux cannot be ignored while modeling or designing a controller for this machine. Mutual flux for MCSRM also varies with rotor position and currents of multiple phases in a similar way as the self flux does. The self flux is mainly affected by the magnitude of the current of its own phase. However, the currents of other phases also affect the self flux because of variation in permeability of the material. Hence, in MCSRMs, there are two quantities (self and mutual flux) varying, which are strong functions of rotor position and magnitude of multiple phase currents.

Let us consider a three phase MCSRM, considering two phases in operation, the net flux for phase A, λ_a, and phase B, λ_b, are defined as follows

$$\lambda_a(i_a, i_b, \theta_e) = \lambda_{aa}(i_a, \theta_e) + \lambda_{ab}(i_a, i_b, \theta_e) \quad (1)$$

$$\lambda_b(i_a, i_b, \theta_e) = \lambda_{bb}(i_b, \theta_e) + \lambda_{ab}(i_a, i_b, \theta_e) \quad (2)$$

where i_a and i_b are the phase currents, λ_{aa} and λ_{bb} are self flux-linkages of phase A and phase B, and λ_{ab} is the mutual flux-linkage, respectively, and θ_e is the electrical rotor position. The phase voltage relationship with the machine parameters when three phases are conducting, that is during the commutation period as shown in Fig. 1, is expressed for a single phase in (3) and for all three phases is shown in (4).

$$v_{aN} = i_a R + \frac{\partial}{\partial i_a}(\lambda_{aa} + \lambda_{ab} + \lambda_{ac})\frac{di_a}{dt} + \frac{\partial \lambda_{ab}}{\partial i_b}\frac{di_b}{dt} + \frac{\partial \lambda_{ac}}{\partial i_c}\frac{di_c}{dt} + \omega_e\left(\frac{\partial \lambda_{aa}}{\partial \theta_e} + \frac{\partial \lambda_{ab}}{\partial \theta_e} + \frac{\partial \lambda_{ac}}{\partial \theta_e}\right) \quad (3)$$

where v_{aN}, v_{bN}, and v_{cN} are the phase voltages, ω_e is the electrical speed, and R is the phase resistance. In case of SRMs, the incremental mutual inductance, $M_{\Delta ab} = \partial \lambda_{ab}/\partial i_a$, can be neglected; however, the incremental mutual inductance of MCSRMs are large and comparable to incremental self inductances, as seen in Fig. 2, hence can not be neglected. There are several underlying drawbacks of this model which need to be overcome so that the modeling of the machine, estimation of machine parameters and the controller design is expedited. These issues are:

- The conventional modeling technique consists of various multiple mutual flux related terms whose accurate estimation is tedious [17]. The more the numbers of terms, the higher are the possibilities of reduced modeling accuracy.
- Since λ_{aa}, λ_{bb} and λ_{ab} are time-varying functions of multiple phase currents and rotor position, parameter extraction using either finite element analysis (FEA) or experiments is laborious.

The challenges call for a new modeling strategy to reduce the computational complexity and make it easier to develop a linearized model, to simplify the estimation of machine parameters and in totality aid in the faster development of current controllers.

III. Machine Model Based on the Net Flux Method

The operation of MCSRMs can be divided into two periods, namely, the normal conduction period and the commutation period. During the normal conduction period, only 2 phases conduct, as shown in Fig. 1. The flux linkage characteristics of each phase is derived based on the concept of net flux method as reported in [7]. In the NFM, two phases of the motor are excited with fixed and equal and opposite magnitude of currents and the net flux of each phase is recorded from FEA or prototype machine.

A. Assumptions for NFM

For the controller development purpose, a simplified model is derived in this paper to capture the major trend of the machine's characteristics, but with slightly reduced accuracy. The following assumptions are made in the process:

0093-9994 (c) 2020 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 November 06,2020 at 04:57:52 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/TIA.2020.3030873, IEEE Transactions on Industry Applications

$$\begin{bmatrix} v_{an} \\ v_{bn} \\ v_{cn} \end{bmatrix} = R \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} + \omega_e \begin{bmatrix} \frac{\partial \lambda_{aa}}{\partial \theta_e} + \frac{\partial \lambda_{ab}}{\partial \theta_e} + \frac{\partial \lambda_{ac}}{\partial \theta_e} \\ \frac{\partial \lambda_{bb}}{\partial \theta_e} + \frac{\partial \lambda_{ab}}{\partial \theta_e} + \frac{\partial \lambda_{bc}}{\partial \theta_e} \\ \frac{\partial \lambda_{cc}}{\partial \theta_e} + \frac{\partial \lambda_{ac}}{\partial \theta_e} + \frac{\partial \lambda_{bc}}{\partial \theta_e} \end{bmatrix} + \begin{bmatrix} \frac{\partial \lambda_{aa}}{\partial i_a} + \frac{\partial \lambda_{ab}}{\partial i_a} + \frac{\partial \lambda_{ac}}{\partial i_a} & \frac{\partial \lambda_{ab}}{\partial i_b} & \frac{\partial \lambda_{ac}}{\partial i_c} \\ \frac{\partial \lambda_{ab}}{\partial i_a} & \frac{\partial \lambda_{bb}}{\partial i_b} + \frac{\partial \lambda_{ab}}{\partial i_b} + \frac{\partial \lambda_{bc}}{\partial i_b} & \frac{\partial \lambda_{bc}}{\partial i_c} \\ \frac{\partial \lambda_{ac}}{\partial i_a} & \frac{\partial \lambda_{bc}}{\partial i_b} & \frac{\partial \lambda_{cc}}{\partial i_c} + \frac{\partial \lambda_{ac}}{\partial i_c} + \frac{\partial \lambda_{bc}}{\partial i_c} \end{bmatrix} \begin{bmatrix} \dot{i_a} \\ \dot{i_b} \\ \dot{i_c} \end{bmatrix} \tag{4}$$

- The NFM is executed only during the normal conduction period to develop the small-signal model. The deviation of flux characteristics during the commutation period from the normal conduction period will not play a significant role. This assumption is the advantage as well as the main limitation of the modeling technique.
- The motor parameters are symmetric and there are no variations because of manufacturing challenges.

B. Non-Linear Model

Considering two phases, phase A and phase B, in operation, the phase A voltage with respect to the neutral of the motor v_{aN}, is given as the per conventional modeling strategy by

$$\begin{aligned} v_{aN} = Ri_a + \omega_e \frac{\partial(\lambda_{aa} + \lambda_{ab})}{\partial \theta_e} + \frac{\partial \lambda_{aa}}{\partial i_a}\frac{di_a}{dt} \\ + \frac{\partial \lambda_{ab}}{\partial i_a}\frac{di_a}{dt} + \frac{\partial \lambda_{ab}}{\partial i_b}\frac{di_b}{dt} \end{aligned} \tag{5}$$

The above expression consists of self and mutual flux terms. The modeling method proposed in this paper is based on the net flux, and hence, the model deals with the net flux, λ_a, of each phase rather than the self flux, λ_{aa}, and mutual flux, λ_{ab}, terms.

In order to derive the relation between the phase net flux linkages and the phase voltages, the normal conduction period is considered such that $i_a = -i_b = I$. The phase voltage, v_{aN}, is given as per NFM by

$$v_{aN} = Ri_a + \omega_e \frac{\partial \lambda_a}{\partial \theta_e} + \frac{\partial \lambda_a}{\partial i_a}\frac{di_a}{dt} \tag{6}$$

Similarly, the phase voltage of phase B, v_{bN}, is given by

$$v_{bN} = Ri_b + \omega_e \frac{\partial \lambda_b}{\partial \theta_e} + \frac{\partial \lambda_b}{\partial i_b}\frac{di_b}{dt} \tag{7}$$

where the partial derivative of λ_x with respect to current and rotor position is termed as net incremental inductance $\psi_{\Delta x}$, and net incremental flux linkage, $\lambda_{\Delta x}$, respectively as illustrated in (8)-(9). Here x is the arbitrary phase.

$$\psi_{\Delta x} = \frac{\partial \lambda_x}{\partial i_x} \tag{8}$$

$$\lambda_{\Delta x} = \frac{\partial \lambda_x}{\partial \theta_e} \tag{9}$$

The net incremental inductance $\psi_{\Delta x}$ of the NFM accounts for both the self and mutual incremental inductance of the conventional modeling strategy. Similarly, the net incremental flux linkage $\lambda_{\Delta x}$ of the NFM accounts for both the self and mutual incremental flux linkage of the conventional modeling strategy. For bipolar MCSRMs, ω_e is related to the mechanical speed ω_m, and the number of rotor poles N_r by the relation in (10).

$$\omega_e(t) = \frac{N_r \omega_m(t)}{2} \tag{10}$$

The phase voltages can also be expressed as in the following using (8)-(9).

$$v_{aN} = Ri_a + \omega_e \lambda_{\Delta a} + \psi_{\Delta a}\frac{di_a}{dt} \tag{11}$$

In (11), there is speed dependent voltage which represents the back-emf (BEMF) of the machine, a resistive drop, and an inductive drop which determines the dynamic behaviour of the phase current. Hence, this shows that by taking the route of the NFM the number of terms required to model are reduced significantly. The net flux, net incremental inductance, and net flux linkage characteristics extracted from the FE are shown in Fig. 3 and the process of extracting these net flux characteristics is illustrated in [7]. It can be observed that the net phase flux, λ_x, is a strong function of rotor position and current magnitude and can be expressed using Fourier series for an arbitrary phase, say x as

$$\lambda_x(i_x, \theta_e) = \sum_{n=1}^{m} \lambda_{nax}(f(i_x)) \sin(n\theta_e) + \lambda_{nbx}(f(i_x)) \cos(n\theta_e) \tag{12}$$

The harmonic spectrum of λ_x, as given in Fig. 4(a), shows that λ_x has a dc and a dominant fundamental component and is true for any slot-pole combination. The harmonic spectrum is obtained using the electrical rotor position θ_e, as a result this conclusion will be valid for any slot-pole combination. Therefore, the net phase flux can be approximated using the dc component and the first order sinusoidal component.

$$\lambda_x(i_x, \theta) \approx \lambda_{x_0}(i_x) + \lambda_{x_1}(i_x) \cos\theta_e \tag{13}$$

Using (8) and (13)

$$\begin{aligned} \psi_{\Delta x}(i_x, \theta) &\approx \frac{\partial \lambda_{x_0}}{\partial i_x} + \frac{\partial \lambda_{x_1}}{\partial i_x} \cos\theta_e \\ &\approx \psi_{\Delta x_0}(i_x) + \psi_{\Delta x_1}(i_x) \cos\theta_e \end{aligned} \tag{14}$$

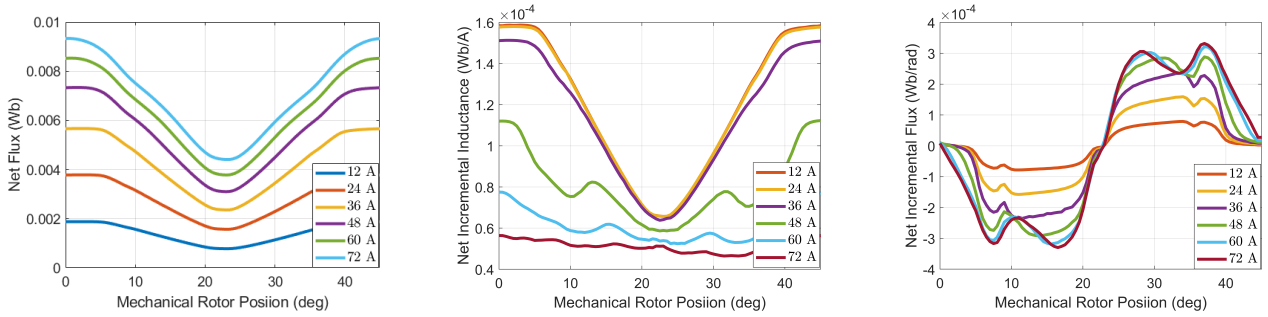
where $\psi_{\Delta x_0} = \frac{\partial \lambda_{x_0}}{\partial i_x}$ and $\psi_{\Delta x_1} = \frac{\partial \lambda_{x_1}}{\partial i_x}$. Similarly, using (9) and (13)

$$\begin{aligned} \lambda_{\Delta x}(i_x, \theta) &\approx \frac{\partial}{\partial \theta_e}\Big(\lambda_{x_0}(i_x) + \lambda_{x_1}(i_x) \cos\theta_e\Big) \\ &\approx -\lambda_{x_1}(i_x) \sin\theta_e \end{aligned} \tag{15}$$

where λ_{x_0} and $\psi_{\Delta x_0}$ are the dc components of net phase flux and net incremental inductance, respectively for an arbitrary phase x. Similarly λ_{x_1} and $\psi_{\Delta x_1}$, and are the magnitudes of the fundamental components of net phase flux and

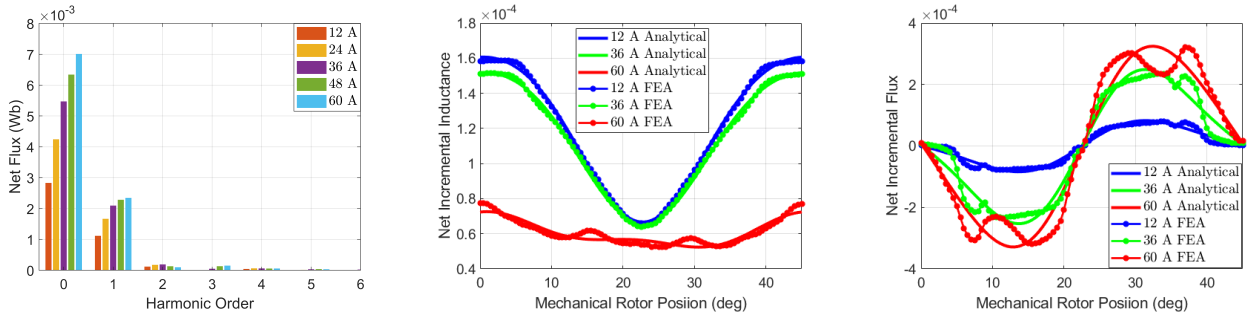
0093-9994 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on November 06,2020 at 04:57:52 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/TIA.2020.3030873, IEEE Transactions on Industry Applications



(a) Flux Linkage, λ_a (b) Net incremental inductance, $\psi_{\Delta a}$ (c) Net incremental flux, $\lambda_{\Delta a}$

Fig. 3. Flux characteristics of 12/8 MCSRM from FE model.



(a) Harmonic spectrum of λ_a (b) Net incremental inductance, $\psi_{\Delta a}$ (c) Net incremental flux, $\lambda_{\Delta a}$

Fig. 4. Comparison of flux characteristics of between the FE model and proposed small-signal model.

net incremental inductance, respectively. $\lambda_{\Delta x}$ and $\psi_{\Delta x}$ are reconstructed using (14) and (15) and compared with the characteristics obtained from the FE model and are shown in Fig. 4(b) and 4(c), respectively. The closeness between the analytical expression and the FE characteristics of the motor validate the approximation made above to ignore higher order harmonics. A point to note is the difference between the MCSRMs and brushless DC machines (BLDCs). Both are excited using bipolar-trapezoidal currents, but in the case of BLDCs, only the fundamental component is dominant, along with other harmonics, and there is no dc component in the flux. On the contrary for MCSRMs, dc and the fundamental components are both significant as discussed above. Therefore, a model for MCSRM considering the dc and fundamental is designed in the stationary reference frame. If MCSRMs are modeled in the synchronous (dq) reference frame, then the dc component of the flux will appear as a sinusoidal quantity, which will not provide any advantage.

The analytical model is helpful to develop a linear time invariant model so that concepts of the control theory can be applied to develop a current control scheme. For modeling MCSRM, phase current is taken as a state variable and net incremental flux, $\lambda_{\Delta x}$, can be expressed as a function of phase current as shown in (16).

$$\lambda_{\Delta x}(i_x, \theta) = \Lambda_{\Delta x}(\theta_e) i_x \tag{16}$$

where $\Lambda_{\Delta x}$ is the BEMF constant and is a function of rotor position. As explained earlier, only the fundamental component is essential to model; therefore, from (15) and (16)

$$\lambda_{\Delta x}(i_x, \theta) \approx \lambda_{x_1} \sin\theta_e \approx \Lambda_{\Delta x_1} i_x \sin\theta_e \tag{17}$$

where $\Lambda_{\Delta x_1}$ is the magnitude of the fundamental of the BEMF constant. Using the above relations, the phase voltage can be expressed as below and the non-linear plant model is shown in Fig. 5.

$$v_{xN} = R i_x + \omega_e(t) i_x \Lambda_{\Delta x_1} \sin\theta_e + (\psi_{\Delta x_0} + \psi_{\Delta x_1} \cos\theta_e) \frac{di_x}{dt} \tag{18}$$

where $\Lambda_{\Delta x_1}$ is the magnitude of the fundamental component the BEMF constant, $\Lambda_{\Delta x}$. Also, note that the motor speed is considered as a time varying quantity.

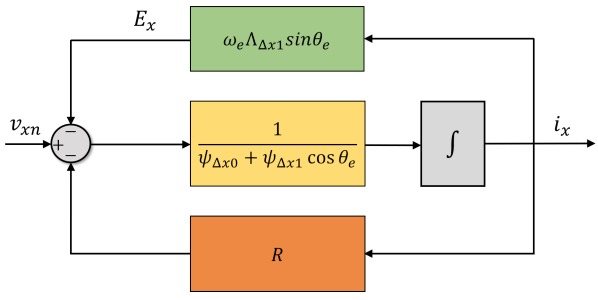


Fig. 5. Non-linear per phase model of MCSRM.

IV. Small-Signal Model

Although a simplified model for the MCSRM is developed in the previous section, the net flux is still non-linear due to the

0093-9994 (c) 2020 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 November 06,2020 at 04:57:52 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/TIA.2020.3030873, IEEE Transactions on Industry Applications

coupled relationship between phase currents and rotor position. Moreover, the BEMF term is product of two time varying quantities, speed and phase currents, which make the model non-linear. In order to linearize the model around an operating point, it is assumed that the operating speed is constant such that

$$\theta_e = \int_0^t \omega_e dt = \omega_e t \tag{19}$$

Motor speed is considered as a modeling parameter and not an explicit function of time, and therefore, is not a state variable in this modeling approach.

A. Linearization

The BEMF and the net incremental inductance are the only non-linear parameters that are to be linearized.

1) Inductive Drop: $\psi_{\Delta x}$ is a function of both current and rotor position. Hence, $\psi_{\Delta x}$ can be expressed as a function of current as

$$\psi_{\Delta x}(i_x, \theta_e) = (\Phi_{\Delta x_0} + \Phi_{\Delta x_1} \cos\theta_e) i_x \tag{20}$$

where $\Phi_{\Delta x_0}$ and $\Phi_{\Delta x_1}$ are the magnitudes of the dc and the fundamental components of the net inductance constant, $\Phi_{\Delta x}$. This constant is utilized to capture the effect of saturation during linearization. Using Taylor series expansion around an operating point of rotor position, θ_{e0} and current magnitude I_{x0}, the incremental inductance is linearized as expressed below.

$$\begin{aligned}\psi_{\Delta_x}\Big|_{\theta_{e0}, I_{x0}} &= (\Phi_{\Delta x_0} + \Phi_{\Delta x_1}\cos\theta_{e0}) I_{x0} \\ &+ (\Phi_{\Delta x_0} + \Phi_{\Delta x_1}\cos\theta_e)(i_x - I_{x0}) \\ &- \Phi_{\Delta x_1} i_x \sin\theta_e (\theta_e - \theta_{e0})\end{aligned} \tag{21}$$

The higher order terms are neglected while using the Taylor series. Net incremental inductance, illustrated in (21), when evaluated at $\theta_{e0} = 0°, i_x = I_{x0}$ and $\theta_{e0} = \pi, i_x = Ix_0$ yields the value of $\psi_{\Delta x}$ as $(\Phi_{\Delta x_0} + \Phi_{\Delta x_1}) I_{x0} = (\psi_{\Delta x_0} + \psi_{\Delta x_1})$ and $(\Phi_{\Delta x_0} - \Phi_{\Delta x_1}) I_{x0} = (\psi_{\Delta x_0} - \psi_{\Delta x_1})$, respectively. This can also be observed from Fig. 3(b). This analysis linearizes the inductive drop term, however the model is still a time-varying as $\psi_{\Delta x}$ is varying with time as shown in (21).

2) BEMF: In this work, the BEMF is modeled as a known sinusoidal disturbance as expressed in (15). Using Taylor series, the BEMF term is linearized as

$$\begin{aligned}E_x\Big|_{\theta_{e0}, I_{x0}} &= \omega_e(-\Lambda_{\Delta x_1} i_x \sin\theta_{e0} - \Lambda_{\Delta x_1}\sin\theta_e (i_x - I_{x0}) \\ &- \Lambda_{\Delta x 1}\cos\theta_e(\theta_e - \theta_{e0}))\end{aligned} \tag{22}$$

BEMF expressed in (22), when evaluated at $\theta_{e0} = \frac{\pi}{2}, i_x = I_{x0}$ and $\theta_{e0} = \frac{3\pi}{2}, i_x = I_{x0}$ will vary from $-\Lambda_{\Delta x_1}\omega_e i_x$ and $+\Lambda_{\Delta x_1}\omega_e i_x$ In one electrical period, the BEMF will range from $-\Lambda_{\Delta x_1}\omega_e i_x$ to $\Lambda_{\Delta x_1}\omega_e i_x$, therefore, can be approximated as a a step disturbance as illustrated in (23).

$$E_x(t) \approx \text{sgn}(\theta - \pi)\omega_e \Lambda_{\Delta x_1} i_x \approx \pm\omega_e \Lambda_{\Delta x_1} i_x \tag{23}$$

Although the sgn function is non-linear, the BEMF can be expressed as a linear function in a piece-wise manner. This makes the model a linear time variant (LTV) system which will be further evaluated in the next section. Using the above relations, the phase voltage can be expressed in a LTV expression as

$$v_{xN} = R i_x \pm \omega_e i_x \Lambda_{\Delta x_1} + (\psi_{\Delta x_0} \pm \psi_{\Delta x 1})\frac{di_x}{dt} \tag{24}$$

When the system is perturbed around the operating point, (V_{xN}, I_x), using small-signal perturbations, $\tilde{i}_x(t)$ and $\tilde{v}_{xN}(t)$ the voltage balance equation becomes

$$\begin{aligned}V_{xN} + \tilde{v}_{xN}(t) = R(I_x + \tilde{i}_x(t)) \pm \omega_e (I_x + \tilde{i}_x(t))\Lambda_{\Delta x_1} \\ + (\psi_{\Delta x_0} \pm \psi_{\Delta x_1})\frac{d(I_x + \tilde{i}_x(t))}{dt}\end{aligned} \tag{25}$$

The small-signal model is then derived as

$$\tilde{v}_{xN}(t) = R\tilde{i}_x(t) \pm \omega_e \tilde{i}_x(t)\Lambda_{\Delta x 1} + (\psi_{\Delta x 0} \pm \psi_{\Delta x 1})\frac{d\tilde{i}_x(t)}{dt} \tag{26}$$

The small-signal model of MCSRM is shown in Fig. 6 in a block diagram form.

B. State Space Representation

The system's performance in the presence of plant parameter variations can be evaluated with the help of a state space representation of the plant model which can be expressed as

$$\frac{d\tilde{\boldsymbol{x}}}{dt} = \boldsymbol{A}\tilde{x} + \boldsymbol{B}\tilde{\boldsymbol{u}} \tag{27}$$

where $\tilde{\boldsymbol{x}}$ is the state vector, and $\tilde{\boldsymbol{u}}$ is the input vector. In this model, the phase currents are considered to be the states and the phase voltages are the inputs. The state and input matrices are defined below.

$$\boldsymbol{A} = -\begin{bmatrix} \frac{R \pm \omega_e \Lambda_{\Delta x_1}}{\psi_{\Delta x 0} \pm \psi_{\Delta x 1}} & 0 & 0 \\ 0 & \frac{R \pm \omega_e \Lambda_{\Delta y_1}}{\psi_{\Delta y 0} \pm \psi_{\Delta y 1}} & 0 \\ 0 & 0 & \frac{R \pm \omega_e \Lambda_{\Delta z_1}}{\psi_{\Delta z 0} \pm \psi_{\Delta z 1}} \end{bmatrix},$$

$$\boldsymbol{B} = \begin{bmatrix} \frac{1}{\psi_{\Delta x 0} \pm \psi_{\Delta x 1}} & 0 & 0 \\ 0 & \frac{1}{\psi_{\Delta y 0} \pm \psi_{\Delta y 1}} & 0 \\ 0 & 0 & \frac{1}{\psi_{\Delta z 0} \pm \psi_{\Delta z 1}} \end{bmatrix},$$

$$\tilde{\boldsymbol{x}} = \begin{bmatrix} \tilde{i}_x & \tilde{i}_y & \tilde{i}_z \end{bmatrix}^T,$$

and

$$\tilde{\boldsymbol{u}} = \begin{bmatrix} \tilde{v}_{xN} & \tilde{v}_{yN} & \tilde{v}_{zN} \end{bmatrix}^T .$$

C. Transfer Function Matrix

The frequency domain representation of the MCSRM model using transfer function matrices can be derived as

$$\tilde{\boldsymbol{i}}(\boldsymbol{s}) = \boldsymbol{G_m}\tilde{\boldsymbol{v}}(\boldsymbol{s}) \tag{28}$$

$$\tilde{\boldsymbol{i}}(\boldsymbol{s}) = \begin{bmatrix} \tilde{i}_x & \tilde{i}_y & \tilde{i}_z \end{bmatrix}^T,$$

$$\tilde{\boldsymbol{v}}(\boldsymbol{s}) = \begin{bmatrix} \tilde{v}_{xN} & \tilde{v}_{yN} & \tilde{v}_{zN} \end{bmatrix}^T$$

0093-9994 (c) 2020 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 November 06,2020 at 04:57:52 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/TIA.2020.3030873, IEEE Transactions on Industry Applications

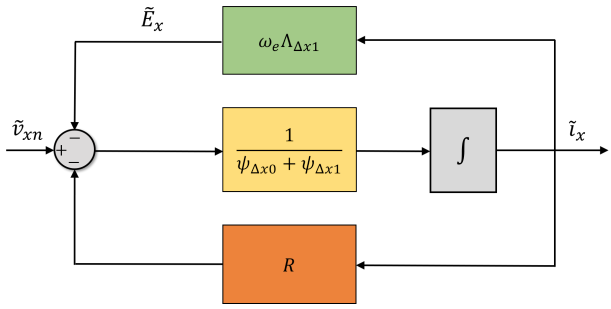


Fig. 6. Small-signal model of MCSRM

where G_m is the input to output transfer matrix, $\tilde{i}(s)$ and $\tilde{\boldsymbol{v}}(\boldsymbol{s})$ are the current and voltage matrices, respectively.

$$\begin{bmatrix}\tilde{i}_x(s)\\ \tilde{i}_y(s)\\ \tilde{i}_z(s)\end{bmatrix}=\begin{bmatrix}G_{mx} & 0 & 0\\ 0 & G_{my} & 0\\ 0 & 0 & G_{mz}\end{bmatrix}\begin{bmatrix}\tilde{v}_{xN}(s)\\ \tilde{v}_{yN}(s)\\ \tilde{v}_{zN}(s)\end{bmatrix} \tag{29}$$

$$G_{mx}(s)=\frac{\tilde{i}_x(s)}{\tilde{v}_{xN}(s)}=\frac{1}{s(\psi_{\Delta x0}\pm\psi_{\Delta x1})+(R\pm\Lambda_{\Delta x1}\omega_e)} \tag{30}$$

Equation (30) shows that the MCSRM plant model is time variant in nature because the net incremental inductance varies with the rotor position. To perform a linear analysis, the plant model is studied for three different scenarios: first, for the maximum value of the net incremental inductance (aligned position), $\psi_{\Delta max}=\psi_{\Delta x0}+\psi_{\Delta x1}$; second, for the minimum value of the net incremental inductance (unaligned position), $\psi_{\Delta min}=\psi_{\Delta x0}-\psi_{\Delta x1}$; third, for the mean value of the net incremental inductance,, $\psi_{\Delta mean}=\psi_{\Delta x0}$.

The frequency response of the plant, $G_{mx}(s)$, obtained from (30) is shown in Fig. 7 for three different values of the $\psi_{\Delta x}$, which are at 36 A and a speed of 500 RPM. The machine parameters are given in Table I.

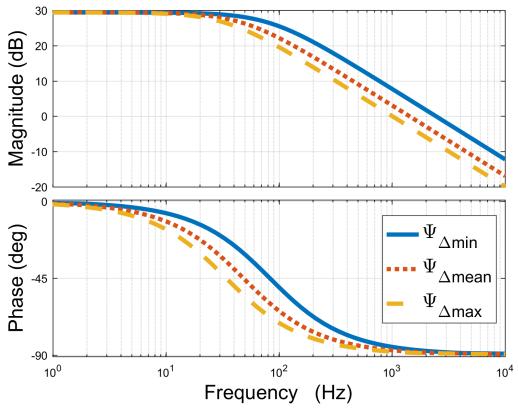


Fig. 7. Frequency response at three different operating points of plant transfer function.

D. Validation of Small-Signal Model

To validate the above model, the frequency response obtained from (30) is compared with the frequency response obtained from the FE model of MCSRM. To obtain the frequency response characteristics from the FE model, the MCSRM should be operating at a fixed speed. Two phases are excited with sinusoidal phase voltages with a fixed magnitude but of different frequencies ranging from 1 Hz to 1000 Hz in discrete steps such that equal but opposite polarity of current flows through them. For each test frequency, the phase current drawn by the machine is recorded. The phase current will have higher order harmonics in addition to the fundamental component of the phase voltage. Therefore, using Fast Fourier Transform (FFT), the magnitude of the fundamental component in the phase current is obtained. The gain of the MCSRM ($|i_x|/|v_x|$) at each frequency and the phase difference between the phase voltage and phase current is similarly calculated at each test frequency.

A comparison between the frequency response of the analytically estimated small-signal model and the frequency response obtained from the FEA is shown in Fig. 8 which shows a reasonable match for a current level of 36 A at 50 and 500 RPM. In the low frequency region, the gain is determined by the phase resistance and the operating speed of the machine. The match between the model and the FE model is close because the model parameters are known precisely. However, there is a difference in the prediction in the higher frequency range, but the trend is still captured by the model. This difference in the high frequency region is due to neglition the higher order harmonics. As the frequency increases, the magnitude of the higher order harmonics become significant and the estimated value of $\Psi_{\Delta x}$ and $\lambda_{\Delta x}$ will deviate.
The frequency response of a small-signal model when compared to the frequency response of the complete non-linear, large signal model is expected to have differences. However, the differences are not significant and the small-signal model is sufficiently valid to capture the machine frequency response, and hence, can be used for building machine models with reduced dependency on FEA and designing current controllers. The current controller design is presented next.

TABLE I
PARAMETERS OF PROTOTYPE MCSRM

Parameter	Value
Rated Power (W)	150
Rated Torque (Nm)	1.5
DC Link Voltage (V)	12
Line-Line Resistance (mΩ)	65
Rated Current (A)	72
Switching Frequency (kHz)	20
Sampling Frequency (kHz)	20
Rated Speed (RPM)	1000

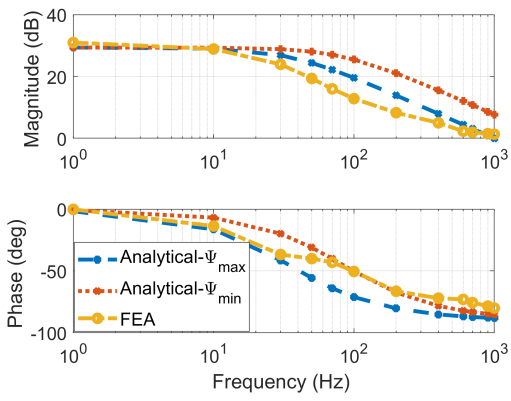
V. Application of Small-Signal Model

A. Plant Modeling and Characterization

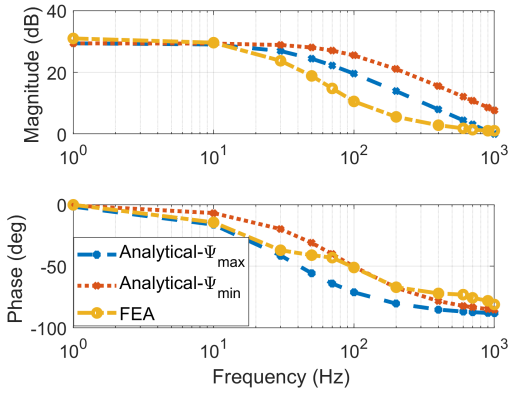
For the MCSRM drive system, the inverter, current sensor feedback and the plant transfer functions, $G_i(s)$, $G_f(s)$ and $G_p(s)$ are defined in (31), (32), and (33), respectively. In the text to follow, the subscript x will be dropped as all the

0093-9994 (c) 2020 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 November 06,2020 at 04:57:52 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/TIA.2020.3030873, IEEE Transactions on Industry Applications



(a) 36 A at 50 RPM



(b) 36 A at 500 RPM

Fig. 8. Validation of the small-signal model using frequency response analysis.

analysis is done on a per-phase basis.

$$G_i(s) = \frac{\tilde{V}(s)}{\tilde{d}(s)} = \frac{\hat{V}_{dc}}{1+sT_d} \tag{31}$$

$$G_f(s) = \frac{\tilde{I}_f(s)}{\tilde{I}(s)} = \frac{k_f}{1+sT_f} \tag{32}$$

where T_d, T_f are the time constants for the pulse width modulator (PWM) and the current measurement delay, respectively, $\hat{V}_{dc}$ is the estimated dc link voltage, and k_f is the gain of the feedback current sensor. The plant as seen by the current controller is given by (33) as

$$G_p(s) = \frac{\tilde{I}_f(s)}{\tilde{d}(s)} = \frac{\hat{V}_{dc}k_f}{(s\Psi_\Delta + R_{eq})(1+sT_d)(1+sT_f)} \tag{33}$$

$$\Psi_\Delta = \psi_{\Delta 0} \pm \psi_{\Delta 1} \tag{34}$$

$$R_{eq} = R \pm \Lambda_{\Delta 1}\omega_e \tag{35}$$

From (33), it can observed that the plant transfer function is dependent on the value of the net incremental inductance ψ_Δ, which is a time varying parameter and makes the plant a LTV system. Hence, in order to observe the maximum variation, the frequency analysis of $G_m(s)$ for three different values of $\psi_{\Delta x}$ are shown in Fig. 7. The three different values correspond to maximum, minimum and mean value of ψ_Δ at a current of 36 A. The MCSRM drive's block diagram is shown in Fig. 9. The input disturbance $\tilde{V}_{dist}$ can be attributed to the modeling uncertainties because of the assumptions of considering only the dominant harmonic given that the machine is operating only in normal conduction period and linearization. The controller, $G_c(s)$, is defined in (36).

$$G_c = K_p + \frac{K_i T_s}{1-z^{-1}} = K_p + \frac{K_i T_s}{1-e^{-T_s s}} = K_p + \frac{K_i}{\hat{s}} \tag{36}$$

$$\hat{s} = \frac{1-e^{-T_s s}}{T_s}. \tag{37}$$

where K_p and K_i are the proportional and integral gains of the controller. In order to analyze the performance of the current controller and the effects of tuning parameters on the system's performance, different transfer functions describing the closed loop response of the system are defined below. The current tracking capability, disturbance rejection, control signal, and the input disturbance rejection are defined by the transfer functions $G_{ct}(s)$, $G_{dr}(s)$, $G_{cr}(s)$, and $G_{id}(s)$. The output and the control effort of the system in terms of reference current and disturbances can be expressed as shown below in (38) and (39), respectively.

$$\tilde{I}(s) = G_{ct}(s)\tilde{I}^*(s) + G_{dr}(s)\tilde{V}_{dist}(s) \tag{38}$$

$$\tilde{V}(s) = G_{cr}(s)\tilde{I}^*(s) + G_{id}(s)\tilde{V}_{dist}(s) \tag{39}$$

$$G_{ct}(s) = \frac{\tilde{I}(s)}{\tilde{I}^*(s)} = \frac{G_c(s)G_m(s)G_i(s)}{1+G_c(s)G_m(s)G_i(s)G_f(s)} \tag{40}$$

$$G_{dr}(s) = \frac{\tilde{I}(s)}{\tilde{V}_{dist}(s)} = \frac{-G_m(s)}{1+G_c(s)G_m(s)G_i(s)G_f(s)} \tag{41}$$

$$G_{cr}(s) = \frac{\tilde{V}(s)}{\tilde{I}^*(s)} = \frac{G_c(s)G_i(s)}{1+G_c(s)G_m(s)G_i(s)G_f(s)} \tag{42}$$

$$G_{id}(s) = \frac{\tilde{V}(s)}{\tilde{V}_{dist}(s)} = \frac{G_c(s)G_m(s)G_i(s)G_f(s)}{1+G_c(s)G_m(s)G_i(s)G_f(s)} \tag{43}$$

For simplicity, k_f is considered to be unity, T_d is equal to the switching time period, $T_s = 50\mu s$ and T_f is considered to be much smaller as compared to T_s and is neglected.

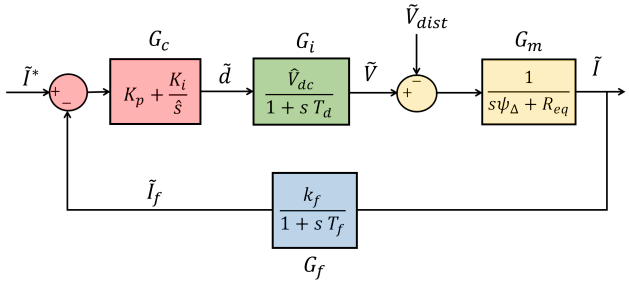


Fig. 9. MCSRM drive block diagram for current controller design.

0093-9994 (c) 2020 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 November 06,2020 at 04:57:52 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/TIA.2020.3030873, IEEE Transactions on Industry Applications

B. Current Controller Design

For the current controller design, the PWM delay and the current feedback delay is considerably small as compared to the motor time constant, and therefore, they can be ignored when deriving relationships between the controller and the plant parameters. The closed loop response of the system can be tuned depending upon how the current controller is designed. Current controller design leading to a second order response of the system is discussed in [16]. In this work, the current controller $G_c(s)$ is designed such that a first order response of the system can be obtained. $G_c(s)$ is designed in terms of motor parameters and the desired bandwidth, ω_b, as given below.

$$G_c(s) = \frac{\omega_b}{\hat{V}_{dc}}(\hat{\Psi}_\Delta + \frac{\hat{R}_{eq}}{\hat{s}}) = K_p + \frac{K_i}{\hat{s}} \tag{44}$$

Hence, from (44) $K_p = \frac{\omega_b}{\hat{V}_{dc}}\hat{\Psi}_\Delta$ and $K_i = \frac{\omega_b}{\hat{V}_{dc}}\hat{R}_{eq}$. where $\hat{\Psi}_\Delta$ and $\hat{R}_{eq}$ are the estimated values of the motor parameters Ψ_Δ and R_{eq}; and ω_b is the frequency at the -3 dB gain and -45°phase, both same for the first order system. ω_b is a tunable parameter for designing the current controller and dictates the bandwidth of the system. Now $G_{ct}(s)$ can be expressed as

$$G_{ct}(s) = \frac{V_{dc}\omega_b(\hat{\Psi}_\Delta\hat{s} + \hat{R}_{eq})}{\hat{V}_{dc}\hat{s}(\Psi_\Delta s + R_{eq}) + \omega_b V_{dc}(\hat{\Psi}_\Delta\hat{s} + \hat{R}_{eq}).} \tag{45}$$

In case of accurate parameter estimation, (45) is reduced to (46), which is a first order system with -3 dB bandwidth at ω_b and DC gain of 0 dB. Gain close to 0 dB and no phase lag is ideally desired for reference tracking.

$$G_{ct}(s) = \frac{\omega_b}{s + \omega_b} \tag{46}$$

Similarly, other transfer functions are expressed in terms of the controller and motor parameters as shown below.

$$G_{dr}(s) = \frac{-\hat{s}\hat{V}_{dc}}{\hat{V}_{dc}\hat{s}(\Psi_\Delta s + R_{eq}) + \omega_b V_{dc}(\hat{\Psi}_\Delta\hat{s} + \hat{R}_{eq})} \tag{47}$$

$$G_{cr}(s) = \frac{V_{dc}\omega_b(\hat{\Psi}_\Delta\hat{s} + \hat{R}_{eq})(\Psi_\Delta s + R_{eq})}{\hat{V}_{dc}\hat{s}(\Psi_\Delta s + R_{eq}) + \omega_b V_{dc}(\hat{\Psi}_\Delta\hat{s} + \hat{R}_{eq})} \tag{48}$$

$$G_{id}(s) = \frac{V_{dc}\omega_b(\hat{\Psi}_\Delta\hat{s} + \hat{R}_{eq})}{\hat{V}_{dc}\hat{s}(\Psi_\Delta s + R_{eq}) + \omega_b V_{dc}(\hat{\Psi}_\Delta\hat{s} + \hat{R}_{eq})} \tag{49}$$

To understand the effect of ω_b on the system's performance, frequency responses are shown in Fig. 10 till the Nquist frequency of 10 kHz, since the control (switching) frequency is 20 kHz. First, the analysis is presented for the case when the machine parameters are $\Psi_\Delta = \psi_{\Delta 0} - \psi_{\Delta 1}$ and $R_{eq} = R + \Lambda_{\Delta 1}\omega_e$. This analysis is carried out assuming accurate estimation of machine parameters. Therefore, $\hat{\Psi}_\Delta = \Psi_\Delta$ and $\hat{R}_{eq} = R_{eq}$. The frequency responses for this case is shown in Fig. 10(a)-(c). With accurate parameter estimation, the bandwidth of the current tracking performance increases with an increase of ω_b; furthermore, the disturbance rejection is higher at lower frequencies but reduces the noise rejection property of the system as higher frequency noise will be transmitted into the system [18].

Second in Fig. 10(d)-(f), three different cases are presented to understand the effect of parameter mismatch on the control performance for a fixed bandwidth of 1000 Hz. The controller design is fixed with $\hat{\Psi}_\Delta = \Psi_{min}$, but the plant's actual net incremental inductance Ψ_Δ is varied from 50% of Ψ_{min} to 100% of Ψ_{min}. This will mimic a real time scenario when the plant's net incremental inductance will change because of rotor position and/or saturation. However, the effect on the current tacking performance and disturbance rejection properties is not very significant, especially when the ω_b is high, thus making the design robust to parameter variations.

In addition to the frequency response analysis, a pole-zero map for the closed loop system is presented in Fig. 11 for parameter mismatch between the plant and the controller. This shows the migration of poles and zeroes of the system for the case when there is an error in the estimation of plant parameters. The closed loop response can change from critically damped to under damped or over damped condition when the net incremental inductance is over estimated or under estimated, respectively. In the entire analysis the effect of PWM delay is not considered as it is outside the scope of this paper and is discussed in [18].

VI. Results and Discussions

A 12/8, 150W prototype MCSRM, as shown in Fig. 12 has been built and tested to validate the proposed small signal model. The motor parameters are given in Table I. An induction motor drive is used as the dyno-motor to rotate the test motor at a constant speed, while shaft torque and speed is obtained from a torque sensor and a position encoder.

A. Static Motor Characteristics

The static flux characteristics is obtained by rotating the prototype test motor (MCSRM) at a fixed speed of 20 RPM. Two phases of the MCSRM are simultaneously supplied with fixed magnitude of dc currents, ranging from 0 to 72 A in increments of 4 A, while the third phase in not conducting. The current is regulated using a 12 V power supply with a current capability of 200 A and is operated in the constant current mode. During this test, phase currents and line-line voltages are recorded. Experimental net flux for each phase is determined from the following voltage equation:

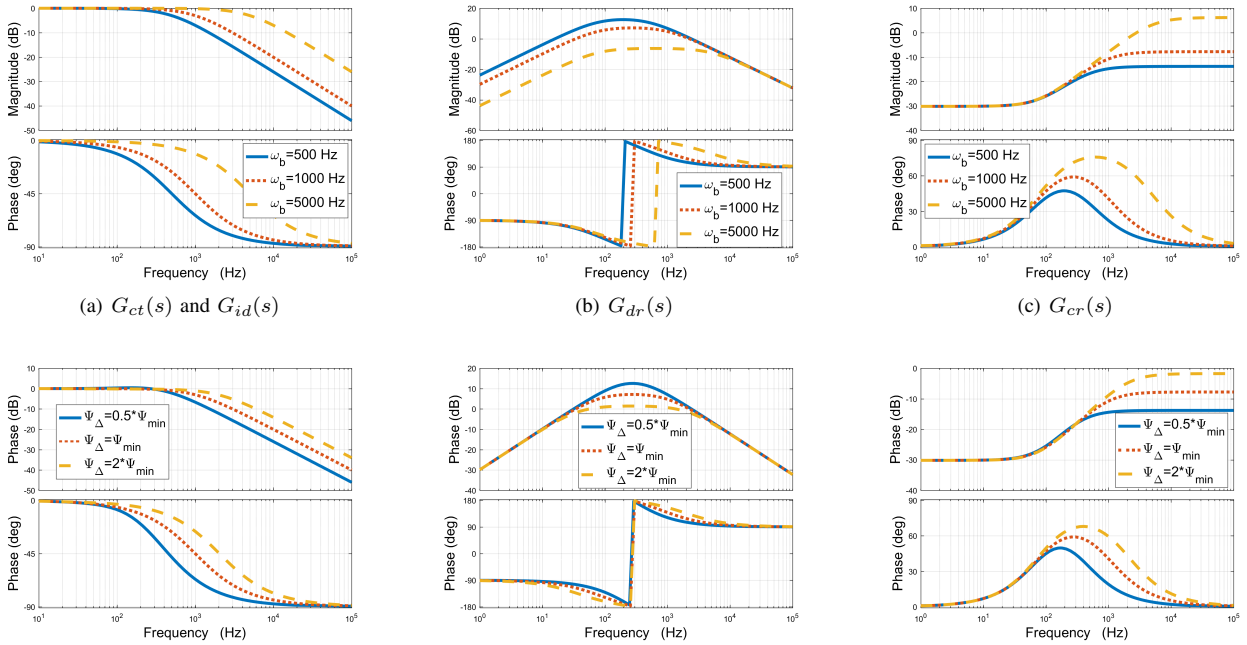
$$\lambda_x(i_x, \theta) = \int_0^{i_x} (v_{xN}(i_x, \theta) - i_x R)\ di_x \tag{50}$$

The static flux characteristics of the 12/8 MCSRM obtained experimentally as well as that obtained using the FE model are presented in Fig. 13 which shows an excellent match. Experimental net incremental inductance and net incremental flux characteristics are calculated from the experimental net flux according to (8) and (9).

Subsequently, experimentally obtained net flux is also post processed to obtain net incremental inductance and net incremental flux analytically by considering only dc and fundamental component using (13)-(15). The close match between the estimated and the measured net incremental inductance and

0093-9994 (c) 2020 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 November 06,2020 at 04:57:52 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/TIA.2020.3030873, IEEE Transactions on Industry Applications



(a) $G_{ct}(s)$ and $G_{id}(s)$ (b) $G_{dr}(s)$ (c) $G_{cr}(s)$

(d) $G_{ct}(s)$ and $G_{id}(s)$ (e) $G_{dr}(s)$ (f) $G_{cr}(s)$

Fig. 10. Frequency response analysis for variation in desired bandwidth, ω_b, and for variation in machine parameters

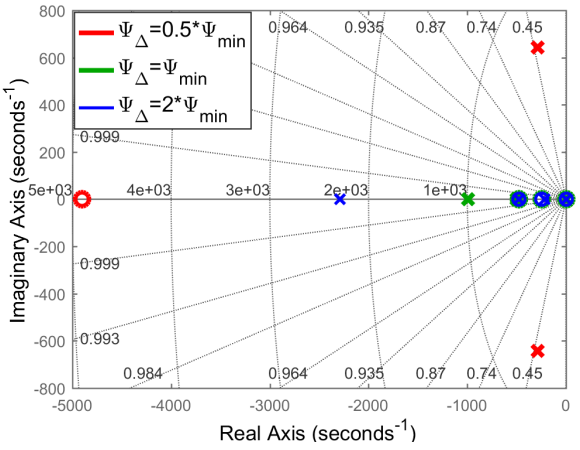


Fig. 11. Pole-zero map to analyze stability of the closed loop system considering parameter mismatch.

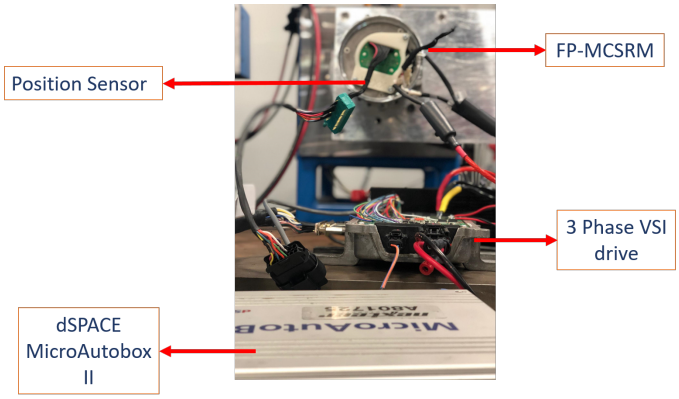


Fig. 12. Experimental test setup for MCSRM drive.

net incremental flux validates the analytical model described in Section III. The differences between the experimental and the analytical results are due to neglecting the higher order harmonics in the analytical estimations of net incremental inductance and net incremental flux because of their insignificant contributions.

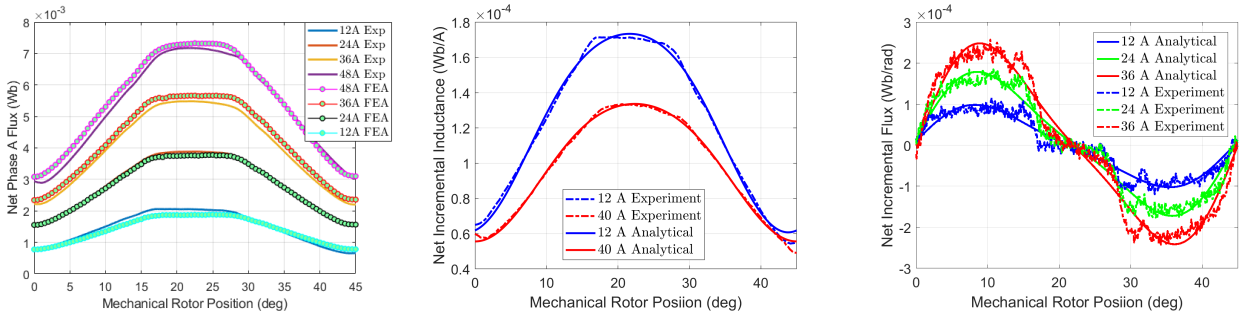
B. Small-Signal Model Validation

To validate the small-signal model, analysis is presented in time domain and frequency domain for the prototype motor. The induction motor drive is operated in speed controlled mode at a fixed speed. The test MCSRM is operated using a 3 phase VSI drive and dSPACE Microautobox-II in torque controlled mode. The block diagram for this setup is shown in Fig. 9.

The PI current controller is commanded with sinusoidal reference currents of 20 A peak amplitude. The objective of the sinusoidal excitation is to extract the frequency response characteristics of the MCSRM plant; a trapezoidal excitation is typically used for torque production in an MCSRM. The frequency of the commanded sinusoidal current is varied from 1 Hz to 1000 Hz in discrete steps while keeping the magnitude fixed. At this time only 2 phases are conducting to mimic the normal conduction mode. The motor is operating at a fixed speed of 50 RPM. The phase voltage commanded by the PI controller and the phase current drawn by the machine

0093-9994 (c) 2020 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 November 06,2020 at 04:57:52 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/TIA.2020.3030873, IEEE Transactions on Industry Applications



(a) Net phase flux , λ_a (b) Net incremental inductance, $\psi_{\Delta a}$ (c) Net incremental flux, $\lambda_{\Delta a}$

Fig. 13. Experimental flux characteristics of 12/8 MCSRM.

are recorded to obtain the frequency response characteristics. The experimental voltage is then fed to the proposed small-signal model of the machine and the performance of the model is compared with the real machine in both time and frequency domains. This analysis is presented next. These phase voltages and phase currents are recorded using the dSPACE Microautobox-II.

1) Time Domain Analysis: The proposed small-signal machine model is setup in Simulink in the form a transfer function given by (30). The time domain plots for the experimental voltage and currents and the phase current predicted by the small-signal model are shown in Fig. 14(a)-(c). The higher order flux harmonics are not accounted for in the inductance model; as a result, the difference between the predicted current and the experimental current increases in the higher frequency region.

2) Frequency Domain Analysis: The FFT of the experimental sinusoidal phase voltage and current is obtained at all the discrete test frequencies. The FFT for phase current and voltage at 100 Hz, 400 Hz, and at 700 Hz is shown in Fig. 14(d)-(f). The magnitude of the gain (i_x/v_x) and the phase difference between the phase voltage and current of the MCSRM is obtained from the FFT. The gain and the phase delay are calculated analytically for the plant model given by (30). The analytical and experimental frequency response characteristics of the MCSRM are then analyzed and the results are shown in Fig. 15. The delay observed in the time domain plots in Fig. 14(a)-(c) translates to the difference in phase lag predicted by the model and that recorded from the experiment.

C. Sensitivity Analysis Using Trapezoidal Currents

The MCSRM is then operated with trapezoidal currents as they depict the actual operation scenario and also to perform the sensitivity analysis of the proposed model. With the trapezoidal shaped currents during the commutation period (see Fig. 1), the assumptions made during modeling will no longer be valid. This is because during the commutation period all the phases will have unequal magnitude of currents. In this case, the PI controller is commanded with trapezoidal current commands and the test is run at different speeds. The test MCSRM is operated using the 3 phase VSI drive and the dSPACE Microautobox-II in the torque controlled mode. The induction motor drive is used as the dyno-motor and is coupled to the MCSRM in order to rotate the MCSRM at a constant speed. Similar to the previous section, the experimental phase voltage is recorded and passed to the analytical small-signal model, given by equation (30), to predict the phase current. The analytically predicted phase current is then compared with the experimental results as shown in Fig. 16(a) and 16(b) for 100 and 500 RPM, respectively. The results indicate that even with the violation of the assumption (during the commutation period), the small-signal model is able to predict the phase currents fairly well. The small-signal model does not incorporate the effects of dead time, PWM and sensor delays, and errors in current sensor which will effect the accuracy of the prediction. It is expected that the predicted current from a small-signal model for a highly non-linear machine model will not be highly accurate.

The above extensive analysis presented shows that the small-signal model is able to predict the behavior behaviour of the MCSRM over a wide operating range. The model is suitable for simulating and analyzing MCSRMs with little to no dependency on the FEA. Furthermore, the model can be extrapolated to design different current control algorithms.

D. Current Controller Analysis

For three phase continuous operation with bipolar-balanced trapezoidal currents, we designed a PI current controller and experimentally verified it for the prototype motor operating at 1000 RPM for 20 A peak phase currents. To show the effect of ω_b on the current tacking performance and disturbance rejection, results are presented for two different ω_b, 160 Hz and 1.6 kHz, and the PI controller's parameters are calculated using (44). From Fig. 17(a), it is observed at a higher speed of 1000 RPM, current controller with lower bandwidth of 160 Hz (1000 rad/s) is not able to track the current command and there is steady state error. The steady state error is minimized when ω_b is increased to 1.6 kHz (10000 rad/s) as shown in Fig. 17(b). As ω_b is increased, the disturbance rejection capability and command tracking performance improves. At 1000 RPM, the higher order harmonics become dominant as their magnitude is directly proportional to the operating speed. Therefore, with increased ω_b, the disturbance rejection capability of the controller improves, which is evident from the frequency response shown in Fig. 10(b). This results in

0093-9994 (c) 2020 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 November 06,2020 at 04:57:52 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/TIA.2020.3030873, IEEE Transactions on Industry Applications

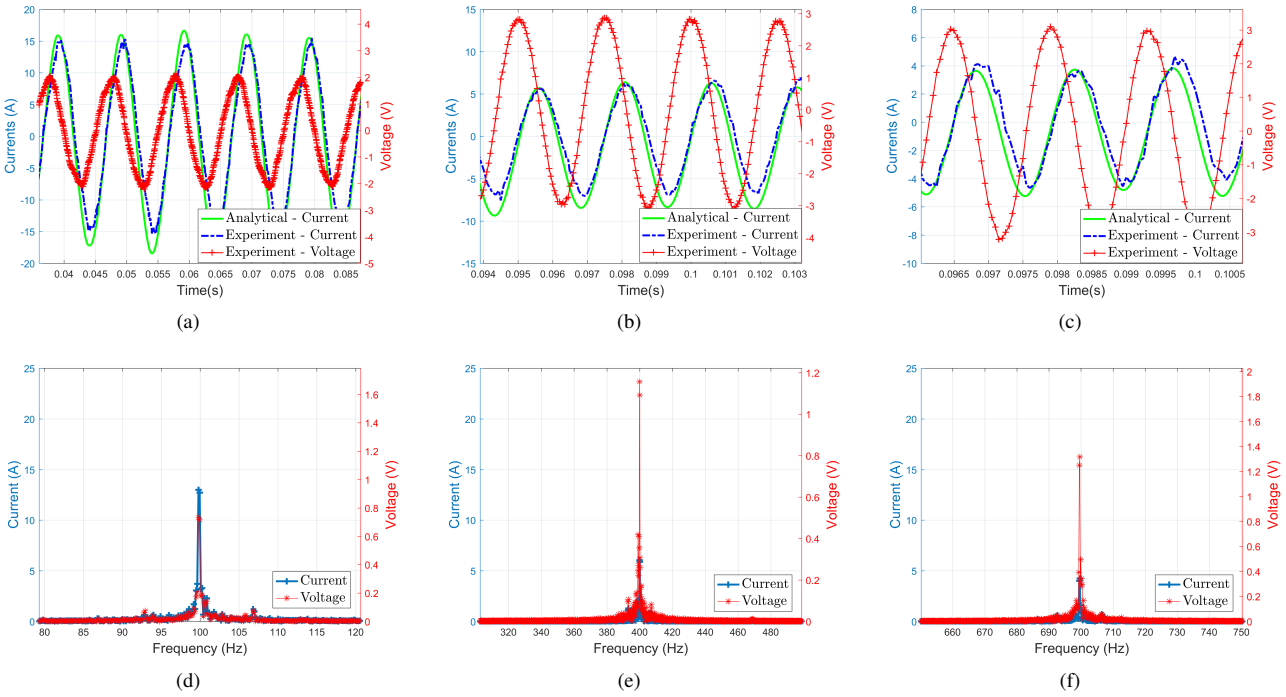


Fig. 14. Time domain plots of sinusoidal currents at (a) 100 Hz, (b) 400 Hz, and (c) 700 Hz and FFT for phase currents and voltages at (d) 100 Hz, (e) 400 Hz, and (f) 700 Hz. at 50 RPM.

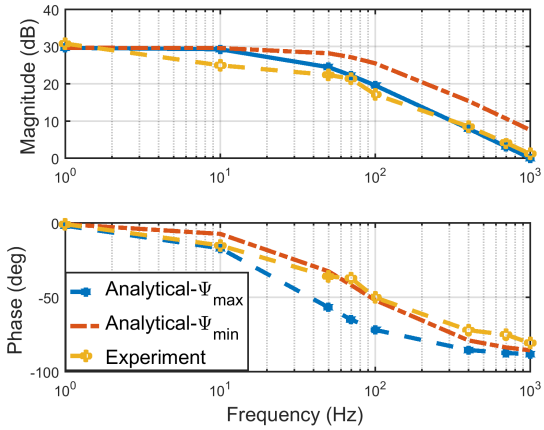


Fig. 15. Frequency response of 150 W MCSRM at 20A, 50 RPM.

enhanced current tracking performance. However, increasing ω_b also increases the noise content as the higher frequency noise can now be transmitted in the system [18].

The dynamic performance of the current controller is tested when the current command is increased in step change from 24 A to 30 A with the motor operating at a fixed speed of 100 RPM as shown in Fig. 17(c). In addition, another test to analyze the dynamic performance of the controller is carried out in which the speed is varied from standstill condition to 1000 RPM, in step changes, with a fixed peak current command of 20 A. The current tracking performance for the instance when the speed is changed from 100 RPM to 500 RPM is shown in Fig. 17(d). These tests are carried out at

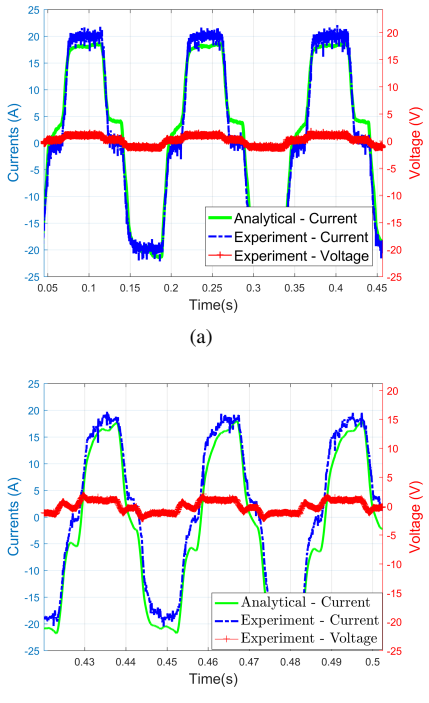


Fig. 16. Performance comparison of small-signal model and prototype machine with trapezoidal currents, with the violation of assumption, at (a) 100 RPM and (b) 500 RPM.

0093-9994 (c) 2020 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 November 06,2020 at 04:57:52 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/TIA.2020.3030873, IEEE Transactions on Industry Applications

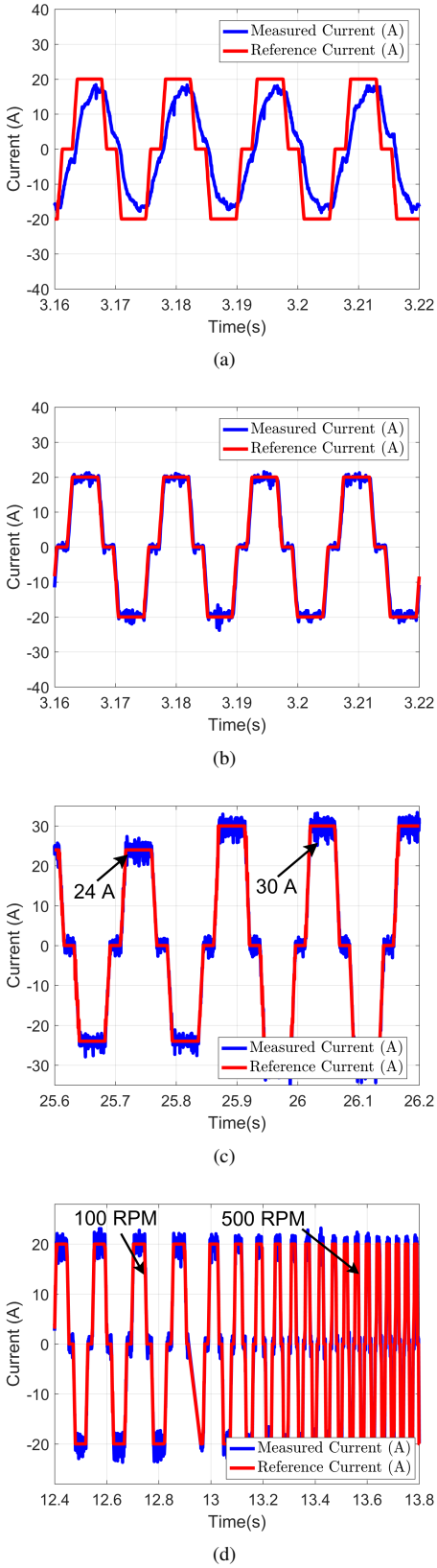


Fig. 17. Current controller's steady state performance at 1000 RPM for (a) ω_b=160 Hz and (b) ω_b=1.6 kHz and dynamic performance at ω_b=1.6 kHz for (c) change in reference current at fixed speed and (d) change in speed for a fixed current command.

$\omega_b = 1.6$ kHz.

VII. Conclusions

A small-signal model based on the net flux method (NFM) for MCSRM provides a simpler model that can be used for controller design using established frequency domain analysis. The NFM avoids estimating self and mutual inductances separately which yields the modeling simplicity. The non-linear characteristics of the motor are analytically represented using Fourier series which are then linearized using Taylor series for developing the small-signal model. The model is developed for a FP-MCSRM on a per-phase basis and is valid for other topologies of MCSRMs also and for both unipolar and bipolar excitations. The small-signal model is validated against the experimental results, which shows excellent accuracy of the analytical model over a wide spectrum of operating points. The NFM based small-signal model allows simplified transfer function representations for the MCSRM plant model which is useful for controller performance evaluation over a wide speed range. An application of the small-signal model is presented by designing a PI current controller for a 150 W MCSRM. The model has been used to analyze the current tracking performance, disturbance rejection, and noise transmission capabilities at different operating points. Similarly, the model can be utilized to design and analyze the performance of other types of current regulators such as proportional-resonant, deadbeat etc.

Acknowledgement

The authors would like to acknowledge the support of Nexteer Automotive Corporation for helping in experimental verification of results and Altair for providing the FE software.

References

[1] P. G. Barrass, B. C. Mecrow, and A. C. Clothier, "Bipolar operation of fully-pitched winding switched reluctance drives," in *1995 Seventh International Conference on Electrical Machines and Drives (Conf. Publ. No. 412)*, Sept. 1995, pp. 252–256.

[2] M. A. Kabir and I. Husain, "Design of mutually coupled switched reluctance motors (MCSRMs) for extended speed applications using 3-phase standard inverters," *IEEE Trans. Energy Convers.*, vol. 31, no. 2, pp. 436–445, June 2016.

[3] Y. Tang, "Switched reluctance motor with fractionally pitched windings and bipolar currents," in *Conference Record of 1998 IEEE Industry Applications Conference. Thirty-Third IAS Annual Meeting (Cat. No.98CH36242)*, vol. 1, Oct. 1998, pp. 351–358 vol.1.

[4] J. D. Widmer, R. Martin, C. M. Spargo, B. C. Mecrow, and T. Celik, "Winding configurations for a six phase switched reluctance machine," in *2012 XXth International Conference on Electrical Machines*, Sept. 2012, pp. 532–538.

[5] A. C. Clothier and B. C. Mecrow, "Inverter topologies and current sensing methods for short pitched and fully pitched winding SR motors," in *APEC '99. Fourteenth Annual Applied Power Electronics Conference and Exposition. 1999 Conference Proceedings (Cat. No.99CH36285)*, vol. 1, Mar. 1999, pp. 416–423 vol.1.

[6] M. A. Kabir and I. Husain, "Concentrated winding segmented rotor switched reluctance machine (SRM) using three-phase standard inverters," in *2015 IEEE Energy Conversion Congress and Exposition (ECCE)*, Sept. 2015, pp. 5567–5572.

[7] S. Mehta, M. A. Kabir, I. Husain, and P. Pramod, "Modeling of mutually coupled switched reluctance motors based on net flux method," *IEEE Trans. Ind. Appl.*, vol. 56, no. 3, pp. 2451–2461, June 2020.

[8] B. C. Mecrow, C. Weiner, and A. C. Clothier, "The modeling of switched reluctance machines with magnetically coupled windings," *IEEE Trans. Ind. Appl.*, vol. 37, no. 6, pp. 1675–1683, Nov. 2001.

0093-9994 (c) 2020 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 November 06,2020 at 04:57:52 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/TIA.2020.3030873, IEEE Transactions on Industry Applications

[9] W. Uddin and Y. Sozer, "Modeling of mutually coupled switched reluctance motors for torque ripple minimization," in *2015 IEEE International Electric Machines Drives Conference (IEMDC)*, May 2015, pp. 1006–1010.
[10] A. J. Visser, D. W. J. Pulle, and R. W. de Doncker, "Virtual short pitch control model for fully pitched switched reluctance machines using a three-phase inverter," in *2016 IEEE International Power Electronics and Motion Control Conference (PEMC)*, Sept. 2016, pp. 711–716.
[11] J. Dong, B. Howey, B. Danen, J. Lin, J. W. Jiang, B. Bilgin, and A. Emadi, "Advanced dynamic modeling of Three-Phase mutually coupled switched reluctance machine," *IEEE Trans. Energy Convers.*, vol. 33, no. 1, pp. 146–154, Mar. 2018.
[12] W. Uddin and Y. Sozer, "Analytical modeling of mutually coupled switched reluctance machines under saturation based on design geometry," *IEEE Trans. Ind. Appl.*, vol. 53, no. 5, pp. 4431–4440, Sept. 2017.
[13] J. M. Kokernak and D. A. Torrey, "Magnetic circuit model for the mutually coupled switched-reluctance machine," *IEEE Trans. Magn.*, vol. 36, no. 2, pp. 500–507, Mar. 2000.
[14] T. W. Jackson, "Analysis and design of a novel controller architecture and design methodology for speed control of switched reluctance motors," Ph.D. dissertation, Virginia Tech, 1996.
[15] Q. Ma, A. EL-Refaie, and J. Lai, "Small-Signal modeling and speed controller design for switched reluctance motor drives," in *2018 IEEE Energy Conversion Congress and Exposition (ECCE)*, Sept. 2018, pp. 4497–4504.
[16] S. Mehta, I. Husain, P. Pramod, and M. A. Kabir, "Small signal model of mutually coupled switched reluctance motors based on net flux method," in *2019 IEEE Energy Conversion Congress and Exposition (ECCE)*. IEEE, Sept. 2019, pp. 7105–7112.
[17] G. J. Li, Z. Q. Zhu, X. Y. Ma, and G. W. Jewell, "Comparative study of torque production in conventional and mutually coupled SRMs using frozen permeability," *IEEE Trans. Magn.*, vol. 52, no. 6, pp. 1–9, June 2016.
[18] S. Mehta, P. Pramod, and I. Husain, "Analysis of dynamic current control techniques for switched reluctance motor drives for high performance applications," in *2019 IEEE Transportation Electrification Conference and Expo (ITEC)*, June 2019, pp. 1–7.



Siddharth Mehta (Student Member, IEEE) received the Bachelor of Technology in Electrical and Electronics Engineering from Vellore Institute of Technology-Chennai, India in 2015 and his Master of Science in Electrical Engineering from North Carolina (NC) State University, Raleigh, NC, USA in 2017. He is currently working towards his Ph. D. in Electrical Engineering from NC State University. He worked with GE Healthcare on medical devices in Bangalore, India; and designed and tested switched reluctance motor drives for steering applications at Nexteer Automotive Corporation, Saginaw, MI, USA. He is currently a Graduate Research Assistant with the Future Renewable Electric Energy Delivery and Management Systems Center, Raleigh, NC. His research interests include modeling and control of electric motor drives and electric machine design. He is a member of IEEE, IEEE-IAS, and reviewer of IEEE journals on electric machines and drives.



Prerit Pramod (Senior Member, IEEE) received his Bachelor of Engineering degree in Electrical Engineering from the Delhi College of Engineering, University of Delhi, India in 2010 and his Master of Science in Electrical Engineering Systems from the University of Michigan, Ann Arbor, Michigan, USA in 2012. He is currently an Engineering Manager in the Advanced Systems and Software Engineering group at Nexteer Automotive Corporation, Saginaw, Michigan, USA, where he is responsible for development of advanced motion control systems. His research interests include the design, modeling and sizing of electrical machines and power converters and the development and analysis of control, estimation and optimization algorithms for electric motor drive systems.



Iqbal Husain (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Texas AM University, College Station, TX, USA, in 1993.
He is the Director with the FREEDM NSF Engineering Center and the ABB Distinguished Professor with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, USA, where he also serves as the Director with the Advanced Transportation Energy Center. Prior to joining the North Carolina State University, he was with the University of Akron, where he built a successful power electronics and motor drives program. He was a Visiting Professor with Oregon State University, Corvallis, OR, USA, in 2001. His expertise is in the areas of power electronics, electric machines, motor drives, and system controls. In the FREEDM center, his research is focused on power electronics integration into power systems and transportation electrification. He has also developed innovative graduate and undergraduate courses on electric and hybrid vehicles and published a textbook on the subject.
Dr. Husain was the recipient of the 2006 SAE Vincent Bendix Automotive Electronics Engineering Award, the 2004 College of Engineering Outstanding Researcher Award, the 2000 IEEE Third Millennium Medal, and the 1998 IEEE-IAS Outstanding Young Member award. He was the Distinguished Lecturer of the IEEE Industry Applications Society for 2012-2013. He is the Editor-in-Chief for the IEEE ELECTRIFICATION MAGAZINE.



Md Ashfanoor Kabir (Member, IEEE) received his Bachelor's and Master's degrees in Electrical and Electronic Engineering (EEE) from Bangladesh University of Engineering and Technology (BUET), Bangladesh in 2009 and 2012, respectively. From 2009 to 2012, he served the Department of EEE, BUET as a lecturer. In 2017, he received his Ph.D. degree in electrical engineering from the department of Electrical and Computer Engineering at North Carolina State University, Raleigh, NC, USA. From 2017 to 2020, he was with ABB Inc. working as a research scientist at their US Corporate Research Center in Raleigh, NC, USA. Currently he is working as a motor design engineer at TurnTide Technologies (formerly Software Motor Company) in Sunnyvale, California.
His research interests include design, modelling and control of electric machines using variable speed drives with mechanical stress, vibration and noise analysis. He focuses on high performance, high reliability designs for industrial and traction, ultra-high speed machines with WBG drives, and integrated motor drives. He has received the NCSU graduate fellowship, ABB fellowship, NCSU research symposium award and grand prize in Appalachian state energy conference. He is a member of IEEE, IEEE-IAS, and reviewer of IEEE and IET journals on electric machines and drives.

0093-9994 (c) 2020 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 November 06,2020 at 04:57:52 UTC from IEEE Xplore. Restrictions apply.

Soil Nailing for Highway Construction

Archita Goyal[1] and **A. K. Shrivastava**[2]
[1] Department of Civil Engineering, Ajay Kumar Garg Engineering College,
27 Kmstone, NH-24, Adhyatmik Nagar, Ghaziabad 201009 UP India
[2] Department of Civil Engineering, Delhi Technological University,
Shahbad Daulatpur, Main Bawana Road, Delhi 110042 India
[1]goyalarchita@akgec.ac.in, [2]aksrivastava@dce.ac.in

Abstract -- **In the construction industry, slope or excavation failure is a common issue and to avoid its risk on human lives and properties, Soil nailing technique is widely used since last two decades. It consists of solid or hollow steel bars which behave as passive reinforcement encased in grout at an inclination to the horizontal into the soil with definite horizontal and vertical spacing. In general, when compared to other traditional stabilization methods like retaining walls, the soil nailing system has the advantages of lower cost, quicker construction procedure and less impact on the adjacent ground structure. To study about soil nailing, its important parameters need to be reviewed. This paper presents applications of soil nailing and some previous researches for improvement of slope stability and stabilization of a vertical cut.**

Keywords: Soil nailing, Slope failure, Slope stabilization, Reinforcement techniques,

I. INTRODUCTION

SOIL Nailing is widely used in civil engineering. In early 1970s, concrete gravity and semi-gravity type earth retaining structures were in vogue to retain the earth structures in highway projects [1]. If excavation support system is required to retain the soil slope then internal bracing for excavation is needed. For this, numerous alternative wall-retaining system have been introduced. In highway projects many of these earth retaining systems like mechanically stabilised earth (MSE) wall, reinforced soil slopes (RSS); anchored wall, diaphragm wall, sheet pile; and *in-situ* reinforced wall system such as soil-nailed wall and micropile walls etc. are used O'Rourke and Jones [2] suggested changes and development in the earth retaining system in highway construction during period 1970-1990. According to them due to growth in underground construction, there is a requirement for advances in construction material and equipment for anchored wall system, advances in the use of polymeric material for MSE wall construction and the use of soil nailing technique to support the excavation. There has been growing demand of these earth retaining systems in highway projects during 25-year period after 1970 and they lead to cost and time savings.

Over the past few years, increasing demand of soil nailing in highway and railway construction was observed particularly in hilly regions. There is large scope of soil nailing in road widening projects due to less impact on moving traffic vehicles. The term 'soil nailing' comes from New Australian Tunnelling Method, where rock bolts were applied to support underground excavation in rock followed by shotcrete. In soil nailing method, closely spaced tendons are inserted in soil slope at some inclination to horizontal to restrain the soil or improve the shear strength of soil. These tendons known as nails behave as passive reinforcement to the soil structure. Origin of technology, types of construction, construction process and design methods, etc. are discussed in this paper.

II. ORIGIN OF SOIL NAILING TECHNOLOGY

Soil nailing has been widely used during last 30 years in various civil engineering projects as it proved to be more cost effective and time saving method than other conventional methods of retaining the earth structure. The origin of soil nailing comes from rock bolting in tunnelling (NAT method) in 1960 and then extending to relatively less cohesive material like soil [3]. After its application in rock, the nailing method extended in less cohesive materials like silt, gravels and sands with application in metro tunnels during 1970 in Frankfurt and in construction of a double tube subway station in Nuremberg. Use of soil nailing technique has been growing rapidly in Europe and North America since the early 1970s. In 1972, the first recorded use of soil nailing was in Europe for an 18 m high cut slope in sand as part of a railway widening project near Versailles, France [4]. The first major research project titled 'Boden-vernagelung' with 'Boden' meaning soil and 'Vernagelung' meaning nailing, thus the technical word soil nailing comes in application and undertaken in Germany from 1975 to 1980 by the University of Karlsruhe and the construction company Bauer [5]. In 1976, U.S. first used soil nailing for the support of a 13.7-meter deep foundation excavation for an extension of Good Samaritan Hospital in Portland, Oregon [6]. In 1986, a 4-year research project named CLOUTERRE was initiated in France to promote soil nailing for temporary and permanent soil nailed walls in excavation involving private and public participants with special emphasis on safety and durability [7]. Meanwhile the new technique soil nailing became popular in North America [8] with the name of 'lateral earth-support system' [9, 10], United Kingdom [11], Australia [12] and Hong Kong [13], Brazil [14], and Japan [15]. The use of soil nailing was reported as a

permanent retaining wall for subway underneath a busy national highway in India [16]. In India, soil nailing technique is also rapidly increasing and design guidelines have been formed by Indian Road Congress (IRC) with the help of Indian Institute of Science, Bangalore [17] wherein till date guidelines given by FHWA are adopted.

III. DESIGN APPROACH OF SOIL NAILING

Soil-nail retaining system are designed to satisfy the requirement of stability, serviceability and durability of their design life. Other issues like the cost, time and the environmental impacts are also important in the design of the soil retention systems. In design approaches, the initial design considerations include wall layout (wall height and wall length), vertical and horizontal spacing of nail, the inclination of nail, orientation of nail and nail length, soil nail material and relevant ground properties. After considering all these parameters, the design engineer has to be evaluate the nail length and maximum nail force. Nail length, nail diameter, and nail spacing typically control external and internal stability of the wall. These parameters could be adjusted during design until all external and internal stability requirements are met. After the initial design is completed, final design progresses where the soil nail wall is tested for external and internal failure modes, seismic considerations and aesthetic qualities. Drainage, frost penetration and external loads such as wind and hydrostatic forces also have to be determined and included in the final examination of the design.

IV. ELEMENTS OF SOIL NAILING

The typical soil-nail system consists of the following basic elements:

(*a*) Reinforcement bar, (*b*) centralizer, (*c*) grout, (*d*) nail head connecters (hex nut, washer and plate), (*e*) facing and (*f*) drainage.

Reinforcement Bar or Nail: Soil nail is a structural member consisting of passive reinforcement used to resist the shear, bending and tension forces. These reinforcements are generally known as nails. They can be hollow or solid type of nails. Solid nails are used when soil conditions are stable to drill hole and grouting, when soil is not able to withstand drilling then hollow nails are used with drilling and grouting process simultaneously. These nails are placed close to one another at some spacing when construction proceeds from top to down. These nails can be driven or pre-bored and grouted type depending on the soil condition. Soil nail should also be corrosion protected. Grouting acts as lower level of corrosion protection in hollow soil nails but in case of solid nails when driven, there is tendency of corrosion when nails get contact with soil and that depends on the corrosion potential of surrounding soil. So, the composite form of grouted, corrosion protected nail is highly dependent on the *in-situ* condition and in turn governs the installation process for these soil nails. Hence based on different installation techniques, soil nails can be classified as: grouted soil nails, self-drilling soil nails, jet-grouted soil nails and driven nails.

Facing: Facing of soil nailing wall acts as an additional support to prevent outward deformation of soil-nailed system. It is also used to improve the aesthetic appearance of the nailed structures. Facing material used should be of non-corroded type as it is exposed to atmosphere. Facing can be temporary or permanent. The facing consists of welded wire mesh carried by shotcreting of cement water and aggregate of fixed ratio as permanent facing. Shotcrete may be wet shotcrete or dry shotcrete. Dry shotcrete is preferred for small construction work. For vertical slope, thicker facing is used in comparison of inclined slope.

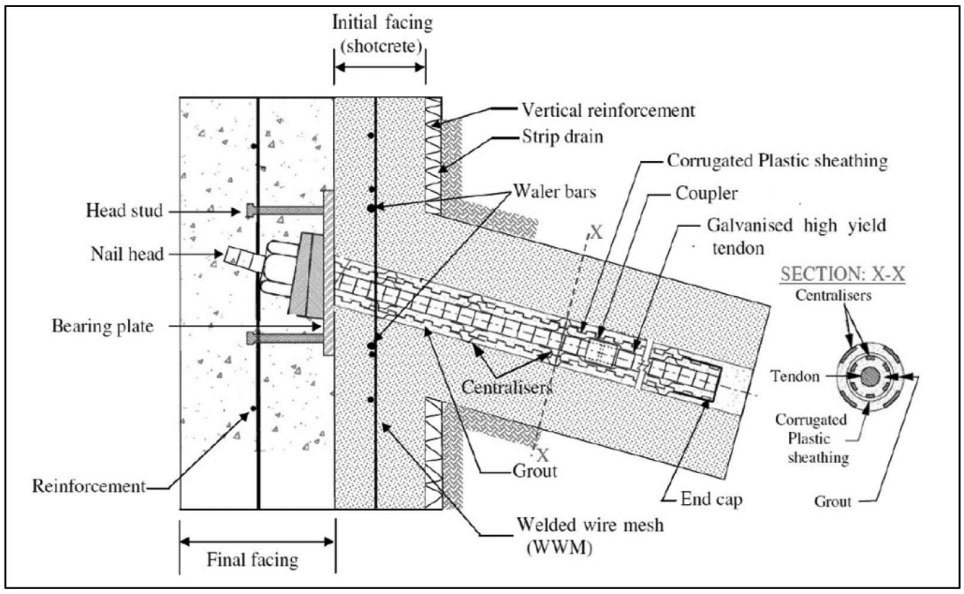


Figure 1. Typical cross-section of a complete soil nail wall.

Grout: When nail is placed in the borehole then to fill the annular spacing between the reinforcement bar and surrounding soil, the grout is injected into the borehole followed by centralizers. Grout acts as corrosion protection to the soil and transfers stresses to the soil to form stabilized bond between nails and soils. Moisture present in soil can corrode the nail by time but with the use of grouting, it prevents nails from corrosion. For grouting neat cement and water mix is used; sometimes cement, sand and water mix also used for the open-hole drilling. Pipe is used for injecting grout to the borehole.

Drainage system: It is necessary to provide a drainage system behind initial facing or adjacent excavation face. It regulates the pore water developed in soil slope and protects the facing and soil nail bonding. Usually vertical geocomposite strip drains are used as a drainage system.

V. ADVANTAGES AND DISADVANTAGES OF SOIL NAILING

Soil nailing technique has various merits and demerits addressed by Geoguide 7, FHWA-NHI-14-007 [2] and by other researchers as follows:

Advantages of soil nailing

- Soil nailing technique is more advantageous than other conventional techniques of slope stabilization *i.e.* retaining wall, ground anchors etc., as it is less disruptive to traffic and causes less environmental impact.
- Soil nailing is top down construction method so relatively faster in installation and uses less construction material.
- Soil nailing method requires less space for movement of equipment as they are comparatively small so can be advantageous for sites at remote location.
- Soil nailing technique does not use bracing to support excavation unlike other techniques so causes less congestion to excavation.
- Soil nailing technique is also well performed in seismic region.
- Facing of soil nailing gives good aesthetics to the structure and more natural appearance than other finishes.
- From cost point of view convention soil nails are more economical than concrete gravity walls and equivalent in cost or more cost-effective than ground anchor walls.

Disadvantages of soil nailing

- Soil nailing method is not recommended where strict wall movement exists, post tensioning of soil nails can be used for this shortcoming but it adds extra cost to the project.
- Soil nailing technique cannot be preferable near ground water condition. High ground water table causes difficulty in excavation during construction and ground water corrodes the reinforcement and grout water ratio disturbs.
- Soil nailing is not suitable for loose soil and cohesionless soil because during the drilling process, un-grouted holes can collapse.
- Underground drainage system, underground cables, pipes get disturbed during soil nailing process. So, for that proper care should be taken during installation of nails.

VI. APPLICATIONS OF SOIL NAILING

Soil nailing method has various applications:

- Slope stabilization
- Landslide prevention
- Deep excavation or vertical cut stabilization
- Stabilization of roadway and highway cuts
- Road widening under existing bridge abutments
- Tunnel portal in steep stratified and unsupported slopes
- Repair and reconstruction of existing retaining structures
- Use for deep vertical excavation in densely populated areas.
- Stabilizing of over steep existing embankments
- Protection and preservation of historical buildings.

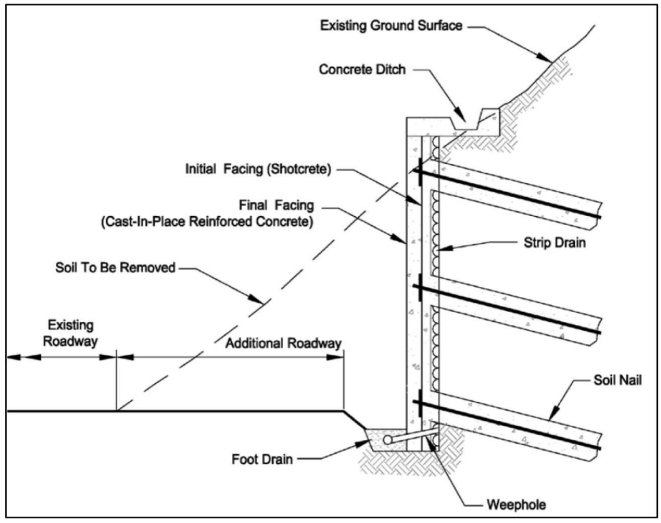


Figure 2. Roadway cut supported with soil nails. Modified after Porterfield *et al*. [25].

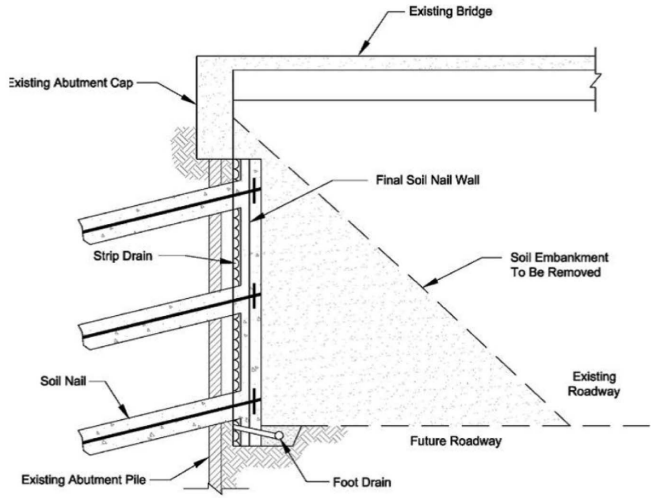


Figure 3: Road widening under existing bridge *et al.* [25]
Source: Porterfield

VII. ADDITIONAL REMARKS ON SOIL NAILING

According to research studies, a nail slope is a composite structure with elements of soil, nail, grout and wall facing. All these parameters affect the stability of soil nailed structure so they need to be checked before the construction proceeds. These parameters are briefly described in this section to study the characteristics and mechanics of soil nailing.

Jewell [18] investigated the effects of orientation of reinforcement bars on the shear strength of soil in a shear box. To maximize the shear strength of soiltrs the reinforcement should be placed in direction of principal tensile strain in soil and if the reinforcement is in direction of the soil compressive strain, there seems to be a decrease in the shear resistance of the soil.

Long *et al.* [19] studied the effect of number of nails on the factor of safety and concluded that it had a great effect on the failure surface location, and showed that as the nail spacing decreased, the factor of safety increased.

Schlosser *et al*. [20] using TALREN software stated that displacement of the wall increases by decreasing the global factor of safety while the later increases by increasing the nail length. Byrne *et al*. (1998) worked on the use of uniform or variable nail length and suggest that uniform nail lengths are used in areas where excessive deformations are not a consideration or concern because of competent ground conditions and absence of nearby structures sensitive to ground movements. Whereas If the wall deformation is critical, then a variable nail system can be selected.

Güler and Bozkurt [21] conducted soil stability analysis using TALREN software and showed that nails installed at angles above horizontal gave more factor of safety and less lateral displacement than nails installed at angle below horizontal.

Mittal and Biswas [22] showed that for cohesion-less soil as the nail length increased, the factor of safety increased. The nail length upto 0.8 times the height of cut is a reasonable length for providing a stable cut, minimum nail length of 0.7 m performs well in field.

Hossain and Islam [23] show that with the increase of inclination, the factor of safety of the slope increases, but after reaching the maximum value further increases of inclination, it decreases. The optimum angle of the nail was found at 30° with the horizontal.

Singh and Shrivastava [24] stated that nails inserted at 0° are more efficient in providing the stability to the slopes as compared to nails inserted at 15° and 30°. Also, nails installed in staggered pattern were found to be most efficient.

REFERENCES

[1] FHWA, "*Geotechnical Engineering Circular No. 7: Soil Nail Walls*", Report No. FHWA0-IF-03- 017, Federal Highway Administration, Washington, USA, 2003.

[2] T.D. O'Rourke and C.J.F.P. Jones, "Overview of Earth Retention Systems: 1970-1990, Design and Performance of Earth Retaining Structures, *American Society of Civil Engineers, Geotechnical Special Publication* No. 25, pp. 22-51, 1990.

[3] L.V. Rabcewicz, "The new Austrian tunneling method", *Water Power*, vol. 16, no.11 pp. 453-457, Nov. 1964.

[4] S. Rabejac and P. Toudic, "Construction of a retaining wall between Versailles-Chantiers and Versailles - Matelots", *Revue générale des chemins de fer, French railway review no. 93*, pp. 232 – 237, 1974.

[5] M.D. Stocker, G.W. Korber, G. Gassler and G. Gudehus, "Soil nailing", *Proc. Int. Conf. Soil Reinfocement*, Paris, vol. 2, pp. 469–474, March 1979 .

[6] FHWA, "Manual for design and construction monitoring of soil nail walls", *FHWA-SA-96-096R*. U.S. Department of Transportation, Federal Highway Administration, Washington, D.C., 1998.

[7] R. Clouterre, "Soil nailing recommendation 1991", An English translation of 'Recommendations CLOUTERRE — 1991, USA Report No. FHWA-SA-93-026, US Department of Transportation, Federal Highway Administration, Washington DC, 1991.

[8] C.K. Shen S. Bang and L.R. Herrman, "Ground movement analysis of earth-support system", *J. of Geotech. and Geoenv. Engg.*, ASCE 107 (GT12), pp. 1609-1624, 1981.

[9] V. Elias and I. Juran, "Soil Nailing for Stabilization of Highway Slopes and Excavations", Publication FHWA-RD-89-198, Washington D.C., Federal Highway Administration,1991.

[10] FHWA, "Soil Nailing Field Inspectors Manual - Soil Nail Walls", *Report No.FHWA-SA-93-068*, Federal Highway Administration, Washington DC, 1994. pp. 665-670, 1981.

[11] R. Jewell and M. Pedley, "Analysis for Soil Reinforcement with Bending Stiffness", *Journal of Geotechnical Engineering*, vol. 118, no. 10, pp. 1505-1528, 1992.

[12] P. Seto, G.W. Won and K.Y. Choi, "Use of soil nailing in stabilization of a freeway embankment", Earth Reinforcement Practice: *Proc. Int. Symp. on Earth Reinforcement Practice*, Fukuoka, Kyushu, Japan, pp. 11-13, 1992.

[13] A.T. Watkins and G.E. Powell, "Soil nailing to existing slopes as landslip preventive works", *Hong Kong Eng*, vol. 20, no. 3, pp. 20-27, 1992.

[14] J.A.R. Ortigao, E.M. Palmeira and A.C. Zirlis, "Experience with soil nailing in Brazil:1970-1994", *Proc. Int. Civil Engineer, Geotechnical Engineering*, vol. 113, pp. 93-106, 1995.

[15] M. Hirano, "Actual status of the application of Soil Nailing to expressway cut slope construction in Japan", Landmarks in earth reinforcement. *Proc. Int. Earth Reinf*.,Kyushu, Japan, pp. 919-934, 2001.

[16] B.R. Srinivasa Murthy, G.L. Sivakumar Babu and A. Srinivas, "Analysis of prototype soil - nailed retaining wall", *Proc. Institution of Civil Engineers-Ground Improvement*, vol. 6, no. 3, pp. 129-136, 2002.

[17] G.L. Siva Kumar Babu and Vikas Pratap Singh, "Appraisal of Soil Nailing Design", *Indian Geotechnical Journal*, vol. 39, no. 1, pp. 10–17, 2009.

[18] R.A. Jewell, "*Some Effects of Reinforcement on the Mechanical Behaviour of Soils*", PhD thesis, University of Cambridge, 1980.
[19] J. Long, "Stability analyses for soil nailed walls", *Geotechnical Special Publication*, vol. 25, pp. 676-691, 1990.
[20] F. Schlosser, P. Unterreiner and C. Plumelle, "French research program CLOUTERRE on soil nailing", *Proc. 1992 ASCE Specialty Conference on Grouting, Soil Improvement and Geosynthetics*, 25-28 Feb. 1992, New Orleans, LA, USA.
[21] E. Güler and C. Bozkurt,"The effect of upward nail inclination to the stability of soil nailed structures", *Geotechnical Engineering for Transportation Projects*, American Society of Civil Engineers: pp. 2213-2220, 2004.
[22] S. Mittal and A.K. Biswas, "River bank erosion control by soil nailing", *Geotechnical and Geological Engineering*, vol. 24, pp. 1821–1833, 2006.
[23] A. Hossain and A. Islam, "Numerical analysis of the effects of soil nail on slope stability", *International Journal of Computer Applications*, vol. 141, pp. 0975-8887, 2016.
[24] S. Singh and A.K. Shrivastava, "Effect of soil nailing on stability of slopes", *International Journal for Research in Applied Science & Engineering Technology*, vol 5, Issue X, pp. 752-763. 2017.
[25] J.A. Potterfield, D.M. Cotton and R.J. Byrne, Soil Nailing Field Inspectors Manual FHWA Report SA-93-068, 1994.



Archita Goyal is pursuing Ph.D in Geotechnical Engineering from Delhi Technological University, Delhi. Obtained M.E. in Geotechnical Engineering (Civil Engineering) in 2014 with Thesis topic, 'Stabilization of Dune Sand Using Bentonite, Cement and Fiber'. Obtained B.Tech. in Civil Engineering from Rajasthan Technical university, Kota in 2010.

Working as Assistant Professor (CE) in AKGEC, Ghaziabad since January 2018. Her previous teaching assignments were at KEC, Ghaziabad, SIRT, Bhopal and JNU, Jaipur. Earlier she worked as Senior Engineer in Larsen & Toubro Ltd. (ECC), Mumbai.
Taught subjects of Geotechnical Engineering, Advance Foundation Design, Geographic Information System, Open Channel Flow, Ground Water Management, Fluid Mechanics, Hydraulics & Hydraulic Machines and Human Values. Organized Guest lectures and workshops.



Dr. Amit Kumar Shrivastava, B.E, M.E, PhD (IIT Delhi), LMIGS, LMISRMTT, MISMGE, has over 23 years of experience. Currently working as Associate Professor, Department of Civil Engineering, Delhi Technological University, Delhi. Areas of interest include geotechnical engineering, strength and deformation behavior of rocks and rock masses, pavement engineering and modern construction practices.

Published over 40 research papers. He is in the panel of Reviewer for International Journals. Supervised one PhD, 25 ME, 30 BE projects besides guiding 3 PhD research scholars. Consultant for various projects of Public Works Department Delhi, Central Public Works Department, New Delhi Municipal Corporation, Delhi Metro Rail Corporation, Municipal Corporation of Delhi, Delhi State Industrial Development Corporation and others. Involved as a Foundation consultant for North India Highest residential building project BRYS and also advisor to expert Technical Committees of Delhi Government, Public Service Commissions, National Institute of Technology and PSUs.

Received appreciation letter from UNESCO for contribution as convener of International workshop on Climate Change and Water Resources in South Asia in August 2010. Recipient of Biennial IGS-H.C. Verma Diamond Jubilee award for Innovative Instrument Design for Year 2010-2011. Delivered key note lecture in Asian Cement Summit-2011. Presented papers at Geo-Congress -2012, California, USA and in EUROCK-2014, Spain in 2014.

2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)
Galgotias University, Greater Noida, UP, India. Oct 2-4, 2020

Speed and Current Controller Modelling for Close Loop Control of Separately Excited DC Motor for Constant Flux and Flux Weakening Operation

Varij Kumar
Electrical Engineering Dept.
Delhi Technological University Shahbad daulatpur
Delhi, India
varijkumar1192@gmail.com

Saurabh Mishra
Electrical Engineering Dept.
Delhi Technological University Shahbad daulatpur
Delhi, India
saurabh.mishra1410@gmail.com

***Abstract*—aforementioned title explains the essence of this paper. This paper revolves around the modelling of controllers (speed/current) of separately excited dc motoring MATLAB/Simulink. There is also mathematical simplification of frequency domain calculations of speed and current in terms of various parameters. Percentage overshoot, settling time and steady state value of speed and current are calculated with input load torque and reference speed variation. Also the flux weakening is done in order to achieve constant power operation at above rated speed.**

Keywords—Symmetric optimum method, controller, sensor, Simulink, gain, time constant, flux weakening.

I. INTRODUCTION

DC MOTOR is an electrical machine, which finds wide range of applications where precision and speed control is required. It can be operated in comparatively easier manner as contrary to ac motors. By varying the voltage and armature current in an open loop control, the DC motor can be operated at any arbitrary speed. The demerit of open loop control resides in inability of getting a prediction based or controlled accuracy that is why speed will be different for load variations on the motor. Due to load variation, there is no feedback, which represents the change in speed. This disadvantage hinders its use as open loop speed control DC motor in areas where constant speed is necessity.

In order to remove this demerit, a closed loop technique is introduced where the tachometer-measured speed is fed back to the speed controller. The speed can be maintained by adjusting terminal armature voltage in accordance with the speed difference caused by the variation in load torque, which results in a fine control, of speed, which is user friendly in a closed loop controller.

II. MODELLING OF DC MOTOR

Separately excited DC machine is governed by equations as mentioned

$$V_a = E_a + i_a R_a + L_a \frac{di_a}{dt} \quad (1)$$

$$T_m = T_L + J\frac{d\omega}{dt} \quad (2)$$

Back emf, $E_a = K_f \omega$ (3)

Torque in the machine is proportional to the flux and the current which is governed by the equation as

$$T = K_f i_a \quad (4)$$

Solving Eq.1 in frequency domain

$$i_a = \frac{Va - Ea}{Ra + sLa} \quad (5)$$

Analysis of moment of inertia and the losses occurred as mechanical losses in electrical machinery plays a major role in improving the performance of the machine and is analyzed via various methods.

Since the exact dimensions of the electrical machine is not in domain of an electrical engineer, the electrical calculation of moment of inertia is done by performing some tests namely torsion oscillation test, pendulum test, retardation test etc. retardation test is a common way of doing the aforementioned.

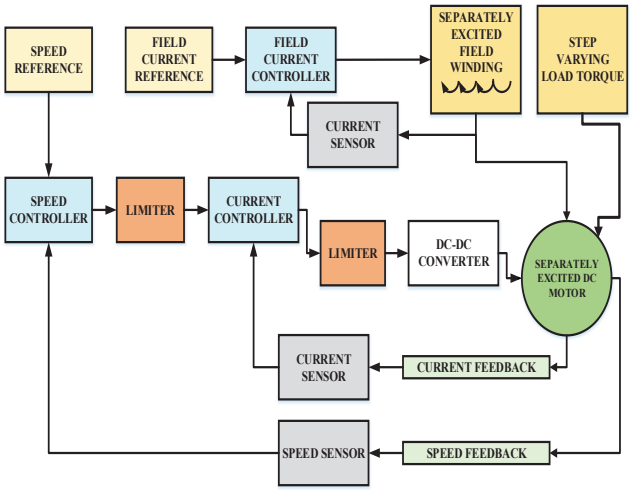


Fig. 1. Model of Closed Loop Speed Control Of Separately Excited Dc Motor

Let N=normal speed in RPM

W=normal angular velocity in rad/s=2πN/60

Rotational losses W=rate of loss of KE of armature

$$\text{Or } W = \frac{d}{dt}\left[\frac{JW^2}{2}\right] = J * w * \frac{dw}{dt} \quad (6)$$

For determining the moment of inertia the machine is run at a speed which is greater than rated speed, and then it is allowed to slow down to stand still. Then the speed versus time in speed dropping time is plotted so as to obtain differential value.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 06:45:39 UTC from IEEE Xplore. Restrictions apply.

This test can be performed on both shunt and dc separately excited motor.

III. MODELLING OF CONTROLLERS AND SENSORS

A. Controller and its significance

An Electrical controller is defined as an electrical instrument employed to minimize the difference between the desired system output and the actual obtained output. The important uses of the controllers are written below:

- Improvement in the steady state accuracy, which is obtained by decreasing the steady state error.
- Complete control over maximum overshoot generated by the system output.
- Reduction in the undesirable offsets produced by the system.
- Enhancement in speed of response of an over damped system.
- Reduction in the noise signals observed in the system channels.
- As soon as steady state accuracy improves, the stability is regained.

B. Transfer function of controller and their respective sensors

According to the nomenclature, Proportional and Integral Controller is a combination of proportional and an integral controller and the output or actuating signal is equal to the summation of proportional and integral of the error signal.

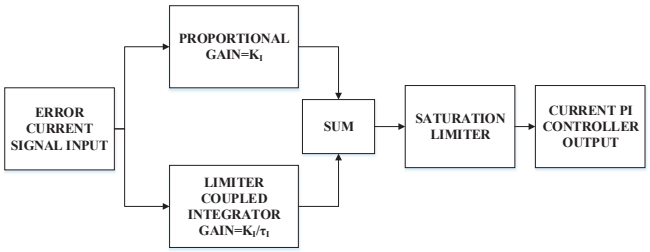


Fig. 2. Current PI controller

The controller is modelled as

$$K_p + \frac{Ki}{s} = \frac{Kps+Ki}{s} = \frac{Kp}{Ki} K_i \frac{\left(1+\frac{Kps}{Ki}\right)}{s\frac{Kp}{Ki}}$$

Time constant $\tau = \frac{Kp}{Ki}$

Gain K = K_p

PI Controller Block = K $\frac{(1+s\tau)}{s\tau}$

Now for Current PI Controller=KI$\frac{(1+s\tau_i)}{s\tau_i}$ (7)

For Speed PI Controller=$K_\omega \frac{(1+s\tau_\omega)}{s\tau_\omega}$ (8)

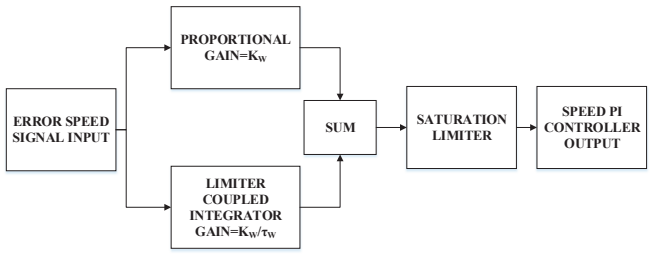


Fig. 3. Speed PI Controller

A current sensor is an electrical device that generates a voltage indicated signal by sensing the electrical current in a wire and the voltage signal generated has amplitude that is directly proportional to the current flowing in the wire. Moreover, a magnetic field is generated near a current carrying conductor. It is important to be noted that all sensors are modelled as first order Transfer function which is quite a precise approximation.

$$\text{Current sensor} = \frac{K_2}{1+s\tau_2} \quad (9)$$

$$\text{Speed sensor} = \frac{K_1}{1+s\tau_1} \quad (10)$$

C. Derivation Of Current Controller Gain And Time Constant Using Mathematical Simplification

Back emf, E is feedforwarded before the power electronics converter in order to reduce the burden of the current controller, feedforward is denoted as

$$\hat{E} = \frac{\boldsymbol{E}}{\boldsymbol{Ka}/(\mathbf{1}+s\boldsymbol{\tau_a})} \quad (11)$$

$\tau_1 = 3\ to\ 5\ times\ \tau_2$

Assume τ_2 = 2.5ms

Pole Zero reduction for simplification

$$\text{Current Controller Time constant } \tau_1 = \frac{L_a}{R_a} \quad (12)$$

$$TF = \frac{\frac{KiKa\frac{1}{Ra}}{s\tau i(1+s\tau a)}}{1+\frac{K2}{s*\tau 2+1} X \frac{KiKa\frac{1}{Ra}}{s\tau i(1+s\tau a)}} = \frac{(1+s\tau_2)KiKa}{K_2KiKa + (1+s\tau_2)R_a s\tau_i(1+s\tau_a)}$$

Applying Symmetric optimum method for

$$(1+s\tau_2)(1+s\tau_a) = 1+s(\tau_a+\tau_2)+s^2(\tau_a*\tau_2)$$
$$= 1+s(\tau_a+\tau_2)$$
$$= (1+s\sigma) \quad (13)$$

Characteristic equation,1+G(s) H(s) =0

$$1+s\tau_2+\frac{\frac{K2KiKa}{Ra}}{s\tau_i(1+s\tau_a)}=0$$

$$s(1+s\sigma)\ \tau_i + K_2K_iK_a/R_a = 0 \quad (14)$$

Compare with $s^2+2\varepsilon w_n+w_n^2=0$ (15)

$$\omega_n = (K_2K_aK_i)/(R\ \sigma\tau_i) \quad (16)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 06:45:39 UTC from IEEE Xplore. Restrictions apply.

$$2\varepsilon\omega_n = 1/\sigma \quad (17)$$

$$\varepsilon = \frac{1}{\sqrt{2}} \quad (18)$$

$$\boldsymbol{K_i = \frac{L_a}{2\sigma K_2 K_a}} \quad (19)$$

Derivation of Current Controller Gain and Time Constant Using Mathematical Simplification

$$I_a / I_a^* = \underline{[K_i/(s\tau_i)]\,[K_a/(1+s\tau_a][1/R_a]}$$

$$[1+[K_i/(s\tau_i)]\,[K_a/(1+s\tau_a]\,[K_2/(1+s\tau_2]\,[1/R_a]] \quad (20)$$

$$= \underline{K_i K_a(1+s\tau_2)/R_a}$$

$$s^2\tau_i\sigma + s\tau_i + [K_i K_a K_2/\tau_i] \quad (21)$$

As $\sigma^2 << 1$ so

$$\mathbf{[I_a / I_a^*] = \frac{(1+s\,\tau_2)}{K_2(1+2\,\sigma s)}} \quad (22)$$

Just like E or feedback voltage was neglected in current controller gain derivation load torque can be neglected in speed controller gain derivation.

Characteristics equation=1+G(s)H(s)

$$\mathbf{G(s)H(s) = \frac{K_s(1+s\tau_s)\,(1+s\,\tau_2)\,K_f\,K_1}{s\tau_s K_2(1+2\,\sigma s)sJ(1+s\,\tau_1)}} \quad (23)$$

Zero remover block or additional pole is required to be added to improve system response $1/(1+s\,\tau_2)$.

$$\mathbf{G(s)H(s) = \frac{K_s(1+s\tau_s)K_f\,K_1}{s\tau_s K_2(1+2\,\sigma s)sJ(1+s\,\tau_1)}} \quad (24)$$

$$(1+2\,\sigma s)(1+s\,\tau_1) = (1+s\delta)$$

$$\mathbf{(\delta) = (2\,\sigma + \tau_1)} \quad (25)$$

$$\mathbf{G(s)H(s) = \frac{K_s(1+s\tau_s)K_f\,K}{s\tau_s K_2(1+s\,\delta)sJ}} \quad (26)$$

Using the idea of pole zero plotting on Bode plot it can be observed

$$\omega_c = (\tau_\omega \delta)^{-0.5} \quad (27)$$

$$\tau_\omega = 4\delta = \tau_s \quad (28)$$

$$|G(s)\,H(s)|_{s=j\omega_c} = 1 \quad (29)$$

ω_c is the crossover frequency

From eqn. 26,27,28,29

$$\mathbf{K_s = K_w = \frac{K_2 J}{2\delta\,K_1 K_f}} \quad (30)$$

D. Converter

To provide a controlled output as per the feedbacks of sensor and controller, the supply was needed to be reformed and hence DC-DC converter has to be used which is modelled in Simulink as a sequence comparator. In an attempt to change a signal from one data type to another keeping the real-world input signal value preserved real world value[RWV], the MATLAB SIMULINK default setting of convert block has been used. The limits imposed when internally inputs are scaled and attempts of generating an equal real-world value output are also in the domain of this block.

Reference sequence is modelled in the form of a rectangular pulse with time as per the converter switching frequency.

Since signal level has to be operated on per unit system and the limiters also ensure the maximum value to unity.

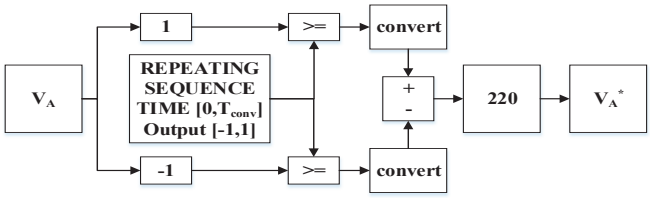


Fig. 4. Converter Subsystem

E. Flux weakening process using field PI controller

The above-discussed process was for below rated speed operation in which flux weakening process does not take place. When the speed of motor goes beyond rated speed, it enters into **constant power region operation**.

This paper introduces the flux weakening process and constant power operation of separately excited dc motor using field current proportional integral controller and sensor. Field current controller has been modelled similar to the armature current controller while current sensor is modelled as first order transfer function block diagram.

The field constant used in emf and torque calculation from speed and current respectively no longer remains constant in flux weakening process.

The two uses of field constant has been modified as per the figure 5 and 6.

Fig 5 shows the operation to obtain motor torque output from Armature current obtained from armature impedance circuit using hyperbolic relation between field current and armature current.

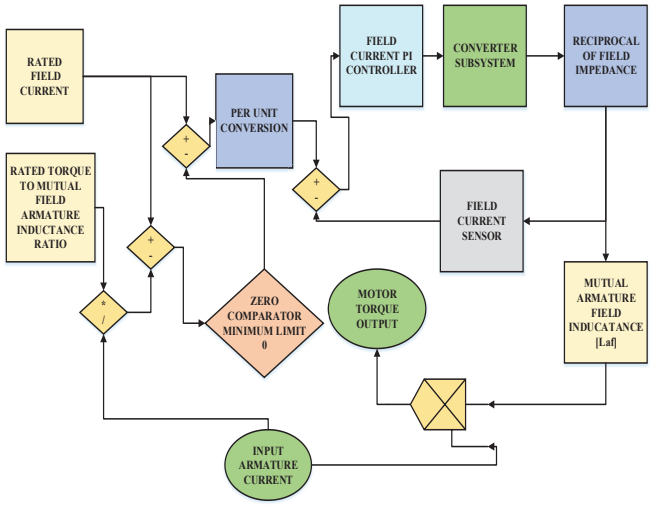


Fig. 5. Field controller fed by armature current

Fig 6 shows the operation to obtain back emf from speed sensor sensed speed using hyperbolic relation between field current and speed of motor.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 06:45:39 UTC from IEEE Xplore. Restrictions apply.

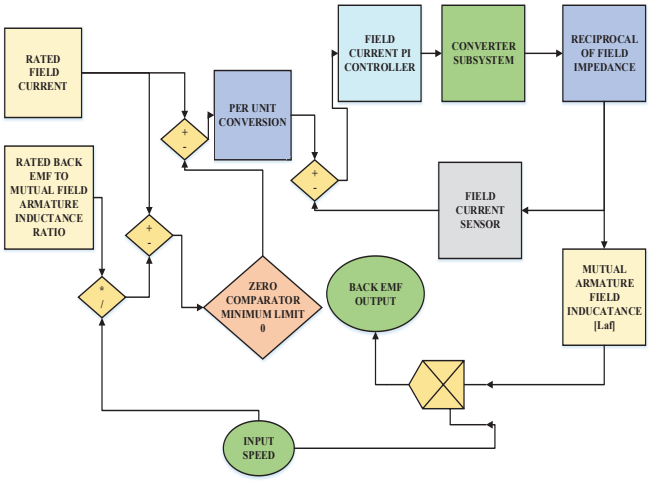


Fig. 6. Field controller fed by motor speed

IV. RESULTS

Part A: - Below Rated Speed Operation

After the entire calculations the Simulink model was drawn with the results of speed, current and torque characteristics with respect to time.

Initially motor's reference is taken as reference speed with torque as full load torque [Fig 8] as it observed slow increment in speed. After a span of 1.5 seconds load torque is reduced by 50% and within 1.2 seconds the actual speed reaches the reference speed. Since current and torque are directly proportional to each other the current curve [Fig 8] resembles similarity.

After 3.5 seconds of starting there is a braking application which implies reference speed to be reduced to zero and controller fed motor takes about 1.2 seconds to stop running.

Meanwhile converter output[Fig 7] also shows some variations as there is change in polarity at 3.5 seconds which verifies that plugging action is taking place.

TABLE I. LOAD TORQUE VARIATION

Initial value	12 N-m
Final value	6 N-m
Step time	1.5 s

TABLE II. REFERENCE SPEED VARIATION

Initial value	157 rad/s
Final value	0 rps
Step time	3.5 s(time of braking)

TABLE III. RESULTS OF CURRENT AND TORQUE

Parameter	Over Shoot	T_s	Value after T_s
Current	1.2499	0.07	11.974A
Torque	1.1538	0.07	12.561N-m
Speed	2.118	2.72(starting) 0.76(braking)	157 rad/s

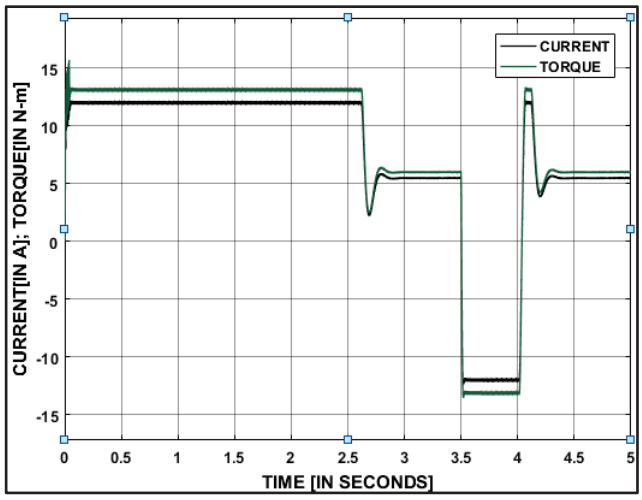


Fig. 7. Current and Torque Characteristics with Time

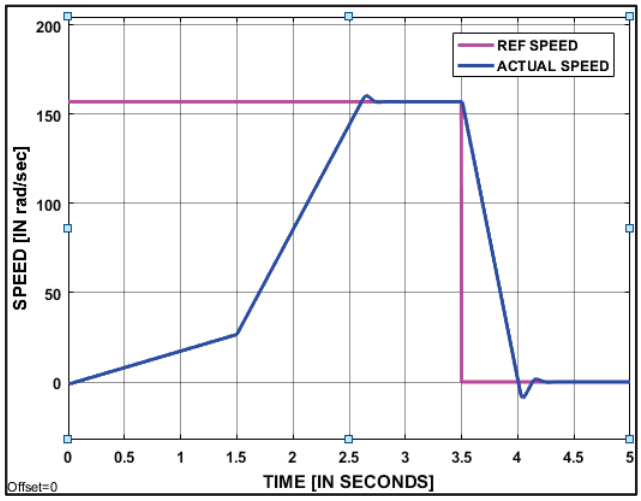


Fig. 8. Speed Characteristics with Time Compared To Ref. Speed

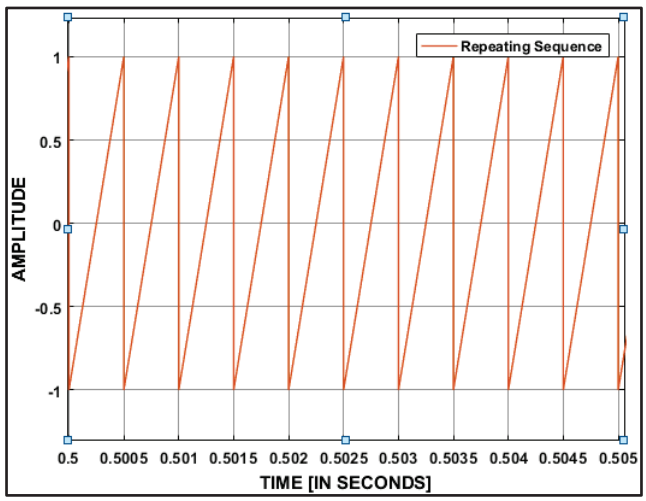


Fig. 9. Repeating sequence

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 06:45:39 UTC from IEEE Xplore. Restrictions apply.

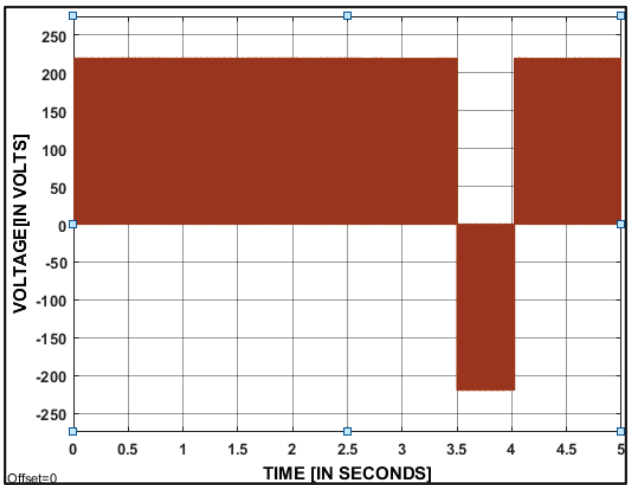


Fig. 10. Converter output

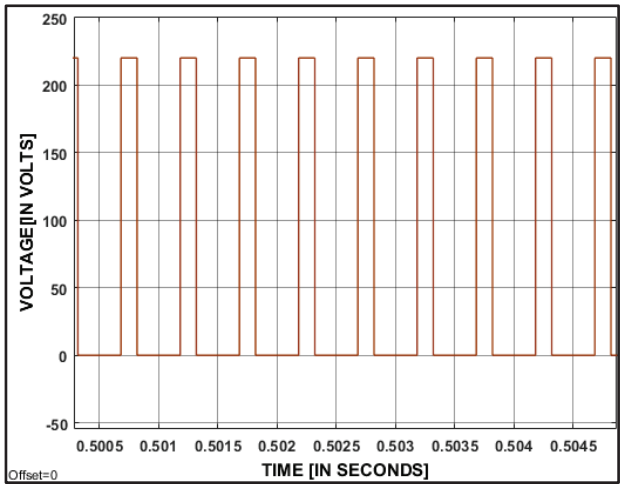


Fig. 11. Converter output [zoomed]

Part B: - Above Rated Speed or Flux Weakening Operation

TABLE IV. REFERENCE SPEED VARIATION WITH TIME

Time	Speed variation
0-10 s	Linearly increasing from 20 to 177 rad/s or 1690 rpm
10-14 s	Constant above base speed operation @ 177 rad/s
14-18 s	Linearly decreasing from 177 rad/s to 114.24 rad/s 157 crossing point 15.35 power stared decreasing
18-20 s	114.24 rad/s speed constant

TABLE V. E.M.F VARIATION OBSERVED WITH TIME

Time	EMF variation
0-8.8 s	Perturbations till 1 s but then Linearly increasing from 35 to 172 Volts
8.8-10s	Since at 8.8 s speed crosses rated speed which weakens flux making emf constant
At 10^{th} s	For only 0.6 s a slight dip ie from 172.7 V to 171 V which justifies real time variation as speed is made constant at 1690 rpm
10.6-14 s	In this period it remains constant at 171 V
At 14^{th} s	Again similar dip from 171 to 169.475 V
14.6-15.35s	As speed has started dedcreasing from 177rad/s but still above rated so emf remains constant at 169.475
15.35-18s	EMF decreasing thereby exiting constant power operation
At 18^{th} s	Since speed became constant there is a dip for 0.2 seconds
18-20s	Below rated speed operation with emf steady state value being 124.5 V

TABLE VI. POWER VARIATION OBSERVED WITH TIME

Time	Power variation
0-8.8 s	Perturbations till 1.9 s but Linearly increasing from 380 to 1210 Watts
8.8-10s	Since at 8.8 s speed crosses rated speed which weakens flux making power constant at 1210 Watts
At 10^{th} s	For only 0.25 s a dip in power ie from 1210 W to 1030 W which justifies real time variation as speed is made constant at 1690 rpm
10.25-14 s	In this period it remains constant at 1030 W
At 14^{th} s	Again similar dip from 1030 W to 860 W
14.25-15.35s	As speed has started dedcreasing from 177rad/s but still above rated so power remains constant at 860 W
15.35-18s	Power decreasing thereby exiting constant power operation and entering constant torque operation
At 18^{th} s	Since speed became constant there is a peak justifying
18-20s	Below rated speed operation with power steady state value being 750 W

TABLE VII. RESULTS OF CURRENT AND TORQUE

Q	Over Shoot	T_s	Value after T_s
I_A	2.17%[1.5^{th} s] 9.09%[18^{th} s]	At each step (10^{th} , 14^{th}, 18^{th}) stabilisation is taking place within 0.5s	From 9 to 6.5 A
T	8.33%[18^{th} s]	0.25 second	6 N-m
ω	0.42%[18^{th} s]	0.14 s	114.2rad/s

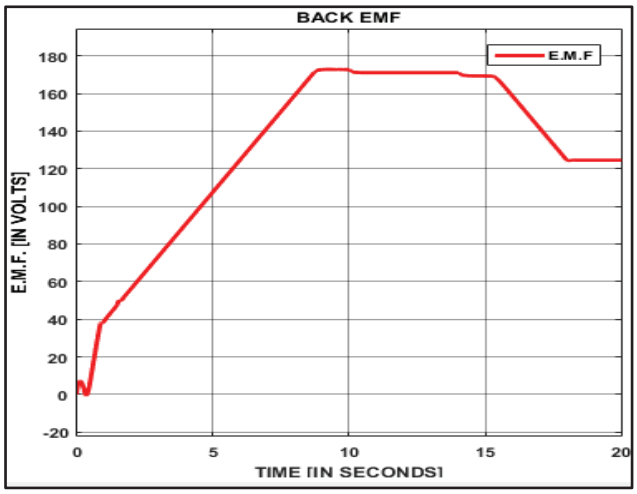


Fig. 12. Back EMF

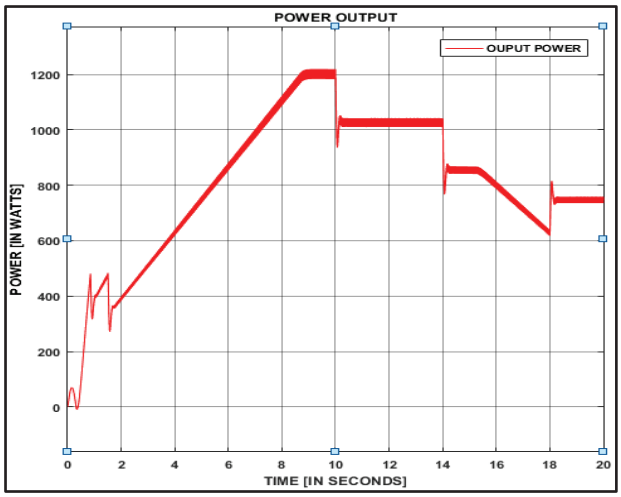


Fig. 13. Output Power

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 06:45:39 UTC from IEEE Xplore. Restrictions apply.

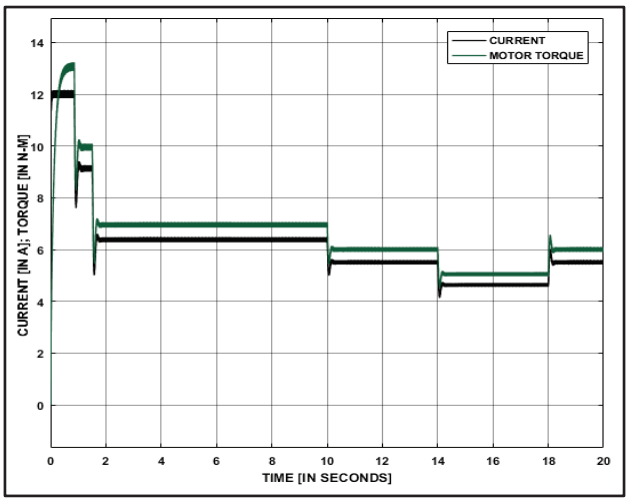


Fig. 14. Current and torque

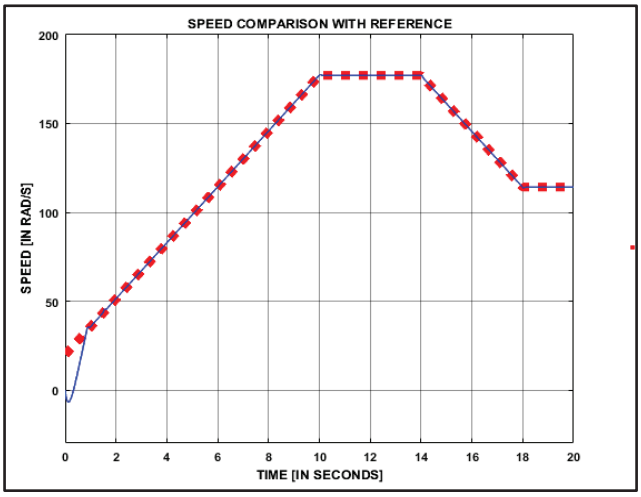


Fig. 15. Speed comparison

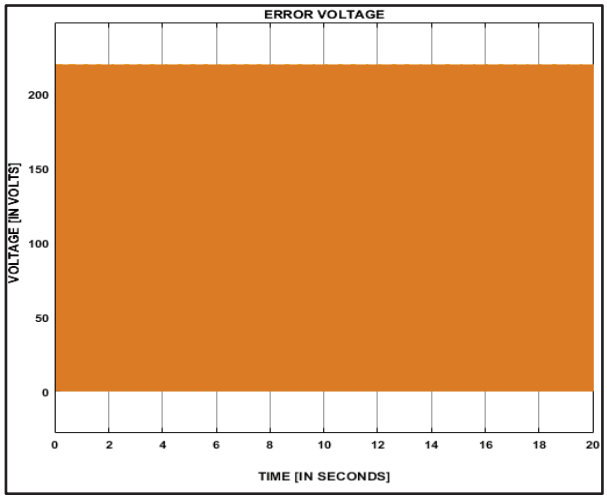


Fig. 16. Converter output

V. PARAMETERS

TABLE VIII. MACHINE PARAMETERS

Armature resistance [r_a]	5 ohms
Armature inductance [l_a]	0.072 henry
Field resistance [R_f]	110 ohms
Field inductance [L_f]	0.01 henry
Mutual armature field inductance[L_{af}]	0.545 henry
Rated field current [I_f]	2 A
Field constant [k_f]	1.09 V/(rad s^{-1})
Moment of inertia [J]	0.0607 kgm^2
Armature voltage [v]	220 V
Rated armature current [i]	12A
Rated speed	157 rad/s or 1500 rpm

TABLE IX. GAINS AND TIME CONSTANTS OF CONTROLLERS AND SENSORS

Converter gain [k_{conv}]	220
Converter time constant [t_{conv}]	0.0005 s
Armature block equivalent gain [k_a]	0.2500
Armature block equivalent time constant [t_a]	0.0180 s
Current sensor gain [k_2]	0.0833
Current sensor time constant [t_2]	0.0025 s
Speed sensor gain [k_1]	0.0064
Speed sensor time constant [t_1]	0.0088 s
σ	0.0030
Current controller gain [k_i]	0.6545
Current controller time constant [t_i]	0.0180 s
δ	0.0148
Speed controller gain [k_w]	24.6200
Speed controller time constant [t_w]	0.0590 s

ACKNOWLEDGMENT

Authors hereby acknowledge all the helping authorities including Department of electrical engineering, TEQIP, IEEE Industrial applications society and GUCON2020 that stood in various phases of writing this research paper.

REFERENCES

[1] Flux-weakening Control Strategy of Separately Excited DC Machine Utilizing Voltage Regulation for Modifying Field Current,' N. Pothi Department of Electrical Engineering University of Phayao Phayao, Thailand

[2] Design of a Flux Weakening Control Scheme for DC Motor Drives Featuring Full Voltage Operation,' S. Bolognani, Member, A. Faggion, L. Sgarbossa.

[3] Intelligent PI Controller for Speed Control of D.C. Motor Brijesh Singh, Surya Prakash, Ajay Shekhar Pandey and S.K. Sinha

[4] Modern Control Systems by Richard C. dorf

[5] Feedback Control Systems, Charles L. Philips & Royce D. Harbor

[6] Devendra K. Chaturvedi, Modeling And Simulation Of System Using Matlab And Simulink, Crc Press Taylor And Francis Group Boca Raton London New York 2010.

[7] B. Szabados, C. D. Dicenzo, N. K. Sinha, "Dynamic Measurements Of The Main Electrical Parameters Of A Dc Machine", Ieee Trans. Ind. Gen. Appl., Vol. Iga-7, Pp. 109-115, Jan/ Feb. 1971.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 06:45:39 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/TAP.2020.3027054, IEEE Transactions on Antennas and Propagation

Stochastic Filtering in Electromagnetics

Rahul Bansal, Sudipta Majumdar and Harish Parthasarthy

***Abstract*—This paper presents the estimation of electric and magnetic fields using the Kalman filter (KF). The electric and magnetic fields in the entire space have been estimated using the scalar and vector potential. For this estimation, the measurements at a sparse discrete set of spatial pixels have been used. To implement the KF, the state space model has been obtained using the wave equations with sources satisfied by the scalar and vector potential. The proposed method has been implemented on a Hertzian dipole antenna. The fields estimated using KF have been compared with the recursive least squares (RLS) method. The KF presents better estimation than RLS, as it is an optimal estimator. This work uses the Kronecker product for compact representation of discretized fields in the form of vectors and partial differential operators in the form of matrices.**

***Index Terms*—Antenna, Kalman filter, electromagnetic field, recursive estimation, Kronecker product.**

I. Introduction

ESTIMATION of electric and magnetic fields is important in various applications. Estimation of magnetic field is used for vehicle guidance and motion control applications, as positioning is based on the magnetic field sensing. Field estimation is also required for base station antenna used in mobile communication [1]. The electromagnetic devices used for diagnostic and therapeutic applications such as transcranial magnetic simulation (TMS) also require the estimation of electric field [2]. Pie *et al.* [3] used field estimation for pedestrian dead reckoning (PDR) algorithm, which calculates the user location. Field estimation is also needed for medical instruments, mine detection, power transmission fields, pipelines, telecommunication lines etc.

Various methods have been used for electric and magnetic field estimation. Belhadj and Ferik [4] used artificial neural networks for electric and magnetic field estimation of live transmission line workers. Wang *et al.* [5] presented electric field estimation in human body exposed to low frequency magnetic field that uses a boundary element procedure. Zenczak [6] proposed electric and magnetic field estimation in distributed and centralized power systems. Paffi *et al.* [7] proposed a method to calculate the electric field induced inside the brain by a TMS method. The electric field induced by TMS coils inside a brain model is obtained by magnetic resonance imaging (MRI) images. This method is integrated with neuro navigation tools, resulting in efficient application. Puthe *et al.* [8] proposed a method that uses three dimensional (3D) modeling of induction process in the earth and source model obtained by spherical harmonics of observed magnetic data. Petrovic *et al.* [9] proposed a method for estimation of low frequency magnetic field of the overhead power line as these low frequency magnetic fields create health risks. To estimate the extremely low frequency fields, Petrovic *et al.* used Biot-Savart law based model that considers the pillar geometry and conductor mutual position and its catenary shape. Kazachenko *et al.* [10] proposed the estimation of photospheric electric fields using magnetic field sequence and Doppler measurement. Doncker *et al.* [11] proposed a statistical approach to deduce the local behaviour of the fields when a set of field values are given. Ekonomou *et al.* [12] presented electric and magnetic field radiated by electrostatic discharges using artificial neural network. This method has the advantage that the electromagnetic field can be calculated easily and accurately by measuring the discharge current only. Azpurua and Ramos [13] reviewed the various interpolatory techniques to estimate the average electromagnetic field to represent the continuous dataset over a map of complete plot area. Colak *et al.* [14] developed a visual software, having 3D screening unit using artificial neural network. Grid data users can estimate electromagnetic field using this software.

Various methods have been used for state estimation. KF is based on stochastic approach for estimation of states. Recently, KF has been used for parameter estimation. Jiang *et al.* [15] used KF for attitude estimation in autonomous vehicles. Lin *et al.* [16] proposed online wildfire propagation estimation using KF. In [17], KF is used for state of charge estimation in supercapacitors. Beside these, KF has wide range of applications in wireless communication [18], unmanned aerial vehicles (UAVs) [19], biomedical sensors [20], chemical sensors [21] and wireless sensors [22] etc. The main advantage of KF is that it takes process and measurement noise into consideration directly. In this paper, we propose a KF based electromagnetic field estimation of Hertzian dipole antenna using vector potential wave equation, and scalar potential wave equation.

Major contributions of this paper are: (i) Wave equation with noisy current source for magnetic vector potential and electric scalar potential have been set up. (ii) Spatial distribution of (i) leads to finite dimensional linear state variable model. (iii) Electric and magnetic fields are expressed in terms of potentials and the measurement model is discretized. (iv) KF is applied to estimate the entire potential from sparse discrete measurements. KF is well known algorithm and has been applied to estimate the state for finite dimensional state variable problem. Maxwell's equations are defined on infinite dimensional systems, therefore very little literature exists that use Maxwell's equations to estimate the field. Our contribution

R. Bansal was with the Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, 110042, India, and he is now with Ajay Kumar Garg Engineering College, Ghaziabad, 201009, India. E-mail: rahulbansal1591@gmail.com

S. Majumdar is with the Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, 110042, India, E-mail korsudipta@rediffmail.com

H. Parthasarthy is Electronics and Communication Engineering, Netaji Subhas University of Technology, New Delhi, 110078, India. E-mail: harisignal@yahoo.com

(*Corresponding author: Sudipta Majumdar*)

0018-926X (c) 2020 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 December 04,2020 at 06:51:51 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/TAP.2020.3027054, IEEE Transactions on Antennas and Propagation

is significant in this respect. In this work, we are able to estimate infinite dimensional state, namely electromagnetic field defined on space-time manifold from only discrete spatial measurements and also, we are able to estimate on real-time basis. This work opens up a new avenue for estimating an electromagnetic field in nonlinear media, i.e. field dependent media, mainly whose permittivity and permeability are nonlinear functions of electromagnetic field.

The paper is organized as follows: State space representation of Maxwell's equation is presented in Section II. Brief introduction to Kalman filter algorithm and RLS filter is presented in Section III and Section IV respectively. Section V presents state space modeling of Hertzian dipole antenna. Implementation of KF to Hertzian dipole is presented in Section VI. Simulation results and conclusions are presented in Section VII and Section VIII respectively.

II. State Space Representation of Maxwell's Equation

Following notations have been used throughout the paper:-

(i) Bold letters denote the vector field defined as the function of time and spatial variables.

(ii) Arrows on letters denote vector valued function of time that are obtained by spatial discretization.

(iii) $\vec{\Phi}(t)$ and $\vec{\Psi}(t)$ represent discretized scalar fields. These discretized versions are $N^3 \times 1$ vectors. So, they are denoted by putting an arrow above them.

(iv) Bold lower case letters denote random vectors $(\mathbf{x}_k, \mathbf{z}_k, \mathbf{v}_k, \mathbf{w}_k.)$

(v) Bold italic lower case letters denote deterministic vectors $(\boldsymbol{u}_k)$.

(vi) Bold italic capital letters denote matrices $(\boldsymbol{Q}_k, \boldsymbol{R}_k, \boldsymbol{F}_k, \boldsymbol{B}_k, \boldsymbol{H}_k)$.

(vii) If $F(t,x,y,z)$ is a scalar or vector field then

$$F(t, n_x, n_y, n_z) \equiv F(t, \Delta\, n_x, \Delta\, n_y, \Delta\, n_z) \quad (1)$$

(viii) e_{n_x} is the $N \times 1$ vector having a 1 at its n_x^{th} position and a 0 at all its other positions.

The electromagnetic field is described by two fields namely, the electric field intensity (**E**) and magnetic field intensity (**H**). Both the quantities depend on the position in space and time. A complete solution of Maxwell's equations results in well known wave equations [23]-[24]. In empty space, the scalar potential (Φ), vector potential (**A**), electric field (**E**), magnetic flux density (**B**), all satisfy the wave equations. The Maxwell's equations can also be expressed in terms of differential equations for the scalar and vector potential. Maxwell's equations are the time dependent first order differential equations. Though, these equations can be solved simultaneously, yet generally, they are reduced to two second order differential equations, known as wave equations. Electric and magnetic fields can be easily evaluated using the scalar and vector potential respectively. The scalar potential and the vector potential, satisfy the wave equation with the source terms $\frac{\rho}{\epsilon_0}$ and $\mu_0\mathbf{J}$ respectively, where ρ is the volume charge density, ϵ_0 is permittivity of free space, μ_0 is vacuum permeability and **J** is the current density. The wave equations of **A** and Φ are obtained using the Lorentz gauge condition ($\nabla \cdot \mathbf{A} = -\frac{1}{c^2}\frac{\partial \Phi}{\partial t}$) [25]. The advantage of Lorentz gauge condition is that the vectorial differential equation for **A** is decoupled into a set of three independent scalar differential equations and there is no mixing of the components (r, θ, ϕ), so that each vector component (**A**) or (Φ) depends only on the source component (**J**) or (ρ). The vector potential at position $\mathbf{r} = (x, y, z)$ satisfies the following wave equation with source:

$$\frac{1}{c^2}\frac{\partial^2 \mathbf{A}(t,\mathbf{r})}{\partial t^2} = \nabla^2 \mathbf{A}(t,\mathbf{r}) + \mu_0 \mathbf{J}(t,\mathbf{r}) + \mathbf{w_J}(t,\mathbf{r}) \quad (2)$$

where c is the velocity of light in free space and $\mathbf{w_J}$ is the noise in the current density field.

Similarly, the scalar potential at position **r** satisfies the following partial differential equation

$$\frac{1}{c^2}\frac{\partial^2 \Phi(t,\mathbf{r})}{\partial t^2} = \nabla^2 \Phi(t,\mathbf{r}) + \frac{\rho(t,\mathbf{r})}{\epsilon_0} - \frac{1}{\mu_0\epsilon_0}\int \nabla \cdot \mathbf{w_J}\, dt \quad (3)$$

The equation of continuity implies

$$\nabla \cdot \left(\mathbf{J} + \frac{\mathbf{w_J}}{\mu_0}\right) = -\frac{\partial}{\partial t}\left(\rho + \epsilon_0 w_\rho\right) \quad (4)$$

where $\mathbf{w}_\rho$ is the noise in the charge density field. Thus by equating the terms of the same order of magnitude, we get

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t} \quad (5)$$

$$\nabla \cdot \mathbf{w_J} = -\mu_0\epsilon_0 \frac{\partial w_\rho}{\partial t} \quad (6)$$

In this work, we have simulated the random current source $\mathbf{w_J}$ and have derived $\mathbf{w}_\rho$ for it using the above formula.

Defining the state vector matrix as

$$\begin{aligned}\xi(t,\mathbf{r}) &= \left[\begin{array}{cccc} \mathbf{A}^T(t,\mathbf{r}) & \frac{\partial \mathbf{A}^T(t,\mathbf{r})}{\partial t} & \Phi(t,\mathbf{r}) & \frac{\partial \Phi(t,\mathbf{r})}{\partial t} \end{array}\right]^T \\ &= \left[\begin{array}{cccc} \mathbf{A}^T(t,\mathbf{r}) & \mathbf{V}^T(t,\mathbf{r}) & \Phi(t,\mathbf{r}) & \Psi(t,\mathbf{r}) \end{array}\right]^T \quad (7)\end{aligned}$$

where

$$\frac{\partial \mathbf{A}(t,\mathbf{r})}{\partial t} = \mathbf{V}(t,\mathbf{r}) \quad (8)$$

and

$$\frac{\partial \Phi(t,\mathbf{r})}{\partial t} = \Psi(t,\mathbf{r}). \quad (9)$$

Equation (2) and (3) can be written in the state-space form as

$$\frac{\partial \mathbf{V}(t,\mathbf{r})}{\partial t} = c^2\nabla^2 \mathbf{A}(t,\mathbf{r}) + c^2\mu_0 \mathbf{J}(t,\mathbf{r}) + c^2 \mathbf{w_J}(t,\mathbf{r}) \quad (10)$$

$$\frac{\partial \Psi(t,\mathbf{r})}{\partial t} = c^2\nabla^2 \Phi(t,\mathbf{r}) + c^2\frac{\rho(t,\mathbf{r})}{\epsilon_0} - \frac{c^2}{\mu_0\epsilon_0}\int \nabla \cdot \mathbf{w_J}\, dt \quad (11)$$

On discretizing the spatial variables into N^3 pixels, we represent the above vector and scalar field at the time t in the form of $N^3 \times 1$ vectors for scalar field and $3N^3 \times 1$ vectors for vector fields.

$$\vec{A}(t) = \sum_{n_x,n_y,n_z=1}^{N} \mathbf{A}(t, n_x, n_y, n_z) \otimes e_{n_x} \otimes e_{n_y} \otimes e_{n_z} \in \mathbb{R}^{3N^3 \times 1} \quad (12)$$

0018-926X (c) 2020 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 December 04,2020 at 06:51:51 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/TAP.2020.3027054, IEEE Transactions on Antennas and Propagation

$$\vec{V}(t) = \sum_{n_x,n_y,n_z=1}^{N} \mathbf{V}(t,n_x,n_y,n_z)\otimes e_{n_x}\otimes e_{n_y}\otimes e_{n_z} \in \mathbb{R}^{3N^3\times 1} \tag{13}$$

$$\vec{\Phi}(t) = \sum_{n_x,n_y,n_z=1}^{N} \Phi(t,n_x,n_y,n_z)\, e_{n_x}\otimes e_{n_y}\otimes e_{n_z} \in \mathbb{R}^{N^3\times 1} \tag{14}$$

$$\vec{\Psi}(t) = \sum_{n_x,n_y,n_z=1}^{N} \Psi(t,n_x,n_y,n_z)\, e_{n_x}\otimes e_{n_y}\otimes e_{n_z} \in \mathbb{R}^{N^3\times 1} \tag{15}$$

Here we are using t as discrete time also. $\otimes$ denotes the Kronecker product [26]-[27]. We can represent the operator ∇^2 by a $N^3\times N^3$ matrix D_0, when acting on scalar field and by a $3N^3\times 3N^3$ matrix $D = I_3\otimes D_0$, when acting on vector field. Specifically,

$$\sum_{n_x,n_y,n_z=1}^{N} \nabla^2\mathbf{x}(t,n_x,n_y,n_z)\otimes e_{n_x}\otimes e_{n_y}\otimes e_{n_z} = D\vec{x}(t) \tag{16}$$

where

$$\vec{x}(t) = \sum_{n_x,n_y,n_z=1}^{N} \mathbf{x}(t,n_x,n_y,n_z)\otimes e_{n_x}\otimes e_{n_y}\otimes e_{n_z} \tag{17}$$

For example if $\mathbf{x}(t,n_x,n_y,n_z)$ is a vector field, then

$$\begin{aligned} D\vec{x}(t) = &\sum_{n_x,n_y,n_z=1}^{N}\Big[\frac{\{\mathbf{x}(t,n_x+1,n_y,n_z)+\mathbf{x}(t,n_x-1,n_y,n_z)-2\mathbf{x}(t,n_x,n_y,n_z)\}}{\Delta^2} \\ &+\frac{\{\mathbf{x}(t,n_x,n_y+1,n_z)+\mathbf{x}(t,n_x,n_y-1,n_z)-2\mathbf{x}(t,n_x,n_y,n_z)\}}{\Delta^2} \\ &+\frac{\{\mathbf{x}(t,n_x,n_y,n_z+1)+\mathbf{x}(t,n_x,n_y,n_z-1)-2\mathbf{x}(t,n_x,n_y,n_z)\}}{\Delta^2}\Big] \\ &\otimes e_{n_x}\otimes e_{n_y}\otimes e_{n_z} \end{aligned} \tag{18}$$

We have defined Δ to be the spatial discretization step size along each of the cartesian direction and hence $x = \Delta\, n_x$, $y = \Delta\, n_y$, $z = \Delta\, n_z$ where n_x, n_y, n_z are integers. Clearly, (18) can be expressed as

$$\begin{aligned} D\vec{x}(t) = &\frac{1}{\Delta^2}\sum_{n_x,n_y,n_z=1}^{N}\mathbf{x}(t,n_x,n_y,n_z)\otimes\big[(e_{n_x+1}+e_{n_x-1} \\ &-2e_{n_x})\otimes e_{n_y}\otimes e_{n_z}+e_{n_x}\otimes(e_{n_y+1}+e_{n_y-1} \\ &-2e_{n_y})\otimes e_{n_z}+e_{n_x}\otimes e_{n_y}\otimes(e_{n_z+1}+e_{n_z-1}-2e_{n_z})\big] \end{aligned} \tag{19}$$

Using

$$\begin{aligned} D_1 = \sum_{n_x,n_y,n_z=1}^{N}&\big[((e_{n_x+1}+e_{n_x-1}-2e_{n_x})\otimes e_{n_y}\otimes e_{n_z}) \\ &\times(e_{n_x}\otimes e_{n_y}\otimes e_{n_z})^T\big] \end{aligned} \tag{20}$$

and similarly for D_2, D_3

$$\begin{aligned} D_2 = \sum_{n_x,n_y,n_z=1}^{N}&\big[(e_{n_x}\otimes(e_{n_y+1}+e_{n_y-1}-2e_{n_y})\otimes e_{n_z}) \\ &\times(e_{n_x}\otimes e_{n_y}\otimes e_{n_z})^T\big] \end{aligned} \tag{21}$$

$$\begin{aligned} D_3 = \sum_{n_x,n_y,n_z=1}^{N}&\big[(e_{n_x}\otimes e_{n_y}\otimes(e_{n_z+1}+e_{n_z-1}-2e_{n_z})) \\ &\times(e_{n_x}\otimes e_{n_y}\otimes e_{n_z})^T\big] \end{aligned} \tag{22}$$

(19) becomes

$$\begin{aligned} D\vec{x}(t) =&\frac{1}{\Delta^2}(I_3\otimes D_1+I_3\otimes D_2+I_3\otimes D_3)\mathbf{x}(t) \\ =&\frac{1}{\Delta^2}(I_3\otimes(D_1+D_2+D_3))\mathbf{x}(t) \\ =&I_3\otimes D_0 \in \mathbb{R}^{3N^3\times 3N^3} \end{aligned} \tag{23}$$

where $D_0 = \frac{1}{\Delta^2}I_3\otimes(D_1+D_2+D_3)\in\mathbb{R}^{N^3\times N^3}$

(8)-(11) can be represented after spatial discretization

$$\frac{\partial\vec{\xi}(t)}{\partial t} = \boldsymbol{F}(t)\vec{\xi}(t)+\boldsymbol{B}^{(1)}(t)\mathbf{J}(t)+\boldsymbol{B}^{(2)}(t)\rho(t)+\boldsymbol{B}^{(3)}(t)\mathbf{w}(t) \tag{24}$$

where

$$\vec{\xi}(t) = \begin{bmatrix} \vec{A}^T(t) & \vec{V}^T(t) & \vec{\Phi}(t) & \vec{\Psi}(t) \end{bmatrix}^T \in \mathbb{R}^{N^3\times 1}$$

$$\boldsymbol{F}(t,\mathbf{r}) = \begin{bmatrix} 0 & \boldsymbol{I} & 0 & 0 \\ c^2(I_3\otimes D_0) & 0 & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{I} \\ 0 & 0 & c^2D_0 & 0 \end{bmatrix}$$

$$\boldsymbol{B}^{(1)}(t) = \begin{bmatrix} 0 & c^2\mu_0 & 0 & 0 \end{bmatrix}^T\otimes\boldsymbol{I}$$

$$\boldsymbol{B}^{(2)}(t) = \begin{bmatrix} 0 & 0 & 0 & \frac{c^2}{\epsilon_0} \end{bmatrix}^T\otimes\boldsymbol{I}$$

$$\boldsymbol{B}^{(3)}(t) = \begin{bmatrix} c^2 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & c^2 \end{bmatrix}$$

$$\mathbf{w}(t) = \begin{bmatrix} \mathbf{w_J} & \mathbf{w}_\rho \end{bmatrix}^T$$

Specifically, the discretized electric field is

$$\vec{E}(t) = \sum_{n_x,n_y,n_z=1}^{N}\mathbf{E}(t,n_x,n_y,n_z)\otimes e_{n_x}\otimes e_{n_y}\otimes e_{n_z} \tag{25}$$

where the actual electric field is

$$\mathbf{E}(t,x,y,z) = -\nabla\Phi(t,x,y,z)-\mathbf{V}(t,x,y,z) \tag{26}$$

so

$$\begin{aligned} \vec{E}(t) = -\frac{1}{\Delta}\sum_{n_x,n_y,n_z=1}^{N}&\begin{bmatrix} \Phi(t,n_x+1,n_y,n_z)-\Phi(t,n_x,n_y,n_z) \\ \Phi(t,n_x,n_y+1,n_z)-\Phi(t,n_x,n_y,n_z) \\ \Phi(t,n_x,n_y,n_z+1)-\Phi(t,n_x,n_y,n_z) \end{bmatrix} \\ &\otimes e_{n_x}\otimes e_{n_y}\otimes e_{n_z}-\sum_{n_x,n_y,n_z=1}^{N}\mathbf{V}(t,n_x,n_y,n_z)\otimes e_{n_x}\otimes e_{n_y}\otimes e_{n_z} \end{aligned} \tag{27}$$

which can be expressed as

$$\vec{E}(t) = -\mathbb{G}_1\Phi(t)-\vec{V}(t) \tag{28}$$

where $\mathbb{G}_1 = \frac{1}{\Delta}\sum_{n_x,n_y,n_z=1}^{N}\begin{bmatrix} ((e_{n_x+1}-e_{n_x})\otimes e_{n_y}\otimes e_{n_z})^T \\ (e_{n_x}\otimes(e_{n_y+1}-e_y)\otimes e_{n_z})^T \\ (e_{n_x}\otimes e_{n_y}\otimes(e_{n_z+1}-e_{n_z}))^T \end{bmatrix}\otimes e_{n_x}\otimes e_{n_y}\otimes e_{n_z}\in\mathbb{R}^{3N^3\times N^3}$ matrix that represents the gradient operator.

0018-926X (c) 2020 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 December 04,2020 at 06:51:51 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/TAP.2020.3027054, IEEE Transactions on Antennas and Propagation

Similarly the magnetic field intensity is

$$\mathbf{H}(t,x,y,z) = \frac{1}{\mu_0}(\nabla \times \mathbf{A}(t,x,y,z)) \tag{29}$$

and its discretized version is

$$\vec{H}(t) = \sum_{n_x,n_y,n_z=1}^{N} \mathbf{H}(t,n_x,n_y,n_z) \otimes e_{n_x} \otimes e_{n_y} \otimes e_{n_z} \tag{30}$$

$$= C\,\vec{A}(t) \tag{31}$$

where $\vec{A}(t) = \sum_{n_x,n_y,n_z=1}^{N} \mathbf{A}(t,n_x,n_y,n_z) \otimes e_{n_x} \otimes e_{n_y} \otimes e_{n_z}$. C is a $3N^3 \times 3N^3$ matrix that represents the curl operator. After estimating the electric scalar and magnetic vector potentials, electric and magnetic field intensities have been estimated using the equation (28) and (31). The expression of the curl operator as a matrix is given by

$$\begin{aligned} C =& \frac{1}{\Delta} \sum_{n_x,n_y,n_z=1}^{N} [e_1[e_3^T \otimes (e(n_x,n_y+1,n_z) - e(n_x,n_y,n_z))^T \\ & - e_2^T \otimes (e(n_x,n_y,n_z+1) - e(n_x,n_y,n_z))^T] \\ & + e_2[e_3^T \otimes (e(n_x,n_y,n_z+1) - e(n_x,n_y,n_z))^T - e_3^T \\ & \otimes ((e(n_x+1,n_y,n_z) - e(n_x,n_y,n_z))^T] \\ & + e_3[e_2^T \otimes (e(n_x+1,n_y,n_z) - e(n_x,n_y,n_z))^T - e_1^T \\ & \otimes (e(n_x,n_y+1,n_z) - e(n_x,n_y,n_z))^T]] \otimes e(n_x,n_y,n_z) \end{aligned} \tag{32}$$

where $e(n_x,n_y,n_z) = e(n_x) \otimes e(n_y) \otimes e(n_z)$, and $e_1 = [1,0,0]^T$, $e_2 = [0,1,0]^T$, $e_3 = [0,0,1]^T$.

Let $\boldsymbol{M} \in \mathbb{R}^{p \times 6N^3}$ be the measurement matrix of the electromagnetic field having p pixels ($p << N^3$). $\boldsymbol{M}$ is a sparse matrix of ones and zeros. The measurement model is then

$$\mathbf{z}(t) = \boldsymbol{M} \begin{bmatrix} \vec{E}(t) \\ \vec{H}(t) \end{bmatrix} \tag{33}$$

$$= \boldsymbol{M} \begin{bmatrix} -\mathbb{G}_1 \vec{\Phi}(t) - \vec{V}(t) \\ C\,\vec{A}(t) \end{bmatrix} + \epsilon_v(t) \tag{34}$$

or equivalently,

$$\mathbf{z}(t) = \boldsymbol{M} \begin{bmatrix} 0 & -\boldsymbol{I} & -\mathbb{G}_1 & 0 \\ C & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \vec{A}(t) \\ \vec{V}(t) \\ \vec{\Phi}(t) \\ \vec{\Psi}(t) \end{bmatrix} + \epsilon_v(t) \tag{35}$$

$$= \boldsymbol{H}\,\vec{\xi}(t) + \epsilon_{\mathbf{v}}(t) \tag{36}$$

where $\epsilon_{\mathbf{v}}(t)$ is the measurement noise.

III. KALMAN FILTER ALGORITHM

KF algorithm is a widely used state estimation technique. KF optimizes the state estimation error energy by considering the measurement and process noise [28]-[31]. The main advantage of KF is that it is a real time algorithm and is therefore computationally cheap like the RLS. Furthermore, KF presents better estimates than the RLS because it takes into account the dynamics of the state and parameter and also, the measurement model while, the RLS uses only the measurement model. State and measurement model for the KF are

$$\mathbf{x}_k = \boldsymbol{F}_{k-1}\mathbf{x}_{k-1} + \boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1} + \mathbf{v}_{k-1} \tag{37}$$

$$\mathbf{z}_k = \boldsymbol{H}_k\mathbf{x}_k + \mathbf{w}_k \tag{38}$$

where $\mathbf{x}_k$ is the state vector, $\boldsymbol{F}_k$ is the state matrix. $\boldsymbol{u}_k$ is the known input. $\mathbf{z}_k$ is the measurement vector. $\mathbf{v}_k$ and $\mathbf{w}_k$ are process and measurement noise respectively. These are Gaussian noise with zero mean and covariance matrices $\boldsymbol{Q}_k$ and $\boldsymbol{R}_k$ respectively. The continuous time state model in (24) can be discretized and cast in the form of (37). The measurement matrix $\boldsymbol{H}_k$ in (38) represents the measurement of a discrete set of spatial pixels of the electromagnetic field.

The state and measurement models (37), (38) in discrete time are obtained from the state and measurement models (24), (36) by discretization in time. Specifically, discretizing (24) in time gives the difference equation

$$\begin{aligned} \xi(t) =& (I + \Delta\boldsymbol{F}(t-1))\xi(t-1) + \Delta\,[\boldsymbol{B}^{(1)}(t-1), \boldsymbol{B}^{(2)}(t-1)] \\ & \begin{bmatrix} \mathbf{J}(t-1) \\ \rho(t-1) \end{bmatrix} + \Delta\,\boldsymbol{B}^{(3)}(t-1)\mathbf{w}(t-1) \end{aligned} \tag{39}$$

We can relate this to the state model (37) by means of the correspondences $k \to t$, $\mathbf{x}_k \to \xi(t)$, $\boldsymbol{F}_{k-1} \to \boldsymbol{I} + \Delta\boldsymbol{F}(t-1)$, $\boldsymbol{B}_{k-1} \to \Delta\,[\boldsymbol{B}^{(1)}(t-1), \boldsymbol{B}^{(2)}(t-1)]$, $\boldsymbol{u}_{k-1} \to \begin{bmatrix} \mathbf{J}(t-1) \\ \rho(t-1) \end{bmatrix}$, $\mathbf{v}_{k-1} \to \boldsymbol{B}^{(3)}(t-1)\Delta\,\mathbf{w}(t-1)$ $\mathbf{w}_{k-1} \to \epsilon_{\mathbf{v}}(t)$, $\boldsymbol{H}_k \to \boldsymbol{M} \begin{bmatrix} 0 & -\boldsymbol{I} & -\mathbb{G}_1 & 0 \\ C & 0 & 0 & 0 \end{bmatrix}$

where the arrow means that the discrete time version is replaced by continuous time counterpart. KF algorithm consists of following steps:

1) *Initialization:-* This step initializes the following parameters: $\boldsymbol{P}_{k-1|k-1}$, $\hat{\mathbf{x}}_{k-1|k-1}$, $\boldsymbol{Q}_{k-1}$ and $\boldsymbol{R}_k$.
2) *Time update:-* It consists of following steps:-
   a) *Prediction of state variable:-*

$$\hat{\mathbf{x}}_{k|k-1} = \boldsymbol{F}_{k-1}\hat{\mathbf{x}}_{k-1|k-1} + \boldsymbol{B}_{k-1}\boldsymbol{u}_k. \tag{40}$$

   b) *Error covariance time update:-*

$$\boldsymbol{P}_{k|k-1} = \boldsymbol{F}_{k-1}\boldsymbol{P}_{k-1|k-1}\boldsymbol{F}_{k-1}^T + \boldsymbol{Q}_{k-1} \tag{41}$$

3) *Measurement update:-* It consists of following steps:-
   a) *Kalman gain is calculated as:-*

$$K_k = \boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^T \left[\boldsymbol{H}_k\boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^T + \boldsymbol{R}_k\right]^{-1} \tag{42}$$

   b) *State estimate measurement is updated using:-*

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \boldsymbol{K}_k^K[\mathbf{z}_k - \boldsymbol{H}_k\hat{\mathbf{x}}_{k|k-1}] \tag{43}$$

   c) *Error covariance measurement is updated as:-*

$$\boldsymbol{P}_{k|k} = \left[\boldsymbol{I} - \boldsymbol{K}_k^K\boldsymbol{H}_k\right]\boldsymbol{P}_{k|k-1} \tag{44}$$

where $(k|k-1)$ and $(k|k)$ represent a prior estimate and post estimate respectively. $\boldsymbol{I}$ is the identity matrix.

0018-926X (c) 2020 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 December 04,2020 at 06:51:51 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/TAP.2020.3027054, IEEE Transactions on Antennas and Propagation

IV. RLS FILTER ALGORITHM

RLS filter is widely used for patrameter estimation which gives the optimal solution in the least square sense [32]-[35]. The desired signal $\mathbf{d}_k$ and optimum solution $\boldsymbol{W}_0 = \begin{bmatrix} W_{0_1} & W_{0_2} ... W_{0_{M-1}} \end{bmatrix}^T$ are mathematically related by

$$\mathbf{d}_k = \mathbf{y}_k^T \boldsymbol{W}_0 + \mathbf{w}_k^{(R)} \tag{45}$$

where $\mathbf{y}_k = [y_k ... y_{k-M+1}]^T$ is the observation vector. $\mathbf{w}_k^{(R)}$ is the zero mean white Gaussian noise. The aim of RLS algorithm is to estimate $\boldsymbol{W}_k$ to minimize the error $\sum_{j=1}^{k} \lambda^{k-j} [\mathbf{d}_j - \boldsymbol{W}_k \mathbf{y}_j^T]^2$. Following steps are involved in the RLS algorithm

$$\zeta_k = \mathbf{d}_k - \mathbf{y}_k^T \boldsymbol{W}_k \tag{46}$$

$$K_k^{(R)} = \frac{\boldsymbol{P}_{k-1}^{(R)} \mathbf{y}_k}{1 + \mathbf{y}_k^T \boldsymbol{P}_{k-1}^{(R)} \mathbf{y}_k} \tag{47}$$

$$\boldsymbol{W}_k = \boldsymbol{W}_{k-1} + \boldsymbol{K}_k^{(R)} \zeta_k \tag{48}$$

$$\lambda_k = 1 - \frac{\left(1 - \mathbf{y}_k^T \boldsymbol{K}_k^{(R)}\right) \zeta_k^2}{\chi} \tag{49}$$

$$\boldsymbol{P}_k^{(R)} = \lambda_k^{-1} \boldsymbol{P}_{k-1}^{(R)} - \lambda_k^{-1} \boldsymbol{K}_k^{(R)} \mathbf{y}_k^T \boldsymbol{P}_{k-1}^{(R)} \tag{50}$$

where ζ is the estimation error, $\boldsymbol{K}_k^{(R)}$ is the RLS gain, λ_k is forgetting factor. $\boldsymbol{P}_k^{(R)}$ is the inverse correlation matrix and χ is constant value.

V. STATE SPACE MODELING OF THE HERTZIAN DIPOLE ANTENNA

Estimation of electric and magnetic field in the space around the antenna is important in antenna analysis. Given an antenna structure together with input excitation, the establishment of current distribution on the antenna structure satisfies the Maxwell's equations everywhere at all times. The antenna analysis has two parts: -

1) Determination of current distribution due to excitation.
2) Field evaluation due to this current distribution in the surrounding space of the antenna. In fact, if we can estimate the electromagnetic field in space, then we can use the boundary conditions to determine the excitation current.

Hertzian dipole is the most basic antenna element [23][25]. It is an elemental antenna consisting of an infinitesimally long wire carrying an alternating current ($I(t)$).

Consider a Hertzian dipole antenna of dl length oriented along z-axis and carrying current $I_0 e^{jwt}$ as shown in Fig. 1. This dipole antenna is driven by a time dependent current density

$$\mathbf{J} = -S.\hat{z}\ \epsilon_0 \cos(\omega t)\, \delta(x)\, \delta(y) \tag{51}$$

and volume charge density

$$\rho = -\frac{S}{\mu_0}\ \frac{1}{\omega} \sin(\omega t)\, \delta(x)\, \delta(y)\, \delta(z) \tag{52}$$

where S is the shape function. Consider a point $P(r_1, \theta, \phi)$ in spherical coordinate system. The magnetic vector potential

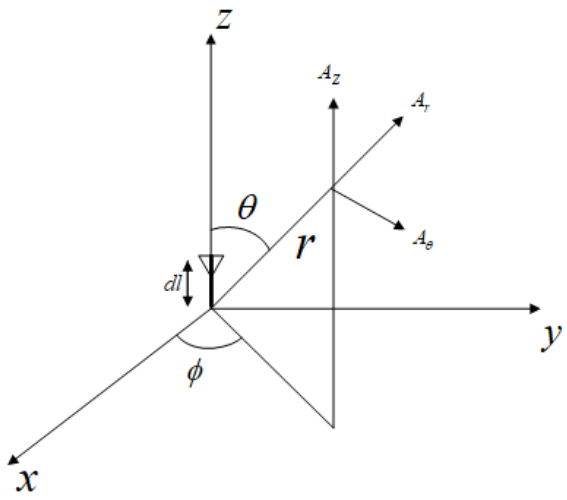


Figure 1. Hertzian dipole antenna.

components at point $P(r_1, \theta, \phi)$ are as given by Green's function

$$A_{r_1} = \frac{\mu_0}{4\pi r_1} I_0\ dl\ e^{j(\omega t - \beta r_1)} \cos(\theta)\ \ \hat{a}_{r_1} \tag{53}$$

$$A_\theta = -\frac{\mu_0}{4\pi r_1} I_0\ dl\ e^{j(\omega t - \beta r_1)} \sin(\theta)\ \ \hat{a}_\theta \tag{54}$$

$$A_\phi = 0 \tag{55}$$

The state space equations are

$$\frac{\partial \vec{A}(t, \mathbf{r})}{\partial t} = \vec{V}(t, \mathbf{r}) \tag{56}$$

$$\begin{aligned} \frac{\partial \vec{V}(t, \mathbf{r})}{\partial t} =& c^2 \left\{ \frac{\mu_0}{4\pi} I_0 dl\ e^{j(\omega t - \beta r_1)} \cos(\theta) \left(\frac{j\beta}{r_1^2} - \frac{\beta^2}{r_1} + \frac{1}{r_1^3} \right) \right\} \hat{a}_r \\ &+ c^2 \left\{ \frac{\mu_0}{4\pi} I_0\ dl\, e^{j(\omega t - \beta r_1)} \sin(\theta) \left(\frac{\beta}{r_1^2} \right) \right\} \hat{a}_\theta \\ &+ c^2 \mu_0 \left\{ -S\ \epsilon_0 \cos(\omega t) \right\} + c^2 \mathbf{w}_{\mathbf{J}}(t) \end{aligned} \tag{57}$$

$$\frac{\partial \Phi(t, \mathbf{r})}{\partial t} = \Psi(t, \mathbf{r}) \tag{58}$$

$$\begin{aligned} \frac{\partial \Psi(t, \mathbf{r})}{\partial t} =& c^2 \left\{ \frac{\omega}{4\pi\epsilon_0} I_0\ dl\ e^{j(\omega t - \beta r_1)} \cos(\theta) \left(-\frac{\beta}{r_1} + \frac{j}{r_1^2} \right) \right\} \\ &+ c^2 \frac{\left\{ -\frac{S}{\mu_0} \cos(\omega t) \right\}}{\epsilon_0} + c^2 \mathbf{w}_\rho(t) \end{aligned} \tag{59}$$

We used Euler-Maruyama method to obtain discrete time state space matrix [36]. Mathematically, $t_k - t_{k-1} = \Delta$ and $\boldsymbol{F}_k = e^{\boldsymbol{F}(t_k - t_{k-1})} \approx \boldsymbol{I} + \boldsymbol{F}\Delta$ and $\boldsymbol{B}_k = \sum_{t_k}^{t_{k-1}} e^{\boldsymbol{F}(t_k - \tau)} \boldsymbol{B} d\tau \approx \boldsymbol{B}\Delta$, where Δ is the sampling time. $\boldsymbol{I}$ is the identity matrix.

VI. APPLYING KF TO HERTZIAN DIPOLE

Differential equation (56)-(59) can be expressed in terms of discrete time state space equation as

$$\mathbf{x}_k = \boldsymbol{F}_{k-1} \mathbf{x}_{k-1} + \boldsymbol{B}_{k-1}^{(1)} \boldsymbol{u}_1 + \boldsymbol{B}_{k-1}^{(2)} \boldsymbol{u}_2 + \boldsymbol{B}_{k-1}^{(3)} \mathbf{w}_{k-1} + \boldsymbol{Z}_k \tag{60}$$

0018-926X (c) 2020 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 December 04,2020 at 06:51:51 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/TAP.2020.3027054, IEEE Transactions on Antennas and Propagation

$$\mathbf{z}_k^{(1)} = \boldsymbol{H}_k^{(1)}\mathbf{x}_k \tag{61}$$

$$\mathbf{z}_k^{(2)} = \boldsymbol{H}_k^{(2)}\mathbf{x}_k \tag{62}$$

where

$$\mathbf{x}_k = \begin{bmatrix} \vec{A} & \vec{V} & \Phi & \Psi \end{bmatrix}^T$$

$$\boldsymbol{F}_{k-1} = \boldsymbol{I} + \Delta \cdot \begin{bmatrix} 0 & \boldsymbol{I} & 0 & 0 \\ c^2(I_3 \otimes D_0) & 0 & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{I} \\ 0 & 0 & c^2 D_0 & 0 \end{bmatrix}$$

$\boldsymbol{B}_{k-1}^{(1)} = \begin{bmatrix} 0 & \Delta c^2\mu_0 & 0 & 0 \end{bmatrix}^T$, $\boldsymbol{u}_1 = -S.\hat{z}\ \epsilon_0 \cos(\omega t)$,
$\boldsymbol{B}_{k-1}^{(2)} = \begin{bmatrix} 0 & 0 & 0 & \frac{\Delta c^2}{\epsilon_0} \end{bmatrix}^T$, $\boldsymbol{u}_2 = -\frac{S}{\mu_0}\cos(\omega t)$,

$$\boldsymbol{B}_{k-1}^{(3)} = \begin{bmatrix} \Delta c^2 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & \Delta c^2 \end{bmatrix},\ \mathbf{w}_{k-1} = \begin{bmatrix} \mathbf{w_J} & \mathbf{w}_\rho \end{bmatrix}^T$$

$$\boldsymbol{Z}_k = \begin{bmatrix} Z_{k_1} & Z_{k_2} & Z_{k_3} & Z_{k_4} \end{bmatrix}^T$$

where

$Z_{k_1} = 0$,
$Z_{k_2} = \Delta c^2 \left\{ \frac{\mu_0}{4\pi} I_0\ dl\ e^{j(\omega t - \beta r_1)} \cos(\theta) \left(\frac{j\beta}{{r_1}^2} - \frac{\beta^2}{r_1} + \frac{1}{r_1^3} \right) \right\} \hat{a}_{r_1}$
$+\Delta c^2 \left\{ \frac{\mu_0}{4\pi} I_0\ dl\ e^{j(\omega t - \beta r_1)} \sin(\theta) \left(\frac{\beta}{r_1^2} \right) \right\} \hat{a}_\theta$,
$Z_{k_3} = 0$,
$Z_{k_4} = \Delta c^2 \left\{ \frac{\omega}{4\pi\epsilon_0} I_0\ dl\ e^{j(\omega t - \beta r_1)} \cos(\theta) \left(-\frac{\beta}{r_1} + \frac{j}{r_1^2} \right) \right\}$,
$\boldsymbol{H}_k^{(1)} = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}$, $\boldsymbol{H}_k^{(2)} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$.

State space representation for Hertzian dipole antenna is

$$\begin{bmatrix} \vec{A}_k \\ \vec{V}_k \\ \vec{\Phi}_k \\ \vec{\Psi}_k \end{bmatrix} = \left[\boldsymbol{I} + \Delta \cdot \begin{bmatrix} 0 & \boldsymbol{I} & 0 & 0 \\ c^2(I_3 \otimes D_0) & 0 & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{I} \\ 0 & 0 & c^2 D_0 & 0 \end{bmatrix} \right] \begin{bmatrix} \vec{A}_{k-1} \\ \vec{V}_{k-1} \\ \vec{\Phi}_{k-1} \\ \vec{\Psi}_{k-1} \end{bmatrix}$$
$$+ \begin{bmatrix} 0 \\ \Delta c^2\mu_0 \\ 0 \\ 0 \end{bmatrix} \boldsymbol{u}_1 + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{\Delta c^2}{\epsilon_0} \end{bmatrix} \boldsymbol{u}_2$$
$$+ \begin{bmatrix} \Delta c^2 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & \Delta c^2 \end{bmatrix} \begin{bmatrix} \mathbf{w_J} \\ \mathbf{w}_\rho \end{bmatrix} + \begin{bmatrix} Z_{k_1} \\ Z_{k_2} \\ Z_{k_3} \\ Z_{k_4} \end{bmatrix} \tag{63}$$

Implementation of KF algorithm on the discrete equations is done by adding process noise $\mathbf{v}_k$ to (60) and measurement noise $\mathbf{w}_k$ to (61) and (62).

$$\mathbf{x}_k = \boldsymbol{F}_{k-1}\mathbf{x}_{k-1} + \boldsymbol{B}_{k-1}^{(1)}\boldsymbol{u}_1 + \boldsymbol{B}_{k-1}^{(2)}\boldsymbol{u}_2 + \boldsymbol{B}_{k-1}^{(3)}\mathbf{w}_{k-1} + \boldsymbol{Z}_k + \mathbf{v}_k \tag{64}$$

$$\mathbf{z}_k^{(1)} = \boldsymbol{H}_k^{(1)}\mathbf{x}_k + \mathbf{w}_k \tag{65}$$

$$\mathbf{z}_k^{(2)} = \boldsymbol{H}_k^{(2)}\mathbf{x}_k + \mathbf{w}_k \tag{66}$$

KF algorithm, as given in Section III, has been applied to equations (64)-(66). Time update and measurement update has been done iteratively after initialization. After estimating the scalar potential and vector potential, electric and magnetic field intensities are computed using (28) and (29) respectively.

Remarks

1) A more direct way to apply the KF on the Hertzian model is to use the partial differential equations

$$\frac{\partial^2 \mathbf{A}(t,x,y,z)}{\partial t^2} = c^2\nabla^2\mathbf{A}(t,x,y,z) + c^2\mu_0\mathbf{J}(t,x,y,z) + c^2\mathbf{w_J}(t,x,y,z) \tag{67}$$

$$\frac{\partial^2 \Phi(t,x,y,z)}{\partial t^2} = c^2\nabla^2\Phi(t,x,y,z) + c^2\frac{\rho(t,x,y,z)}{\epsilon_0} + c^2\mathbf{w}_\rho(t,x,y,z) \tag{68}$$

where

$$\mathbf{J}(t,x,y,z) = I_0.\hat{z}\ cos(\omega t)\,\delta(x)\,\delta(y)\,\Theta\left(\frac{z}{\delta l}\right) \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{69}$$

where

$$\Theta\left(\frac{z}{\delta l}\right) = \begin{cases} 1 & \text{if } \left|\frac{z}{\delta l}\right| \le \frac{1}{2} \\ 0 & \text{if } \left|\frac{z}{\delta l}\right| > \frac{1}{2} \end{cases}$$

and

$$\rho(t,x,y,z) = -\int_0^t \nabla\cdot\mathbf{J}\,dt \tag{70}$$

$$= \frac{I_0}{\omega} sin(\omega t) \left(\delta\left(z - \frac{dl}{2}\right) - \delta\left(z + \frac{dl}{2}\right) \right) \times \delta(x)\,\delta(y) \tag{71}$$

and used smooth approximation for the δ-function.

2) We have taken the non-random component of the current and charge signals of a fixed frequency ω so that these signals are sinusoids in time. Then, $I_0\,sin(\omega\,t+\phi)$ or $I_0\,cos(\omega\,t+\phi)$ etc. have been replaced by $Im\{I_0\,e^{j\phi}\cdot e^{j\,\omega t}\}$ and $Re\{I_0\,e^{j\phi}\cdot e^{j\,\omega t}\}$. This makes it easier to calculate time derivatives. If however, the non-random signal components were not sinusoids, then it would be better to work entirely in the time domain.
3) The formulas used in this work indicate that $O\left(\frac{1}{r}\right)$, $O\left(\frac{1}{r^2}\right)$, $O\left(\frac{1}{r^3}\right)$ terms, are all present in the field expression. So, we have carried out simulations in the intermediate zone, neither far nor near field zone. The formula used in this work are based on the exact retarded potential expressions

$$\mathbf{A}(t,\mathbf{r}) = \frac{\mu_0}{4\pi}\int \frac{\mathbf{J}\left(t - \frac{|\mathbf{r}-\mathbf{r}'|}{c}, \mathbf{r}'\right)}{|\mathbf{r}-\mathbf{r}'|} d^3\mathbf{r}' \tag{72}$$

in the time domain and equivalently in the frequency domain

$$\mathbf{A}(\omega,\mathbf{r}) = \frac{\mu_0}{4\pi}\int \frac{\mathbf{J}(\omega,\mathbf{r}')\,e^{-\left(\frac{j\omega|\mathbf{r}-\mathbf{r}'|}{c}\right)}}{|\mathbf{r}-\mathbf{r}'|} d^3\mathbf{r}' \tag{73}$$

rather than on the far field formula

$$\mathbf{A}(\omega,\mathbf{r}) \approx \frac{\mu_0}{4\pi}\frac{e^{-\left(\frac{j\omega\mathbf{r}}{c}\right)}}{\mathbf{r}}\int \mathbf{J}(\omega,\mathbf{r}')\,e^{-\left(\frac{j\omega\,\hat{\mathbf{r}}\cdot\mathbf{r}'}{c}\right)} d^3\mathbf{r}' \tag{74}$$

4) Comparing the computational complexity between the KF and the RLS approaches to estimation, such a comparison can only be made for parameter estimation, since the RLS is used for parameter estimation and not

0018-926X (c) 2020 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 December 04,2020 at 06:51:51 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/TAP.2020.3027054, IEEE Transactions on Antennas and Propagation

for state estimation. Take for example a linear model

$$\mathbf{z}[n+1] = \boldsymbol{H}[n+1]\,\boldsymbol{\theta}_1 + \mathbf{v}[n+1], n \geq 0 \tag{75}$$

where $\boldsymbol{\theta}_1 \in \mathbb{R}^{M\times 1}$, $\mathbf{z}[n] \in \mathbb{R}^{p\times N}$, $p << M$. The least squares (LS) method for estimating $\boldsymbol{\theta}_1$ based on measurement data upto the n is given by

$$\hat{\boldsymbol{\theta}}_1[n] = \underset{\boldsymbol{\theta}_1}{\operatorname{argmin}} \sum_{m=0}^{n} ||\mathbf{z}[m] - \boldsymbol{H}[m]\boldsymbol{\theta}_1||^2 \tag{76}$$

$$= \left[\sum_{m=0}^{n} \boldsymbol{H}[m]^T\boldsymbol{H}[m]\right]^{-1} \left[\sum_{m=0}^{n} \boldsymbol{H}[m]^T\mathbf{z}[m]\right] \tag{77}$$

The RLS is obtained by setting

$$\boldsymbol{R}[n] = \sum_{m=0}^{n} \boldsymbol{H}[m]^T\boldsymbol{H}[m] \in \mathbb{R}^{M\times M} \tag{78}$$

and using the matrix inversion lemma to get

$$\boldsymbol{R}[n+1]^{-1} = \boldsymbol{R}[n]^{-1} - \boldsymbol{R}[n]^{-1}\boldsymbol{H}[n+1]^T(\boldsymbol{I}_p + \boldsymbol{H}[n+1] \times \boldsymbol{R}[n]^{-1}\boldsymbol{H}[n]^T)^{-1}\boldsymbol{H}[n+1]\boldsymbol{R}[n]^{-1} \tag{79}$$

Per iteration, the major computational cost comes from inverting a $p \times p$ matrix which involves $O(p^3)$ multiplications. On the other hand, if we use the KF to estimate θ_1 based on the dynamical equation

$$\boldsymbol{\theta}_1[n+1] = \boldsymbol{\theta}_1[n] + \varepsilon[n] \tag{80}$$

then we get the iteration $\hat{\boldsymbol{\theta}}_1[n+1|n]$ in terms of $\hat{\boldsymbol{\theta}}_1[n|n]$.

$$\hat{\boldsymbol{\theta}}_1[n+1|n+1] = \hat{\boldsymbol{\theta}}_1[n|n] + \boldsymbol{K}[n+1](\mathbf{z}[n+1] - \boldsymbol{H}[n+1]\hat{\boldsymbol{\theta}}_1[n|n]) \tag{81}$$

where

$$\boldsymbol{K}[n+1] = \boldsymbol{P}_{\boldsymbol{\theta}_1}[n+1|n]\boldsymbol{H}[n+1]^T(\boldsymbol{H}[n+1] \boldsymbol{P}_{\boldsymbol{\theta}_1}[n+1|n]\boldsymbol{H}[n+1]^T + \boldsymbol{R}[n+1])^{-1} \tag{82}$$

and again cost per iteration is $O(p^3)$ for inverting the $p \times p$ matrix. However, KF has an additional equation involving getting $\boldsymbol{P}_{\boldsymbol{\theta}_1}[n+1|n+1]$ from $\boldsymbol{P}_{\boldsymbol{\theta}_1}[n|n]$. In this case, the RLS may be better. However, KF can take noise into account and other evolution in the parameter dynamics which the RLS can not. In this work, to use the RLS, we must assume that the electromagnetic fields are known everywhere as functions of some unknown parameter vector $\boldsymbol{\theta}_1$ and linearize around some given $\boldsymbol{\theta}_0$ to obtain linear model for the field measurement given $\delta\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1 - \boldsymbol{\theta}_0$. The dynamics of the measurement, namely the Maxwell's equations are not involved in the RLS algorithm involving $\delta\boldsymbol{\theta}_1$ estimation.

5) If we wish to use the LS method for the parameter $\boldsymbol{\theta}_1$ in the model

$$\mathbf{x}_{k+1} = \boldsymbol{F}(\boldsymbol{\theta}_1)\mathbf{x}_k + \boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{u}_k + \mathbf{w}_{k+1} \tag{83}$$

$$\mathbf{z}_k = \boldsymbol{H}(\boldsymbol{\theta}_1)\mathbf{x}_k + \mathbf{v}_k \tag{84}$$

then we would first solve

$$\mathbf{x}_k \approx \boldsymbol{F}(\boldsymbol{\theta}_1)^k\mathbf{x}_0 + \textstyle\sum_{m=0}^{k-1} \boldsymbol{F}(\boldsymbol{\theta}_1)^{k-1-m}\boldsymbol{G}(\boldsymbol{\theta}_1)\boldsymbol{u}_m + \text{noise} \tag{85}$$

$$\mathbf{z}_k = \boldsymbol{H}(\boldsymbol{\theta}_1)\boldsymbol{F}(\boldsymbol{\theta}_1)^k\mathbf{x}_0 + \sum_{m=0}^{k-1} \boldsymbol{H}(\boldsymbol{\theta}_1)\boldsymbol{F}(\boldsymbol{\theta}_1)^{k-1-m}\boldsymbol{G}(\boldsymbol{\theta}_1)\boldsymbol{u}_m + \text{noise} \tag{86}$$

and then consider

$$\hat{\boldsymbol{\theta}}_1 = \underset{\boldsymbol{\theta}_1}{\operatorname{argmin}} \sum_{k=1}^{N} ||\mathbf{z}_k - \boldsymbol{H}(\boldsymbol{\theta}_1)\boldsymbol{F}(\boldsymbol{\theta}_1)^k\mathbf{x}_0 - \sum_{m=0}^{k-1} \boldsymbol{H}(\boldsymbol{\theta}_1)\boldsymbol{F}(\boldsymbol{\theta}_1)^{k-1-m}\boldsymbol{G}(\boldsymbol{\theta}_1)\boldsymbol{u}_m||^2 \tag{87}$$

which would be a highly nonlinear problem that is not recursively implementable.

The RLS algorithm can be applied to (37) for estimating the model parameters provided that we have accurate state measurements. For example, it can be applied in our problem to estimate μ_0, ϵ_0. More generally, if

$$\boldsymbol{F}_{k-1} = \sum_{m=0}^{p} \boldsymbol{\theta}_{1_m}\boldsymbol{G}_{k-1\ m} = \boldsymbol{F}_{k-1}(\boldsymbol{\theta}_1) \tag{88}$$

$$\boldsymbol{B}_{k-1} = \sum_{m=0}^{p} \boldsymbol{\theta}_{1_m}\boldsymbol{B}_{k-1\ m} = \boldsymbol{B}_{k-1}(\boldsymbol{\theta}_1) \tag{89}$$

in (37), then $[\boldsymbol{\theta}_{1_m}]_{m=0}^{p}$ may be estimated using the RLS by minimizing

$$\sum_{j=1}^{k} \lambda^{k-j}\, ||\mathbf{x}_k - \boldsymbol{F}_{k-1}(\boldsymbol{\theta}_1)\,\mathbf{x}_{k-1} - \boldsymbol{B}_{k-1}(\boldsymbol{\theta}_1)\,\boldsymbol{u}_{k-1}||^2 = \varepsilon_k(\boldsymbol{\theta}_1) \tag{90}$$

6) Fig. 18 and Fig. 19 show another simulation with (69) replaced by

$$\mathbf{J}(t,x,y,z) = I_0.\hat{z}\ e^{-\gamma\, t}\ cos(\omega t)\,\delta(x)\,\delta(y)\ \Theta\left(\tfrac{z}{\delta l}\right) \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{91}$$

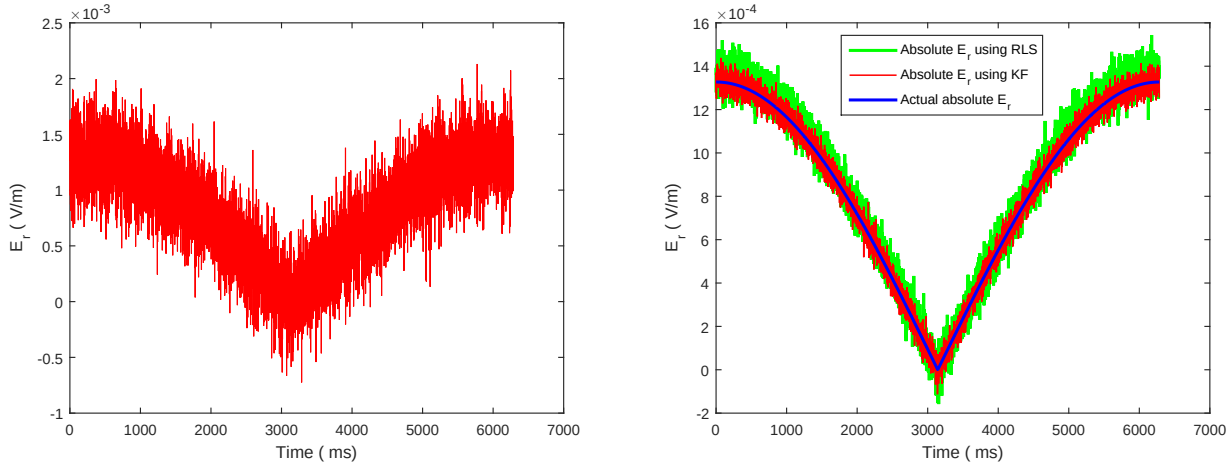
This signal represents a sinusoid modulated by a decaying exponential.

VII. Simulation Results

The derived expressions have been implemented in MATLAB software. Initial values assigned to the KF are:- the covariance of process noise $\boldsymbol{Q}_k$ = diag[0.0001 0.001 0.0001 0.01], the covariance of measurement noise, $\boldsymbol{R}_k$ = 0.01 and the predicted covariance, $\boldsymbol{P}(0)$ = diag[0.0001 0.0001 0.0001 0.0001]. The number of samples N used are 6284. The input frequency used is 10 MHz. Fig. 2(a) shows the noisy electric field output and Fig. 2(b) shows estimated electric field magnitude using KF and RLS. The magnitude and phase estimation of (r, θ) components of the electric field are shown in Fig. 2 - Fig. 5 using KF and RLS for near field. Fig. 6 and Fig. 7

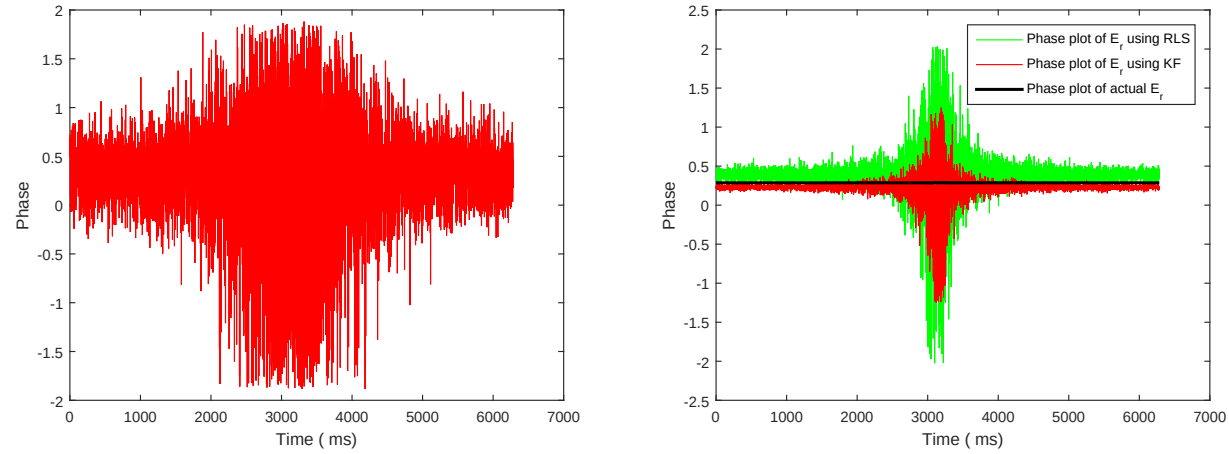
0018-926X (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 06:51:51 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/TAP.2020.3027054, IEEE Transactions on Antennas and Propagation



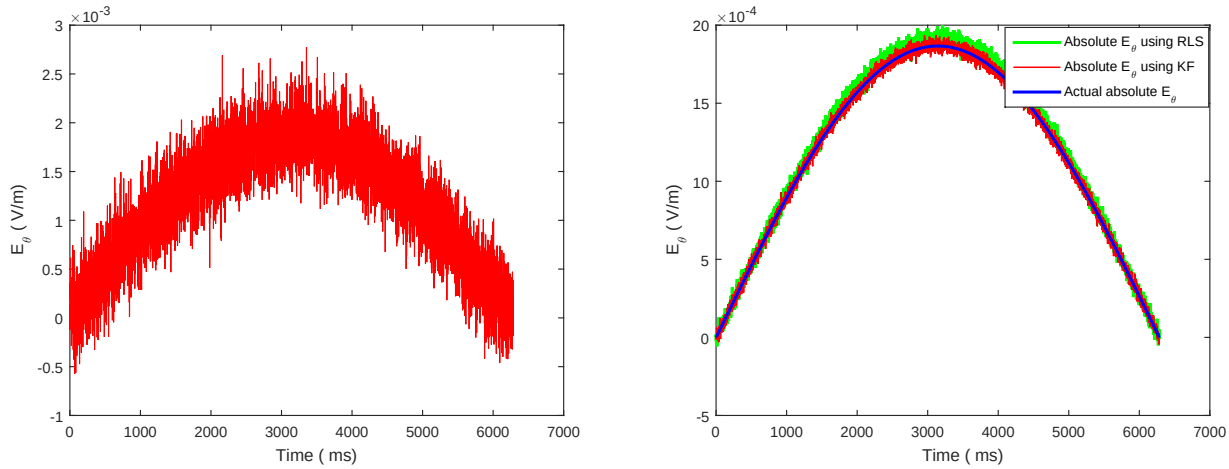
(a) Noisy electric field intensity (E_r). (b) Estimated electric field intensity (E_r).

Figure 2. Electric field intensity (E_r) using noisy driving source (white Gaussian noise of zero mean and 0.00025 variance) for near-field.



(a) Phase plot of noisy electric field intensity (E_r). (b) Phase plot of estimated electric field intensity (E_r).

Figure 3. Phase plot of electric field intensity (E_r) using noisy driving source (white Gaussian noise of zero mean and 0.00025 variance) for near-field.

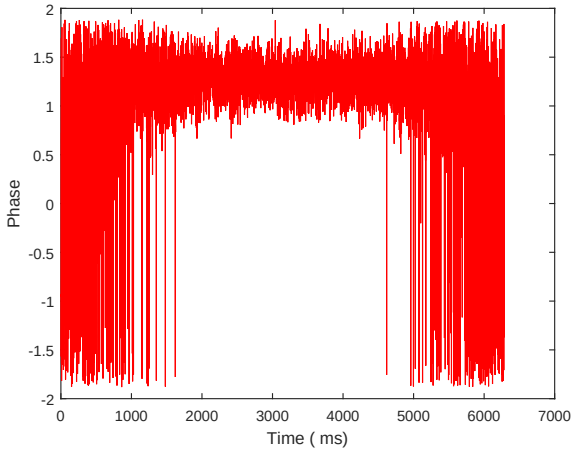


(a) Noisy electric field intensity (E_θ). (b) Estimated electric field intensity (E_θ).

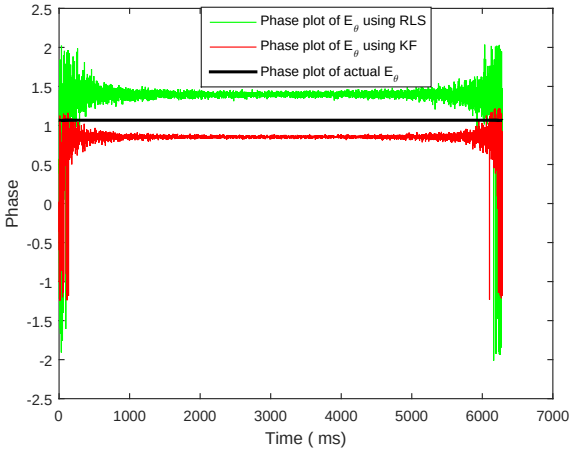
Figure 4. Electric field intensity (E_θ) using noisy driving source (white Gaussian noise of zero mean and 0.00025 variance) for near-field.

0018-926X (c) 2020 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 December 04,2020 at 06:51:51 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/TAP.2020.3027054, IEEE Transactions on Antennas and Propagation

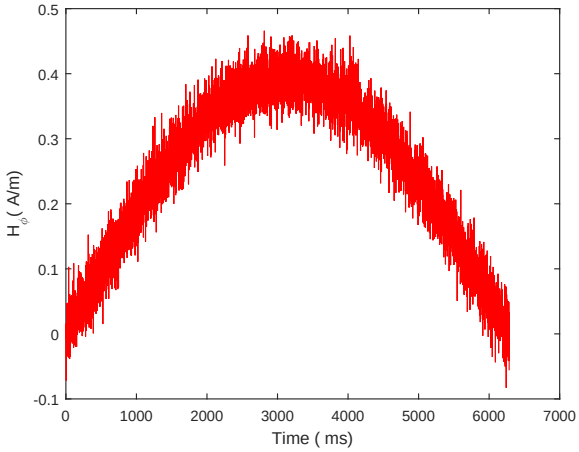


(a) Phase plot of noisy electric field intensity (E_θ).

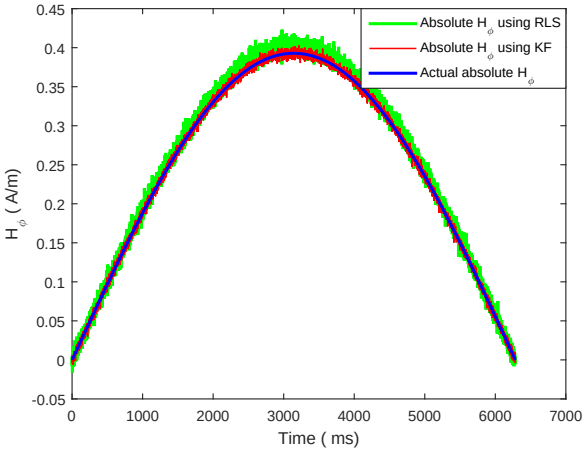


(b) Phase plot of estimated electric field intensity (E_θ).

Figure 5. Phase plot of electric field intensity (E_θ) using noisy driving source (white Gaussian noise of zero mean and 0.00025 variance) for near-field.

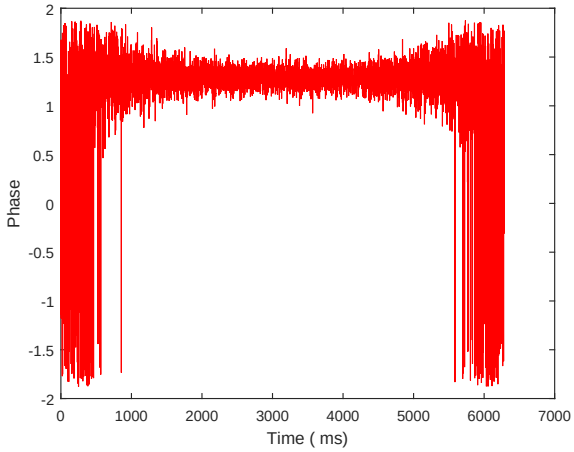


(a) Noisy magnetic field intensity (H_ϕ).

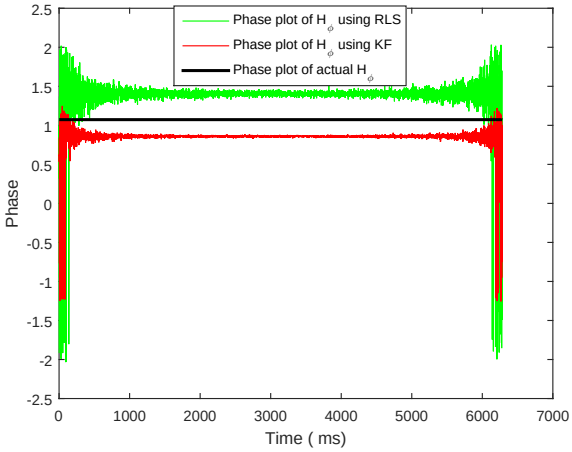


(b) Estimated magnetic field intensity (H_ϕ).

Figure 6. Magnetic field intensity (H_ϕ) using noisy driving source (white Gaussian noise of zero mean and 0.00025 variance) for near-field.



(a) Phase plot of noisy magnetic field intensity (H_ϕ).

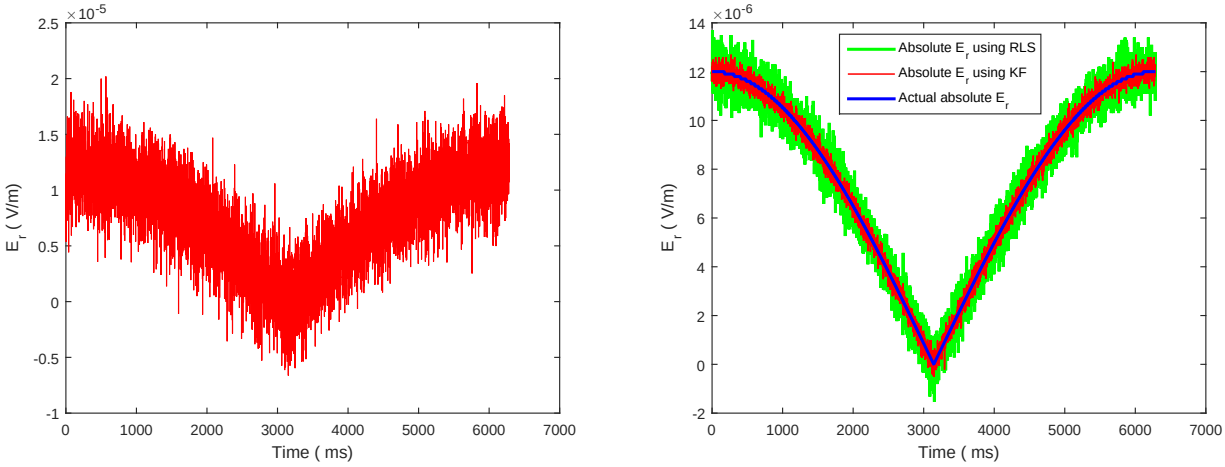


(b) Phase plot of estimated magnetic field intensity (H_ϕ).

Figure 7. Phase plot of magnetic field intensity (H_ϕ) using noisy driving source (white Gaussian noise of zero mean and 0.00025 variance) for near-field.

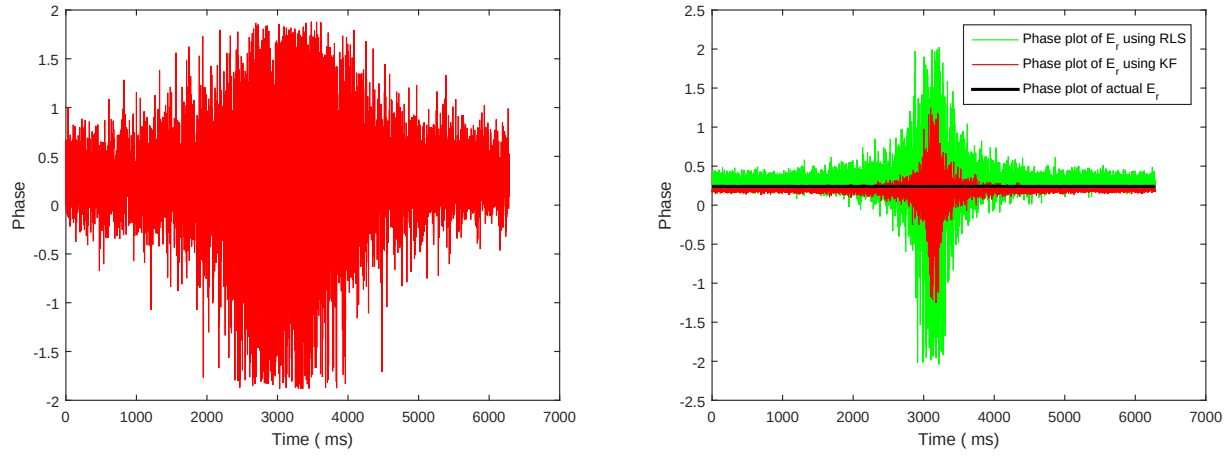
0018-926X (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 06:51:51 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/TAP.2020.3027054, IEEE Transactions on Antennas and Propagation



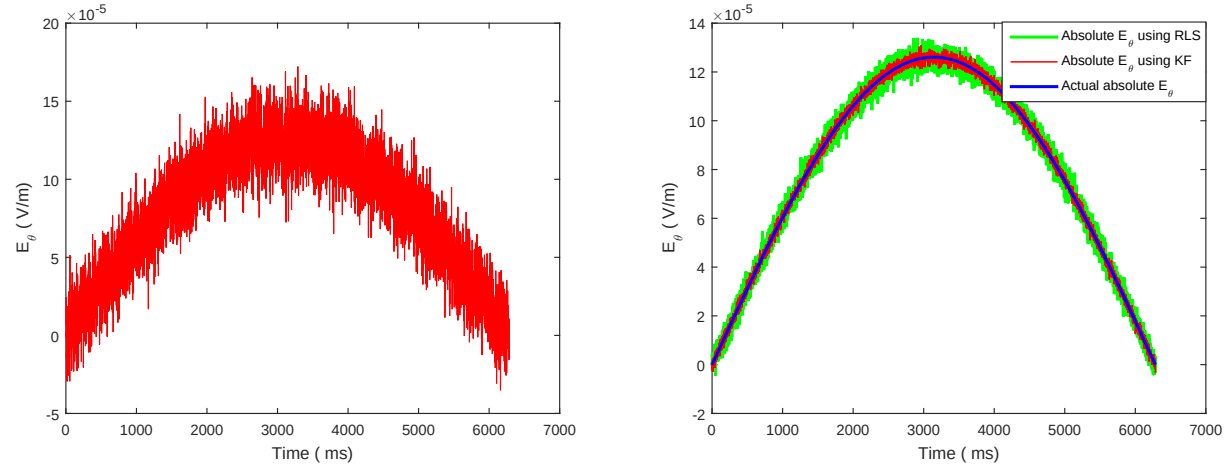
(a) Noisy electric field intensity (E_r). (b) Estimated electric field intensity (E_r).

Figure 8. Electric field intensity (E_r) using noisy driving source (white Gaussian noise of zero mean and 0.00025 variance) for far-field.



(a) Phase plot of noisy electric field intensity (E_r). (b) Phase plot of estimated electric field intensity (E_r).

Figure 9. Phase plot of electric field intensity (E_r) using noisy driving source (white Gaussian noise of zero mean and 0.00025 variance) for far-field.

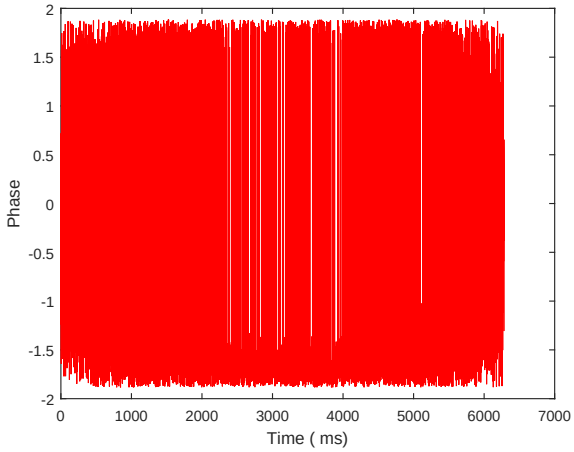


(a) Noisy electric field intensity (E_θ). (b) Estimated electric field intensity (E_θ).

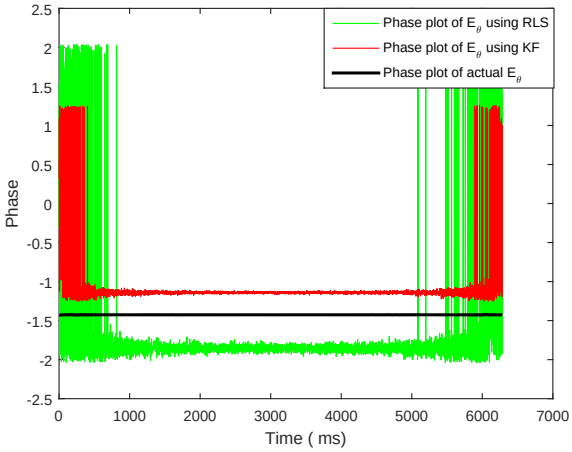
Figure 10. Electric field intensity (E_θ) using noisy driving source (white Gaussian noise of zero mean and 0.00025 variance) for far-field.

0018-926X (c) 2020 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 December 04,2020 at 06:51:51 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/TAP.2020.3027054, IEEE Transactions on Antennas and Propagation

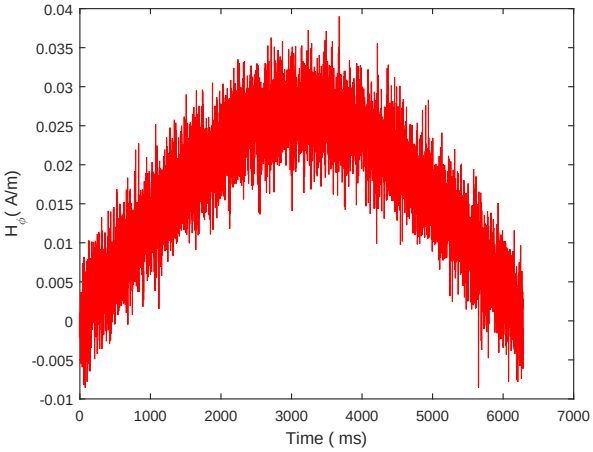


(a) Phase plot of noisy electric field intensity (E_θ).

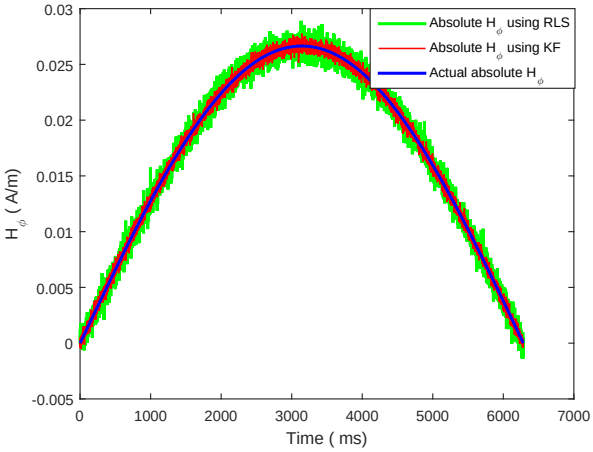


(b) Phase plot of estimated electric field intensity (E_θ).

Figure 11. Phase plot of electric field intensity (E_θ) using noisy driving source (white Gaussian noise of zero mean and 0.00025 variance) for far-field.

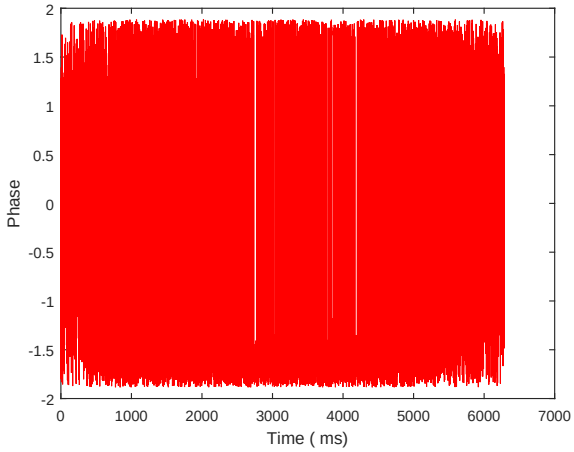


(a) Noisy magnetic field intensity (H_ϕ).

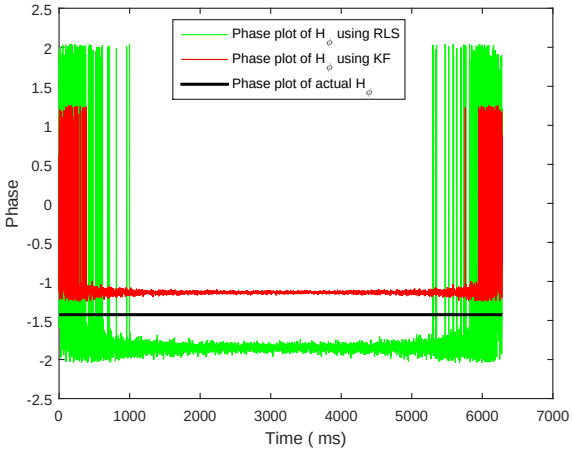


(b) Estimated magnetic field intensity (H_ϕ).

Figure 12. Magnetic field intensity (H_ϕ) using noisy driving source (white Gaussian noise of zero mean and 0.00025 variance) for far-field.



(a) Phase plot of noisy magnetic field intensity (H_ϕ).

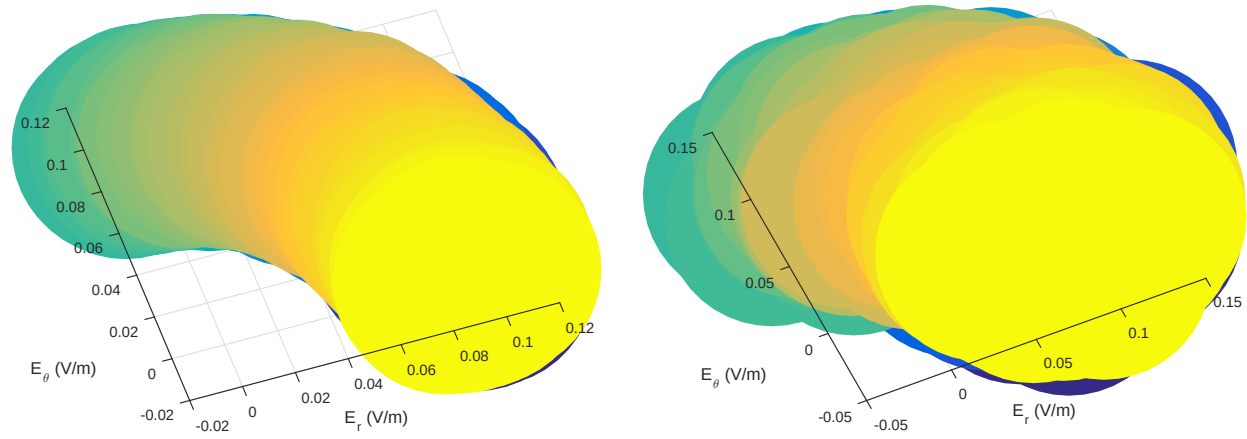


(b) Phase plot of estimated magnetic field intensity (H_ϕ).

Figure 13. Phase plot of magnetic field intensity (H_ϕ) using noisy driving source (white Gaussian noise of zero mean and 0.00025 variance) for far-field.

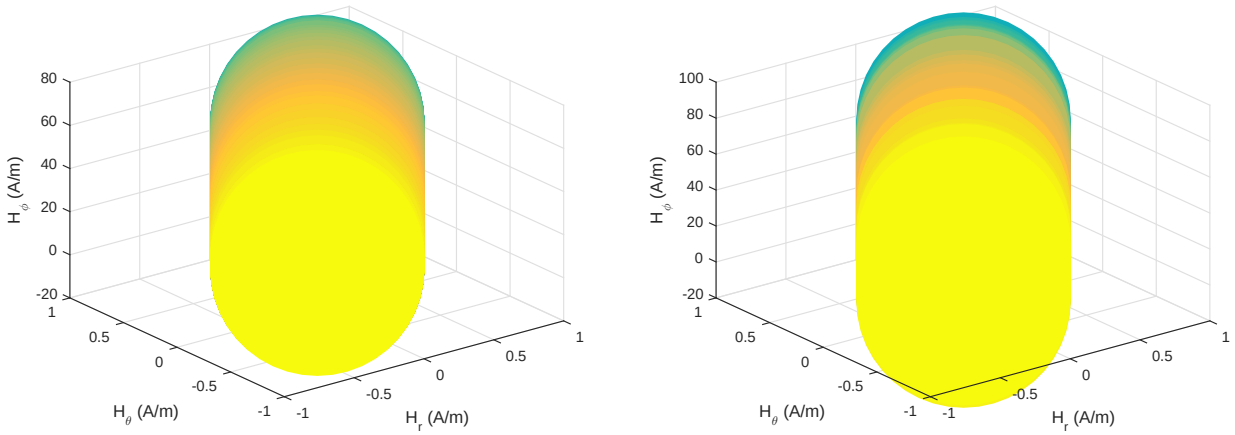
0018-926X (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 06:51:51 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/TAP.2020.3027054, IEEE Transactions on Antennas and Propagation



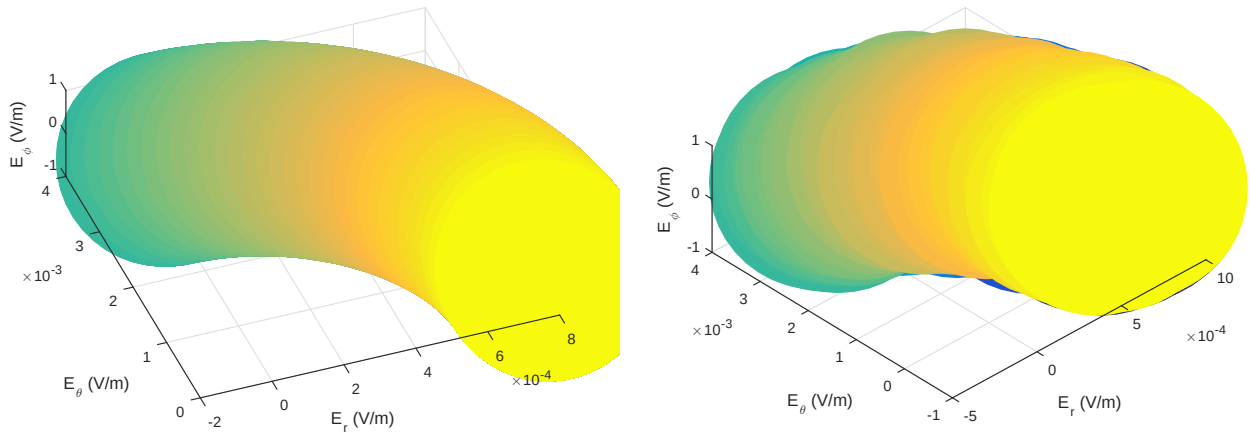
(a) (b)

Figure 14. (a): Estimated E-field using KF for near field. (b): Estimated E-field using RLS for near field.



(a) (b)

Figure 15. (a): Estimated H-field using KF for near field. (b): Estimated H-field using RLS for near field.

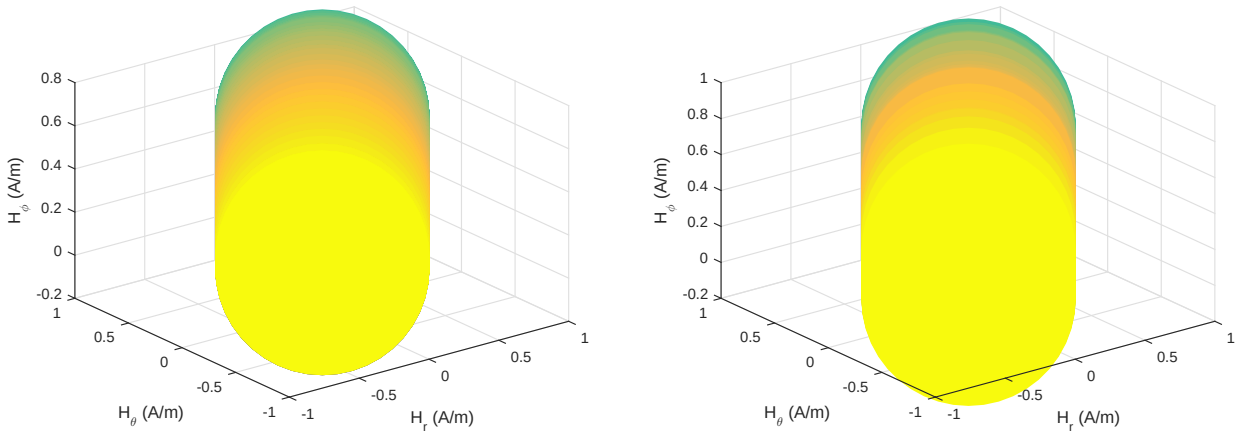


(a) (b)

Figure 16. (a): Estimated E-field using KF for far field. (b): Estimated E-field using RLS for far field.

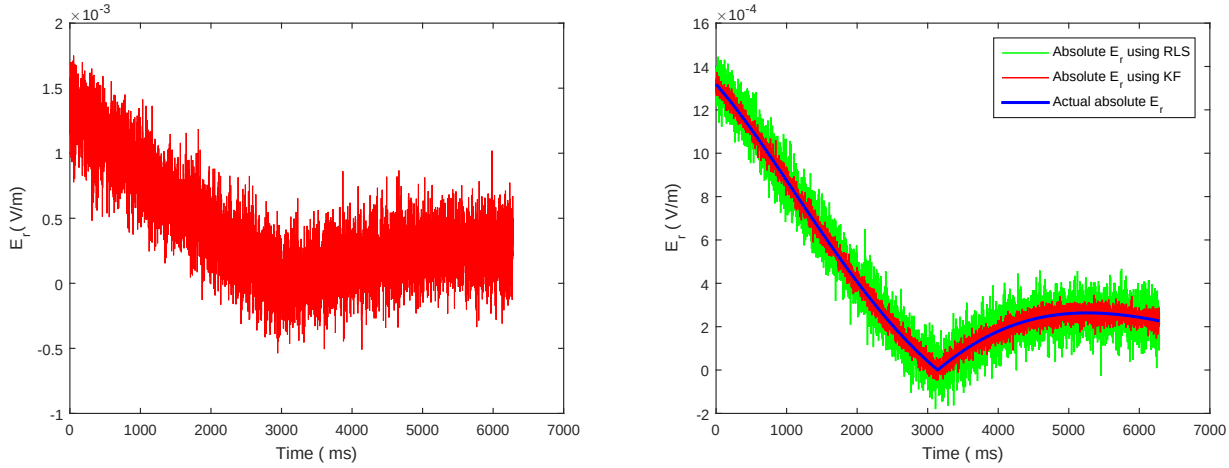
0018-926X (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 06:51:51 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/TAP.2020.3027054, IEEE Transactions on Antennas and Propagation



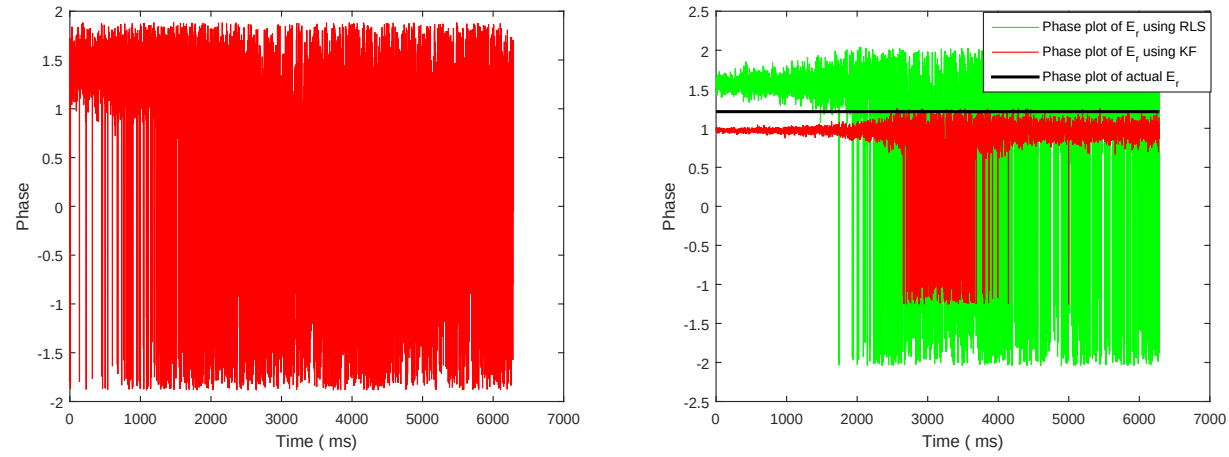
(a) (b)

Figure 17. (a): Estimated H-field using KF for far field. (b): Estimated H-field using RLS for far field.



(a) Noisy electric field intensity (E_r). (b) Estimated electric field intensity (E_r).

Figure 18. Modulated electric field intensity (E_r) using noisy driving source (white Gaussian noise of zero mean and 0.00025 variance) for near-field.



(a) Phase plot of noisy electric field intensity (E_r). (b) Phase plot of estimated electric field intensity (E_r).

Figure 19. Phase plot of modulated electric field intensity (E_r) using noisy driving source (white Gaussian noise of zero mean and 0.00025 variance) for near-field.

0018-926X (c) 2020 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 December 04,2020 at 06:51:51 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/TAP.2020.3027054, IEEE Transactions on Antennas and Propagation

Table I
RMSE OF ELECTRIC FIELD AND MAGNETIC FIELD ESTIMATION USING KF AND RLS FOR NEAR-FIELD.

Parameters	Gaussian noise at driving source	Magnitude estimation using KF	Phase estimation using KF	Magnitude estimation using RLS	Phase estimation using RLS
E_r	$\mu = 0,\ \sigma^2 = 0.00025$	1.92×10^{-5}	0.0030	3.41×10^{-5}	0.0043
	$\mu = 0,\ \sigma^2 = 0.00050$	5.23×10^{-5}	0.0051	5.65×10^{-5}	0.0064
E_θ	$\mu = 0,\ \sigma^2 = 0.00025$	4.71×10^{-7}	0.0009	9.21×10^{-7}	0.0021
	$\mu = 0,\ \sigma^2 = 0.00050$	7.03×10^{-7}	0.0013	1.88×10^{-6}	0.0038
H_ϕ	$\mu = 0,\ \sigma^2 = 0.00025$	0.0102	0.0029	0.0109	0.0045
	$\mu = 0,\ \sigma^2 = 0.00050$	0.0124	0.0032	0.0146	0.0051

Table II
RMSE OF ELECTRIC FIELD AND MAGNETIC FIELD ESTIMATION USING KF AND RLS FOR FAR-FIELD.

Parameters	Gaussian noise at driving source	Magnitude estimation using KF	Phase estimation using KF	Magnitude estimation using RLS	Phase estimation using RLS
E_r	$\mu = 0,\ \sigma^2 = 0.00025$	1.92×10^{-7}	0.0022	7.75×10^{-7}	0.0046
	$\mu = 0,\ \sigma^2 = 0.00050$	4.45×10^{-7}	0.0048	9.61×10^{-7}	0.0061
E_θ	$\mu = 0,\ \sigma^2 = 0.00025$	6.84×10^{-7}	0.0068	9.58×10^{-7}	0.0091
	$\mu = 0,\ \sigma^2 = 0.00050$	8.12×10^{-7}	0.0087	1.54×10^{-6}	0.0098
H_ϕ	$\mu = 0,\ \sigma^2 = 0.00025$	7.06×10^{-4}	0.0079	1.23×10^{-3}	0.0122
	$\mu = 0,\ \sigma^2 = 0.00050$	7.29×10^{-4}	0.0097	1.49×10^{-3}	0.0231

show the magnitude and phase estimation of (ϕ) component of magnetic field for near field. Fig. 8 - Fig. 11 show the magnitude and phase estimation of (r, θ) components of the electric field for far field. The magnitude and phase estimation of (ϕ) component of magnetic field are shown in Fig. 12 - Fig. 13 for far field. Fig. 14 - Fig. 17 show the three dimensional plot of estimated electric and magnetic fields for near and far field using KF and RLS methods. Fig. 18 shows the estimated field for modulated signal. If a modulated noisy signal is used in place of a time harmonic one in the current source, as for example an exponentially decaying sinusoid, then the vector and scalar potentials will also decay with time. Hence, we expect transient behaviour of the KF output whose amplitude decays to zero or if the process and measurement noise processes are strong, the KF output will not decay to zero but to some small values that fluctuate around zero, i.e., the limiting estimate has a small finite variance just as it happens in the least mean square (LMS) algorithm. If the current source signal consists of harmonics plus decaying harmonics, then the limiting KF estimate will show some transient behaviour, but will ultimately converge to a harmonic process with some small noise with only those frequencies appearing in the harmonic process components that do not have any decaying amplitude in the corresponding input current source components. This fact is clear from the Fig. 18 and Fig. 19. Following mathematical expression is used to compute root mean square error (RMSE)

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N} e_k^2} \tag{92}$$

where e_k is the difference between the theoretical value and the estimated value. N is the total number of samples used.

Table 1 - Table 2 show the RMSE for electric and magnetic field estimation using KF and compares the result with RLS and theoretical value. The KF presents better accuracy as compared to the RLS method, as it is an optimal estimator that minimizes the least square of the estimation error.

VIII. CONCLUSIONS

In this paper, we estimated the electric field intensity and magnetic field intensity of the Hertzian dipole antenna using KF algorithm. For this, stochastic based state space equations are obtained using scalar and vector potential wave equations. These state space equations can be used for state and parameter estimation of any antenna. We have estimated the field intensity in the Hertzian dipole antenna using KF and RLS algorithm. Estimation using KF gives better results as this algorithm takes process noise and measurement noise into the consideration. The main advantage in applying the KF for estimating the electromagnetic field in space-time is that the KF is a real time estimation algorithm and moreover, it takes the advantage of apriori dynamics of the field in the entire space-time zone to provide estimation over this entire zone by only taking measurements over a sparse zone. The basic philosophy of the proposed work is that by using sparse measurements of the electromagnetic fields, we are able to retrieve on a real time basis, the entire electromagnetic field, because we make use of the extra information available to us in the form of the noisy Maxwell's equations.

ACKNOWLEDGMENT

First author thanks CSIR India.

0018-926X (c) 2020 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 December 04,2020 at 06:51:51 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/TAP.2020.3027054, IEEE Transactions on Antennas and Propagation

REFERENCES

[1] Y. Sugimoto, H. Arai, T. Maruyama, M. Nasuno, M. Hirose and S. Kurokawa, "Fast far-field estimation method by compact single cut near-field measurements for electrically long antenna array," *IEEE Trans Antennas Propag.*, vol. 66, no. 11, pp. 5859-5868, 2018.

[2] A. Arduino, O. Bottauscio, M. Chiampi, I. Laakso and L Zilberti, "Computational Low-Frequency Electromagnetic Dosimetry Based on Magnetic Field Measurements," *IEEE J. Electromagn*, RF, Microw. Med. Biol., vol. 2, no. 4, pp. 302-309, 2018.

[3] L. Pei, D. Liu, D. Zou, R. L. F. Choy, Y. Chen and Z He, "Optimal heading estimation based multidimensional particle filter for pedestrian indoor positioning," *IEEE Access* , vol. 6, pp. 49705-49720, 2018.

[4] C. A. Belhadj and S. El-Ferik, "Electric and magnetic fields estimation for live transmission line right of way workers using artificial neural network. In 2009 15th International Conference on Intelligent System Applications to Power Systems 2009, pp. 1-6.

[5] W. Wang, O. Bottauscio, M. Chiampi, D. Giordano and L. Zilberti, "A procedure to estimate the electric field induced in human body exposed to unknown magnetic sources," *Radiation protection dosimetry*, vol. 154, no. 2, pp. 157-163, 2012.

[6] M. Zenczak. Method of estimating the exposure of the natural environment to 50 Hz electric and magnetic fields in power systems with distributed and centralized generations," Archives of electrical engineering, vol. 65, no. 2, pp. 295-304, 2016.

[7] A. Paffi, F. Camera, F. Carducci, G. Rubino, P. Tampieri, M. Liberti and F. Apollonio, "A computational model for real-time calculation of electric field due to transcranial magnetic stimulation in clinics," *Int. J. Antenn. Propag.*, 976854, pp. 1-11, 2015.

[8] C. Puthe, C. Manoj and A. Kuvshinov, "Reproducing electric field observations during magnetic storms by means of rigorous 3-D modelling and distortion matrix co-estimation," *Earth, Planets and Space*, vol. 66, no. 1, 162, 2014.

[9] G. Petrovic, T. Kilic and T. Garma, "Measurements and estimation of the extremely low frequency magnetic field of the overhead power lines," *Elektronika ir Elektrotechnika*, vol. 19, no. 7, pp. 33-37, 2013.

[10] M. D. Kazachenko, G. H. Fisher and B. T. Welsch, "A comprehensive method of estimating electric fields from vector magnetic field and Doppler measurements," *Astrophys. J.*, vol. 795, no. 1, p.17, 2014.

[11] P. H. De Doncker, J. M. Dricot, R. Meys, M. Helier and W. Tabbara, "Electromagnetic fields estimation using spatial statistics," *Electromagnetics*, vol. 26, no. 2, pp. 111-122, 2006.

[12] L. Ekonomou, G. P. Fotis, T. I. Maris and P. Liatsis, "Estimation of the electromagnetic field radiating by electrostatic discharges using artificial neural networks," *Simul. Model. Pract. Th.*, vol. 15, no. 9, pp. 1089-1102, 2007.

[13] M. Azpurua and K. D. Ramos, "A comparison of spatial interpolation methods for estimation of average electromagnetic field magnitude," *Prog. Electromagn. Res. M.*, 14: pp. 135-145, 2010.

[14] I. Colak, I. Kosalay and A. Inan, "Estimation of magnetic field distributions in substation centers using artificial neural networks," *J. Sci. Ind. Res*., vol. 66, no. 10, pp. 819-827, 2007.

[15] R. Jiang, X. Liu, H. Wang and S. S. Ge, "Secure estimation for attitude and heading reference systems under sparse attacks," *IEEE Sensors J.*, vol. 19, no. 2, pp. 641-649, 2019.

[16] Z. Lin, H. H. Liu and M. Wotton, "Kalman filter-based large-scale wildfire monitoring with a system of UAVs," *IEEE Trans. Ind. Electron.*, vol. 66, no. 1, pp. 606-615, 2019.

[17] W. Houlian and Z. Gongbo, "State of charge prediction of supercapacitors via combination of Kalman filtering and backpropagation neural network," *IET Electr. Power. App.*, vol. 12, no. 4, pp. 588-594, 2018.

[18] Z. Wu, J. Li, J. Zuo and S. Li, "Path planning of UAVs based on collision probability and Kalman filter," *IEEE Access*, 6, pp. 34237-34245, 2018.

[19] J. Wu, G. Shi, B. D. Anderson and K. H. Johansson, "Kalman filtering over fading channels: zero-one laws and almost sure stabilities," *IEEE Trans. Inf. Theory*, vol. 64, no. 10, pp. 6731-6742, 2018.

[20] O. Salem, A. Serhrouchni, A. Mehaoua and R. Boutaba, "Event detection in wireless body area networks using Kalman filter and power divergence," *IEEE Trans. Netw. Service Manag.*, vol. 15, no. 3, pp. 1018-1034, 2018.

[21] M. Weiss, M. S. Wiederoder, R. C. Paffenroth, E. C. Nallon, C. J. Bright, V. P. Schnee, S. McGraw, M. Polcha and J. R. Uzarski, "Applications of the Kalman filter to chemical sensors for downstream machine learning," *IEEE Sensors J.*, vol. 18, no. 13, pp. 5455-5463, 2017.

[22] Y. Huang, W. Yu, M. Usman, S. Wang and A. Garcia-Ortiz, "Analysis of optimal reconstruction methods based on incomplete information from sensor nodes using Kalman filter," *IEEE Sensors J.*, vol 18, no. 16, pp. 6889-6902, 2018.

[23] E. C. Jordan and K. G. Balmain, Electro-Magrzetic Waves and Radiating Systems. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

[24] M. Idemen, "The Maxwell's equations in the sense of distributions," *IEEE Trans. Antennas Propag.*, vol. 21, no. 5, pp. 736-738, 1973.

[25] M. N. O. Sadiku, Elements of Electromagnetics. Oxford University Press, 2014.

[26] W. H. Greub, 1967. Multilinear Algebra. Springer, Berlin, Heidelberg, 1967.

[27] J. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. ciruits syst.*, vol. 25, no. 9, pp.772-781, 1978.

[28] S. Haykin, Kalman filtering and neural networks. John Wiley & Sons, 2004.

[29] J. Irwin and J. Hung, "Kalman estimator and sample mean," *IEEE Trans. Autom. Control*, vol. 12, no. 4, pp. 472-472, 1967.

[30] C. Wells, An approximation of the Kalman filter equations. *IEEE Trans. Autom. Control*, vol. 13, no. 4, pp. 445-445, 1968.

[31] M. Mintz, A note on minimax estimation and Kalman filtering. *IEEE Trans. Autom. Control*, vol. 14, no. 5, pp. 588-590, 1969.

[32] S. Haykin, Adaptive filter theory. Pearson Education India, 2005.

[33] P. Swerling, "Modern state estimation methods from the view point of the method of least squares," *IEEE Trans. Autom. Control*, vol. 16, no. 6, pp. 707-119, 1971.

[34] M. Gevers and T. Kailath, "An innovations approach to least-squares estimation–Part VI: Discrete-time innovations representations and recursive estimation," *IEEE Trans. Autom. Control*, vol. 18, no. 6, pp. 588-600, 1973.

[35] T. Kailath, "Some new algorithms for recursive estimation in constant linear systems," *IEEE Trans. Inf. Theory*, vol. 19, no. 6, pp. 750-760, 1973.

[36] P. E. Kloeden and E. Platen, Numerical solution of stochastic differential equations. Springer-Verlag, 1992

0018-926X (c) 2020 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 December 04,2020 at 06:51:51 UTC from IEEE Xplore. Restrictions apply.

2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)
Galgotias University, Greater Noida, UP, India. Oct 2-4, 2020

Study of Estimating Dynamic State Jacobian Matrix and Dynamic System State Matrix Based on PMU

Annu Ahlawat
Department of Electrical engineering
Delhi Technological University
Delhi, India
annuahlawat014@gmail.com

S.T Nagarajan
Department of Electrical Engineering
Delhi Technological University
Delhi, India
stnagarajan@dce.ac.in

Abstract— **Our power system is dynamic in nature, changes occur very rapidly. Stability condition of power system need to be assessed continuously to prevent collapse of power grid. Stability of power system can be estimated with dynamic state Jacobian Matrix and dynamic system state matrix with help of Synchrophasor measurements. In this paper, a Synchrophasor measurement-based method has been studied to calculate the dynamic state Jacobian matrix and Dynamic system state matrix in varying load conditions. A case study on WSCC 9-bus, 3-machine system has been performed in Digsilent Power factory Simulator Software for simulation of ambient conditions. The critical Eigen values are estimated for the System state matrix to show a good measure of proximity of stability.**

Keywords— Phasor measurement unit, dynamic Jacobian matrix, Dynamic system state matrix, Eigen values, Power system stability.

I. INTRODUCTION

The electric power is dynamic in nature, due to rapid variations in the system conditions. Therefore, to secure the electric system it is highly necessary to continuous monitoring the system stability conditions to maintain the system efficiency and reliability. To observe the power system dynamic stability, dynamic state Jacobian matrix and system state matrix plays very important role [1].

For estimation of these matrices it is assumed that system model is well known and system provides coherent set of measurements. As power system is a dynamic system, there are chances of erroneous calculation in the state estimation results, due to corrupted information of the system data, leading to erroneous calculation of Jacobian matrix and system state matrix. The deployment of Wide Area Measurement System (WAMS) advanced technologies gives opportunity to avoid these problems [2].

The Phasor Measurement Units (PMU) provides high speed, real time measurement-based approach for reliable and accurate monitoring, analysis and control of electric power system. The advantages of these synchronized measurement based methods is high speed and completely model free. Therefore, there is no need of system model information such as network topology and parameters for calculation of Dynamic state Jacobian matrix and system state matrix. These matrices are calculated by using real time synchronized information provided by the Phasor Measurement Unit (PMU) based measurements.

Phasor Measurement Unit (PMU) provides measurement of phasors of voltage and current, system frequency and ROCOF (rate of change of frequency) with real time tag [3]. Therefore, based on PMU measurements there are many techniques to calculate the generator rotor angle and generator rotor speed for monitoring transient stability and real time controlling of the power system [4].

The method studied in this paper for estimation of matrices utilizes the Lyapunov equation initially proposed by [5]. In this study, WSCC (Western System Coordinating Council) 3 machine, 9 bus system has been simulated in Digsilent Power factory software and the system is introduced with Stochastic Load Variations [6][7][8]. We have assumed that all the generating buses are having connected PMUs which provides the real time synchronized measurements of rotor angle and rotor speed. Further in this method, we assume that the inertias M of all the generators is well known. For calculation of dynamic system state matrix A we also require the knowledge of Damping coefficient D.

In this paper, a study has been carried out to show the proposed method in [5] to estimated state Jacobian matrix in comparison with the Jacobian matrix calculated using network model. In this work, we also estimate and plot the Eigen values to measure the proximity of instability [9]. The results show that it is possible to monitor the stability margin of the power system with the real time PMU measurements.

II. BAKGOURND OF THE METHOD

In system state estimation the general power system dynamic model is represented as,

$$\dot{x}=a(x,y) \tag{1}$$

$$0=b(x,y) \tag{2}$$

where, a and b are continuous functions. The generator dynamics is represented by equation (1), and electrical transmission system and the internal static behaviour of passive devices is represented by equation (2) in which x is state variable describes generator rotor angle and rotor speed and y is algebraic variable describes bus voltage and bus voltage angles [10].

In this paper, small signal stability is considered in which power system shows ambient oscillations around steady state. Therefore, under this condition classical generator model is considered which represents dynamics of all connected generators in coherent conditions. [11]

On considering Swing equation, equation (1)-(2) written as, [12][13]:

$$\dot{\delta} = \omega \tag{3}$$

$$M\,\dot{\omega} = P_m - P_e - D\omega \tag{4}$$

Where,

$$P_e = \sum_{j=1}^{n} E_i E_j (G_{ij}\mathrm{Cos}(\delta) + B_{ij}\mathrm{Sin}(\delta_{ij})) \tag{5}$$

$\boldsymbol{\delta} = [\delta_1,...\delta_n]^T$ generator rotor angles,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 06:27:18 UTC from IEEE Xplore. Restrictions apply.

$\boldsymbol{\omega} = [\omega_1,...\omega_n]^T$ generator rotor speeds,

$\boldsymbol{P}_m=[P_{m1},...P_{mn}]^T$ generators input mechanical power,

$\boldsymbol{Pe}= [P_{e1},...P_{en}]^T$ generators output electrical power,

M=diag(M_1,...M_n) generator inertia constants,

D = diag(D_1,...D_n) generator damping factors,

E_i =[E_1 ,…,E_n]^T emf magnitude behind the transient reactance;

$Y_{ij}\angle\varphi_{ij} = G_{ij} + jB_{ij}$ is the reduced admittance matrix of the system which also includes generators impedances and rotor saliency is assumed to be neglected. The estimation of the reduced admittance matrix is provided in the Annexure A [14].

In this paper, we assume that the designed power system is experiencing the stochastic load variations [15]. To reveal the random load variations the power system model is now represented as,

$$\dot{\delta} = \omega \tag{6}$$

$$M\,\dot{\omega} = P_m - P_e - D\omega - E^2 G \sum \mathcal{E} \tag{7}$$

Where,

G = diag ([G_{11},...,G_{nn}]) equivalent conductance ;

Σ = diag([σ_1,...,σ_n]) variance of load variation ;

$\mathcal{E}$ = [$\dot{\omega}_1$,..., $\dot{\omega}_n$]^T standard Gaussian random variable.

As we are considering ambient oscillations due to random load variations, the aforementioned system is linearized around steady state and equations are as followings:

$$\begin{bmatrix}\dot{\delta}\\ \dot{\omega}\end{bmatrix} = \begin{bmatrix}\frac{d\dot{\delta}}{d\delta} & \frac{d\dot{\delta}}{d\omega}\\ \frac{d\dot{\omega}}{d\delta} & \frac{d\dot{\omega}}{d\omega}\end{bmatrix}\begin{bmatrix}\delta\\ \omega\end{bmatrix} + \begin{bmatrix}\frac{d\dot{\delta}}{d\mathcal{E}}\\ \frac{d\dot{\omega}}{d\mathcal{E}}\end{bmatrix}\mathcal{E} \tag{8}$$

$$\begin{bmatrix}\dot{\delta}\\ \dot{\omega}\end{bmatrix} = \begin{bmatrix}0 & I_n\\ -M^{-1}\frac{dP_e}{d\delta} & -M^{-1}D\end{bmatrix}\begin{bmatrix}\delta\\ \omega\end{bmatrix} + \begin{bmatrix}0\\ -M^{-1}E^2G\mathcal{E}\end{bmatrix}\mathcal{E} \tag{9}$$

Let, x= $\begin{bmatrix}\delta\\ \omega\end{bmatrix}$, A= $\begin{bmatrix}0 & I_n\\ -M^{-1}\frac{dP_e}{d\delta} & -M^{-1}D\end{bmatrix}$ and,

B= $\begin{bmatrix}0\\ -M^{-1}E^2G\mathcal{E}\end{bmatrix}$

Where,

I_n : Identity matrix, with n being the size of the system,

The equation (9) describes the set of stochastic differential equations which follows Vector Ornstein Uhlenbeck process [16] as followings:

$$\dot{x} = Ax + B\mathcal{E} \tag{10}$$

If the obtained state matrix A of the system is stable, then the stationary covariance matrix, C_{xx} can be used to satisfy the following Lyapunov equation of the model [17]:

$$AC_{xx} + C_{xx}A^T = -BB^T \tag{11}$$

Where,

$$C_{xx} = \begin{bmatrix}C_{\delta\delta} & C_{\delta\omega}\\ C_{\omega\delta} & C_{\omega\omega}\end{bmatrix}$$

On solving equation (9) and (11), the results are obtained as followings: (for complete derivation refer appendix B)

$$\left(\frac{\partial P_e}{\partial\delta}\right) = M\,C_{\omega\omega}\,C_{\delta\delta}^{-1} \tag{12}$$

$$D = \frac{1}{2}M^{-1}\Sigma^2 / C_{\omega\omega} \tag{13}$$

From the equation (12), it can be concluded that Jacobian Matrix can be calculated by using covariance matrices $C_{\delta\delta}$ and Cωω and inertias, M of all the connected generators. As M is generally known and covariance matrices are calculated from the data extracted based on PMU measurements. It is also noticed that the estimation of Jacobian matrix using equation (12) does not require the network topology and parameters. The standard deviation of load variations $\sum$ is not included in the equation (12) indicates that the estimation of Jacobian matrix using proposed technique does not affected by the noise disturbances.

For estimation of system state matrix, A the additional Damping factor, D of the generators must be known. In some cases of classical model D is not known, which can be calculated using (13). However, to estimate D some of the network parameters need to be known such as generator emf, conductance and variance of load variation. In this paper, we have assumed that damping factor is known so that it is easy to calculate the state matrix, A.

A. Determination of Sample Covariance Matrices

The covariance matrix of generator rotor speed is defined as:

$$C_{\omega\omega} = \begin{bmatrix}C_{\omega_1\omega_1} & C_{\omega_1\omega_2} & \cdots & C_{\omega_1\omega_n}\\ C_{\omega_2\omega_1} & C_{\omega_2\omega_2} & \cdots & C_{\omega_2\omega_n}\\ \vdots & \vdots & \ddots & \vdots\\ C_{\omega_n\omega_1} & C_{\omega_n\omega_2} & \cdots & C_{\omega_n\omega_n}\end{bmatrix} \tag{14}$$

The element entries is defined as followings:

$$C_{\omega_i\omega_j} = E\left[(\omega_i - \bar{\omega}_i)\,(\omega_j - \bar{\omega}_j)\right] \tag{15}$$

Where, $\bar{\omega}_i$ is the mean of ω_i and E is the emg's of generator.

In this paper, the window size is selected around 200sec. As limited PMUs are connected so it is difficult to estimate the covariance matrix. Therefore, covariance matrix is replaced by sample covariance matrix whose entries are independent of network parameters. The entries of sample covariance matrix is defined as,

$$Q_{\omega_i\omega_j} = \frac{1}{N-1}\sum_{k-1}^{N}(\tilde{\omega}_{ki} - \overline{\tilde{\omega}_i})(\tilde{\omega}_{kj} - \overline{\tilde{\omega}_j}) \tag{16}$$

Where, N is the sample size. In this paper, we assumed that the PMU provides samples at the rate of 50 samples per sec.

Similarly, the sample covariance matrix of generator rotor angle is defined as,

$$Q_{\delta_i\delta_j} = \frac{1}{N-1}\sum_{k-1}^{N}(\tilde{\delta}_{ki} - \overline{\tilde{\delta}_i})(\tilde{\delta}_{kj} - \overline{\tilde{\delta}_j}) \tag{17}$$

As inertia, M of all generators is known, therefore the state Jacobian matrix is defined by using equation (12) as follows:

$$\left(\frac{\partial P_e}{\partial\delta}\right) = M\,Q_{\omega\omega}Q_{\delta\delta}^{-1} \tag{18}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 06:27:18 UTC from IEEE Xplore. Restrictions apply.

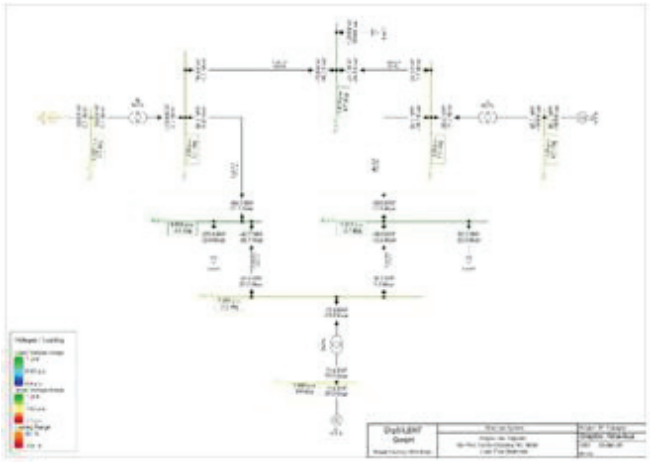


Fig. 1. WSCC 9 bus, 3 machine system simulation in Digsilent Power factory software

III. CASE STUDY: WSCC 3 MACHINE, 9 BUS SYSTEM

The WSCC (Western System Coordinating Council) 3 machine, 9 bus system has been simulated in Digsilent power factory software as shown in fig.1. and the designed system is introduced with Stochastic Load Variations.

IV. THE SYSTEM MODEL METHOD

The stochastic load variation is considered in System model by using the center-of-inertia (COI) approach which is represented as below [18]:

$$\dot{\tilde{\delta}}_i = \tilde{\omega}_i \tag{19}$$

$$M_i \dot{\tilde{\omega}}_i = P_{mi} - P_{ei} - \frac{M_i}{M_T} P_{coi} - D_i \tilde{\omega}_i + \sigma_i \varepsilon_i \tag{20}$$

Where $\tilde{\delta}_i$ and $\tilde{\omega}_i$ shows the center-of-inertia reference frame of the classical generator models, which are obtained by using following equations:

$$M_T = \sum_{i=1}^{3} M_i \tag{21}$$

$$\delta_0 = \frac{1}{M_T} \sum_{i=1}^{3} M_i \delta_i \tag{22}$$

$$\omega_0 = \frac{1}{M_T} \sum_{i=1}^{3} M_i \omega_i \tag{23}$$

$\tilde{\delta}_i = \delta_i - \delta_0$ and $\tilde{\omega}_i = \omega_i - \omega_0$ for i=1,2,3

$$P_{ei} = \sum_{j=1}^{3} E_i E_j \left(G_{ij} Cos(\tilde{\delta}_i - \tilde{\delta}_j) + B_{ij} Sin(\tilde{\delta}_i - \tilde{\delta}_j)\right) \tag{24}$$

$$P_{coi} = \sum_{i=1}^{3} (P_{mi} - P_{ei}) \tag{25}$$

$$P_{coi} = \sum_{i=1}^{n} P_{mi} - \sum_{i=1}^{n} E_i^2 G_{ii} - 2\sum_{i=1}^{n-1} \sum_{j}^{n} E_i E_j G_{ij} Cos\delta_{ij} \tag{26}$$

Using equation (9), the State Matrix using Model based method is obtained as following:

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -M^{-1}\left(\frac{\partial P_e}{\partial \tilde{\delta}}\right)_{coi} & & \frac{-D_1}{M_1} & 0 \\ & & 0 & \frac{-D_2}{M_2} \end{bmatrix} \tag{27}$$

Where, $J = -M^{-1}\left(\frac{\partial P_e}{\partial \tilde{\delta}} + \frac{M}{M_T} \frac{\partial P_{coi}}{\partial \delta}\right)$, for i=1,2

For diagonal elements (i=j), $\left(\frac{\partial P_e}{\partial \tilde{\delta}}\right)_{coi}$ is calculated as following:

$$\left[\left(\frac{\partial P_e}{\partial \tilde{\delta}}\right)_{coi}\right] = \sum_{k \neq i} E_i E_k \left[-G_{ik} Sin(\tilde{\delta}_i - \tilde{\delta}_k) + B_{ik} Cos(\tilde{\delta}_i - \tilde{\delta}_k)\right] + \frac{M_i}{M_T} \frac{\partial P_{coi}}{\partial \delta} \tag{28}$$

For off-diagonal elements (i≠j), $\left(\frac{\partial P_e}{\partial \tilde{\delta}}\right)_{coi}$ is calculated as,

$$\left[\left(\frac{\partial P_e}{\partial \tilde{\delta}}\right)_{coi}\right] = E_i E_j \left[G_{ik} Sin(\tilde{\delta}_i - \tilde{\delta}_k) - B_{ik} Cos(\tilde{\delta}_i - \tilde{\delta}_k)\right] + \frac{M_i}{M_T} \frac{\partial P_{coi}}{\partial \delta} \tag{29}$$

where , $\frac{\partial P_{coi}}{\partial \tilde{\delta}_i} = 2 \sum_{k \neq i}^{3} E_i E_k \left(G_{ik} Sin(\tilde{\delta}_i - \tilde{\delta}_k)\right)$

TABLE I. PARAMETER VALUES OF ALL THE 3 GENERATORS IN PER UNITS

Generating unit 1	Generating unit 2	Generating unit 3
$P_{mech1} = 0.72$	$P_{mech2} = 1.63$	$P_{mech3} = 0.85$
$Emf_1 = 1.057$	$Emf_2 = 1.050$	$Emf_3 = 1.060$
$M_1 = 0.630$	$M_2 = 0.340$	$M_3 = 0.160$
$D_1 = 0.630$	$D_2 = 0.340$	$D_3 = 0.160$

V. THE PROPOSED METHOD FLOW CHART

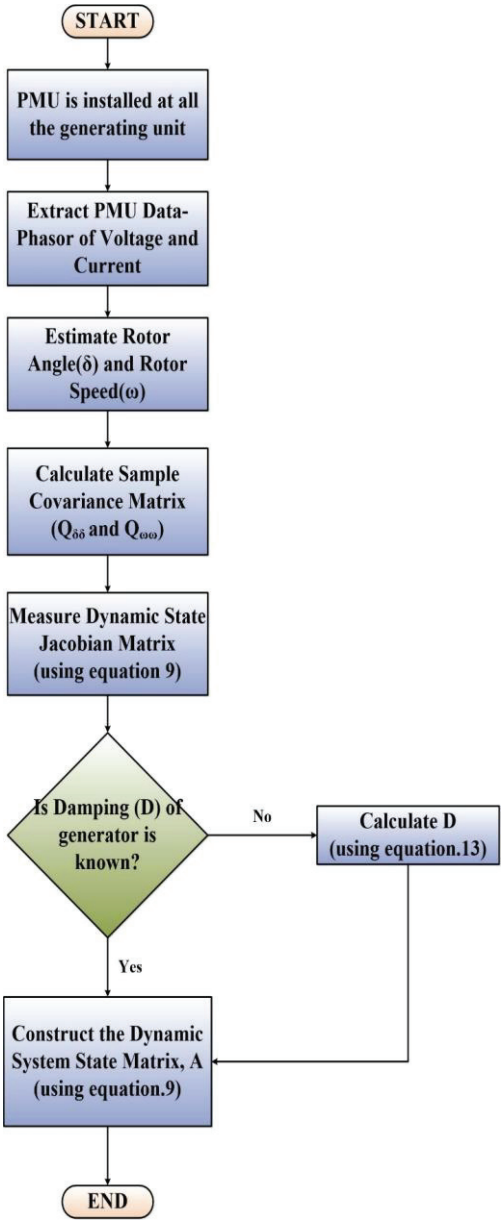


Fig. 2. Flow Chart of the proposed algorithm

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 06:27:18 UTC from IEEE Xplore. Restrictions apply.

VI. Results, Error And Plots

A standard deviation of stochastic load variations (σ= 0.01) is introduced in the simulated system. The trajectories obtained of the system rotor angle and rotor speed is oscillating around their respective steady state values.

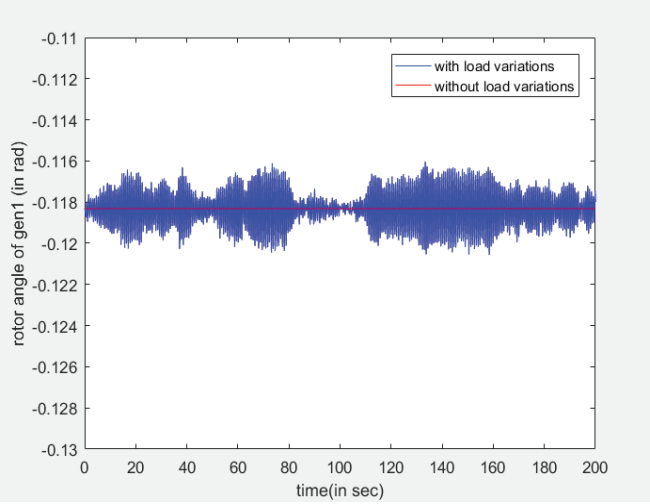


Fig. 3. Trajectory of generator-1 rotor angle ($\widetilde{\delta_1}$)

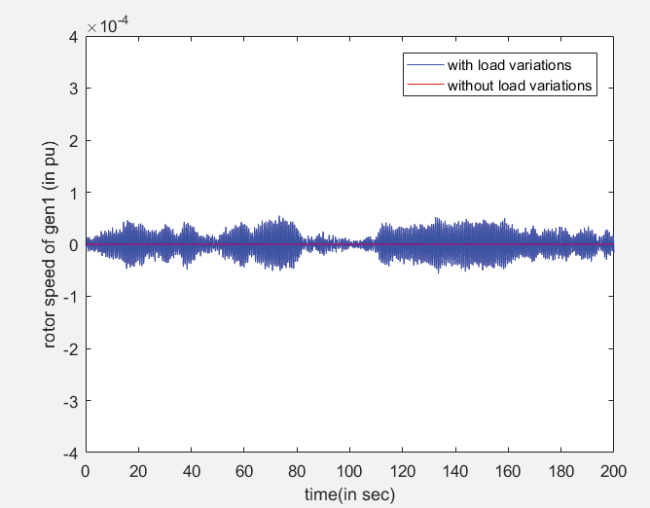


Fig. 4. Trajectory of generator-1 rotor speed ($\widetilde{\omega_1}$)

First, we calculate the Jacobian Matrix and system state matrix using model-based method, considering no load variations (σ = 0). Therefore, the Jacobian Matrix and system state matrix, A is obtained as:

$$\left(\frac{\partial P_e}{\partial \tilde{\delta}}\right)_{coi} = \begin{bmatrix} 6.3861 & 1.9868 \\ 3.2017 & 4.1392 \end{bmatrix} \tag{30}$$

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -10.1366 & -3.1536 & -1 & 0 \\ -9.41676 & -12.174 & 0 & -1 \end{bmatrix} \tag{31}$$

Now, the Jacobian Matrix and system state matrix, A is estimated using the proposed approach which is based on PMU measurements. The system is provided with stochastic load variations of σ = 0.01p.u. The sampling rate of data obtained from PMU is 50 samples per sec for 200sec.

Hence, the Jacobian matrix is calculated using equation (18):

$$\left(\frac{\partial P_e}{\partial \tilde{\delta}}\right)^{P}_{coi} = \begin{bmatrix} 6.2099 & 1.7363 \\ 3.4000 & 4.0649 \end{bmatrix} \tag{32}$$

$$A^{P} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -9.85698 & -2.75603 & -1 & 0 \\ -10.0000 & -11.9555 & 0 & -1 \end{bmatrix} \tag{33}$$

As we have observed that the estimated matrices are close to each other. Now the error is estimated as:

$$\frac{\left\| \left(\frac{\partial P_e}{\partial \tilde{\delta}}\right)^{P}_{coi} - \left(\frac{\partial P_e}{\partial \tilde{\delta}}\right)_{coi} \right\|_F}{\left\| \left(\frac{\partial P_e}{\partial \tilde{\delta}}\right)_{coi} \right\|_F} = 4.38\% \tag{34}$$

Where, $\| \ \|_F$ is known as Frobenius Norm which is used to calculate the matrix error percent.

Similarly we can also estimate the error for Dynamic System state Matrix, A as:

$$\frac{\left\| \left(\frac{\partial P_e}{\partial \tilde{\delta}}\right)^{P}_{coi} - \left(\frac{\partial P_e}{\partial \tilde{\delta}}\right)_{coi} \right\|_F}{\left\| \left(\frac{\partial P_e}{\partial \tilde{\delta}}\right)_{coi} \right\|_F} = 3.64\% \tag{35}$$

The estimated error indicated that the proposed method is provides the accurate results of calculated Jacobian matrix for stability purpose.

The Eigen Values are calculated of the both the estimated state matrix and plotted to show the proximity of instability by observing the right most eigen values.

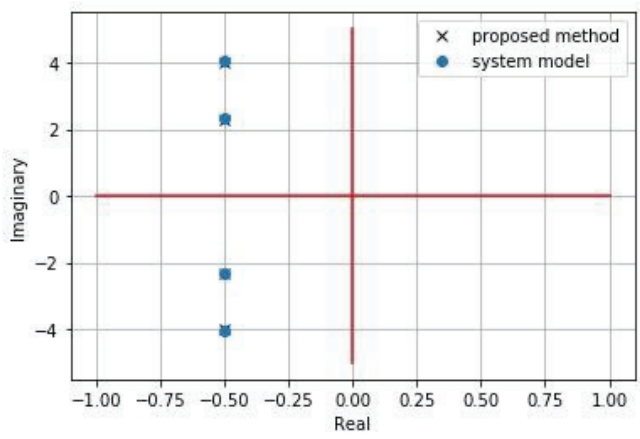


Fig. 5. Plotting of the eigenvalues of state matrix estimated by both the methods

VII. Conclusion

In this paper, we have estimated the dynamic state Jacobian matrix and system state matrix for analysis of ambient stability of the system. . For this purpose the standard WSCC 9bus, 3 machine system is simulated and the system is introduced with stochastic load variations. These matrices are estimated in near real time based on PMU data and then compared with the matrices calculated based on system model to show the reliability and accuracy of the proposed method The Eigen values are calculated and plotted to estimate the proximity of instability.

VIII. Appendix A

To calculate the Reduced admittance matrix the network admittance is reduced to all the generator buses and elimination of all other buses.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 06:27:18 UTC from IEEE Xplore. Restrictions apply.

The node voltage equation with ground reference is given as:

$$[I_{bus}] = [Y_{bus}]\,[V_{bus}] \tag{36}$$

$$\begin{bmatrix} I_1 \\ \vdots \\ I_n \\ I_{n+1} \\ \vdots \\ I_{n+p} \end{bmatrix} = \begin{bmatrix} Y_{11} & \cdots & Y_{1n} & Y_{1(n+1)} & \cdots & Y_{1(n+p)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & \cdots & Y_{nn} & Y_{n(n+1)} & \cdots & Y_{n(n+p)} \\ Y_{(n+1)1} & \cdots & Y_{(n+1)n} & Y_{(n+1)(n+1)} & \cdots & Y_{(n+1)(n+p)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ Y_{(n+p)1} & \cdots & Y_{(n+p)n} & Y_{1(n+p)(n+1)} & \cdots & Y_{(n+p)(n+p)} \end{bmatrix} \begin{bmatrix} V_1 \\ \vdots \\ V_n \\ V_{n+1} \\ \vdots \\ V_{n+p} \end{bmatrix}$$

Where, I_{bus} is the current injected in the bus, V_{bus} is the voltage of the bus with respect to ground and Y_{bus} is the bus admittance matrix.

To eliminate all buses other than the generator terminal buses, make there injected bus current to be zero.

Let us assume that $[I_1,\ldots, I_n]$ are the generating buses, equation (37) is rewrite as followings:

$$\begin{bmatrix} I_n \\ 0 \end{bmatrix} = \begin{bmatrix} Y_{pp} & Y_{np}^t \\ Y_{np} & Y_{nn} \end{bmatrix} \begin{bmatrix} V_n \\ V_p \end{bmatrix} \tag{37}$$

$$I_n = [Y_{pp} - Y_{np}^t Y_{nn}^{-1} Y_{np}]\, V_n \tag{38}$$

Therefore, the reduced admittance matrix is defined as,

$$Y_r = [Y_{pp} - Y_{np}^t Y_{nn}^{-1} Y_{np}] \tag{39}$$

IX. Appendix B

Let J= Jacobian matrix $=\frac{dP_e}{d\delta}$, using aforementioned equations (9),(10) and (11),

$$\begin{bmatrix} 0 & I_n \\ -M^{-1}J & -M^{-1}D \end{bmatrix} \begin{bmatrix} C_{\delta\delta} & C_{\delta\omega} \\ C_{\omega\delta} & C_{\omega\omega} \end{bmatrix} + \begin{bmatrix} C_{\delta\delta} & C_{\delta\omega} \\ C_{\omega\delta} & C_{\omega\omega} \end{bmatrix} \begin{bmatrix} 0 & -M^{-1}J \\ I_n & -M^{-1}D \end{bmatrix} =$$

$$(-1) \begin{bmatrix} 0 \\ -M^{-1}E^2G\mathcal{E} \end{bmatrix} [0 \quad -M^{-1}E^2G\mathcal{E}] \tag{40}$$

$$\begin{bmatrix} I_n C_{\omega\delta} & I_n C_{\omega\omega} \\ -M^{-1}(JC_{\delta\delta} + DC_{\omega\delta}) & -M^{-1}(JC_{\delta\omega} + DC_{\omega\omega}) \end{bmatrix} + \begin{bmatrix} I_n C_{\delta\omega} & M^{-1}(JC_{\delta\delta} + DC_{\delta\omega}) \\ I_n C_{\omega\omega} & M^{-1}(JC_{\omega\delta} + DC_{\omega\omega}) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & M^{-2}E^4G^2\mathcal{E}^2 \end{bmatrix} \tag{41}$$

On solving above equation,

$$C_{\delta\omega} + C_{\omega\delta} = 0 \tag{42}$$

$$C_{\delta\omega} + C_{\delta\omega}^T = 0 \tag{43}$$

$$C_{\omega\omega} - M^{-1}(JC_{\delta\delta} + DC_{\delta\omega}^T) = 0 \tag{44}$$

$$M^{-1}(JC_{\delta\omega} + DC_{\omega\omega}) + (C_{\delta\omega}^T J^T + C_{\omega\omega}D)M^{-1} = M^{-2}G^2E^4\mathcal{E}^2 \tag{45}$$

On solving equation (45), J is obtained as followings:

$$J = MC_{\omega\omega}C_{\delta\delta}^{-1} + DC_{\delta\omega}C_{\delta\delta}^{-1} \tag{46}$$

On putting equation (46) in (45),

$$C_{\delta\omega} = 0 \tag{47}$$

$$C_{\delta\delta} = \left(\frac{\partial P_e}{\partial \delta}\right)^{-1} M\, C_{\omega\omega} \tag{48}$$

$$C_{\omega\omega} = \frac{1}{2} M^{-1}D^{-1}\Sigma^2 \tag{49}$$

References

[1] P. W. Sauer and M. A. Pai, "Power system steady-state stability and the load-flow Jacobian," IEEE Trans. Power Syst., vol. 5, no. 4, pp. 1374–1383, 1990.

[2] J. D. L. Ree, V. Centeno, J. S. Thorp, and A. G. Phadke, "Synchronized phasor measurement applications in power systems," IEEE Transaction Smart Grid, vol. 1, no. 1, pp. 20–27, 2010.

[3] Monti, C. Muscas and F. Ponci, "Phasor Measurement Units and Wide Area Monitoring Systems," Elsevier, 2016.

[4] A. D. Angel, M. Glavic, and L. Wehenkel, "Using artificial neural networks to estimate rotor angles and speeds from phasor measurements" Proceedings of intelligent systems applications to power systems, ISAP03, 2003.

[5] X. Wang, J. W. Bialek and K. Turitsyn, "PMU-Based Estimation of Dynamic State Jacobian Matrix and Dynamic System State Matrix in Ambient Conditions," in IEEE Transactions on Power Systems, vol. 33, no. 1, pp. 681-690, Jan. 2018.

[6] O. A, Theresa, M. L. Crow, "An analysis of power system transient stability using stochastic energy functions". International Transactions on Electrical Energy Systems, vol. 23, no. 2, pp. 151-165, 2013.

[7] C. O. Nwankpa, S. M.Shahidehpour, and Z. Schuss, "A stochastic approach to small disturbance stability analysis". IEEE Transactions on Power systems, vol. 7, no. 4, pp. 1519-1528, 1992.

[8] H. Mohammed, and C. O. Nwankpa, "Stochastic analysis and simulation of grid-connected wind energy conversion system". IEEE Transactions on Energy Conversion, vol, 15, no. 1, pp. 85-90, 2000.

[9] B. Gao, G. K.Morison, and P. Kundur, "Voltage stability evaluation using modal analysis". IEEE Transactions on Power Systems, vol. 7, no. 4, pp. 1529-1542, 1992.

[10] Wang, X., & Chiang, H. (2014). Analytical Studies of Quasi Steady-State Model in Power System Long-Term Stability Analysis. IEEE Transactions on Circuits and Systems I: Regular Papers, 61, 943-956.

[11] J. H. Chow, Power system coherency and model reduction. London: Springer, 2013.

[12] H. D. Chiang, Direct Methods for Stability Analysis of Electric Power Systems-Theoretical Foundation, BCU Methodologies, and Applications. New Jersey: John Wiley & Sons, Inc, 2011.

[13] J. Machowski, J. Bialek, and J. R. Bumby, Power system dynamics and stability. John Wiley & Sons, 1997.

[14] P.M. Anderson, Power System Control and Stability. Hoboken: Wiley- IEEE Press,2003,p. 38-39,44.

[15] C. O. Nwankpa, S. M.Shahidehpour, and Z. Schuss, A stochastic approach to small disturbance stability analysis. IEEE Transactions on Power systems, vol. 7, no. 4, pp. 1519-1528, 1992.

[16] C. Gardiner, Stochastic Methods: A Handbook for the Natural and Social Sciences. Springer Series in Synergetics. Springer, Berlin, Germany, 2009.

[17] G. Ghanavati, P. D. H. Hines, and T. I. Lakoba, Identifying useful statistical indicators of proximity to instability in stochastic power systems. IEEE Transactions on Power Systems, vol. 31, no. 2, pp. 1360-1368, 2016.

[18] O. A, Theresa, M. L. Crow, An analysis of power system transient stability using stochastic energy functions. International Transactions on Electrical Energy Systems, vol. 23, no. 2, pp. 151-165, 2013.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 06:27:18 UTC from IEEE Xplore. Restrictions apply.

Volume 8 Issue 2 (2020) 106-109 ISSN 2347 - 3258
International Journal of Advance Research and Innovation

Study of Trapping and Intermixing of Delhi Drains for Rejuvenation of the River Yamuna

Prateek Shukla, SK Singh, Anunay Gour
Department of Environmental Engineering, Delhi Technological University, Delhi-42, India

Article Info
Article history:
Received 15 February 2020
Received in revised form
29 May 2020
Accepted 8 June 2020
Available online 15 June 2020

Keywords: Dissolved Oxygen, Intermixing of Sewage, Sewage Treatment Plant

Abstract: This study examines the presence of pollution of Yamuna in the city of Delhi, from two perspectives: (i) understanding the concept of trapping drains and (ii) establishing small scale Sewage Treatment plants at the point of outfall. With events occurring over the last century making river Yamuna polluted, the study focuses on the latest methods and technologies available to avoid intermixing of sewage and storm water drains to reduce pollution. Due to mixing of storm water and sewage, the desired qualities of the river water get degraded and the amount of dissolved oxygen decrease considerably. If this water is allowed to pass through a sewage treatment plant then it will lead to increase in time of treatment due to increase in discharge. The increased discharge can also affect the efficiency of the treatment plant. The study also includes the theory of tributaries of river Yamuna, Major Drainage problems in the region, Quality of River water.

1. Introduction

Yamuna is one of the largest tributary of Ganga and also one of the most important rivers of northern India. It passes through Uttarakhand, Himachal Pradesh, Haryana, Delhi, and Uttar Pradesh. It merges with the Ganga at Allahabad in Uttar Pradesh. At one time, it was the major lifeline for the people of the area, but today it is one of the most polluted rivers of the country which is being degraded day by day. The Yamuna starts getting polluted by pesticides and fertilizers as it enters Haryana, still most of the pollution occurs in Delhi. More than 18 million people live in Delhi. Yet it does not have a proper sewage disposal system. The capital lacks not only in collecting sewage from the residential areas but also in avoiding mixing of industrial and residential waste water.

Nineteen major drains from Delhi open into the Yamuna. At one time, these carried rainwater. But because of the poor sewage disposal system, water carrying sewage is discharged into these drains, from where it finds its way to the river. This leads to increase in the pollutant concentration in the river thereby decreasing its dissolved oxygen capacity considerably.

Delhi, along a stretch, the Yamuna is choked by water hyacinth a weed. This is an example of eutrophication. Dead fishes are also found in the river as soon as the monsoon begins. This is due to the sudden increase in pesticide and other pollutant levels. Industrial wastes also find their way into the river from large industrial units (22 in Haryana, 42 in Delhi and 17 in Uttar Pradesh) and many small industrial units. This industrial waste is harmful to people and the environment. Surprisingly, though Delhi constitutes only 2% of the catchment area, it is responsible for 80% of the pollution of the river.

2. Quality of River Water

Although the Yamuna River flows only for 54 KM from Palla to Badarpur through Delhi, the 22 KM stretch from Wazirabad to Okhla, which is less than 2% of the river length of 1370 Km from Yamnotri to Allahabad, accounts for about 76% of the pollution in the river. During the dry season, spreading over nearly nine month of the year, the river has no fresh water downstream of Wazirabad and the only flow available is sewage, both treated and untreated, flowing through 22 drains that join the river Yamuna all through its journey from end to end within Delhi. The setting up of Sewage Treatment Plants (STPs) and Common Effluent Treatment Plants (CETPs) is done to ensure that the sewage/industrial effluent does not find its way to drains which eventually joins the river, besides several other measures.

3. Measurement Units

The flow of all the drains was measured in MLD (millions of liters per day) and MGD (millions of gallons per day).



Fig. 1: Civil Military Drain



Fig. 2: Tonga Stand Drain

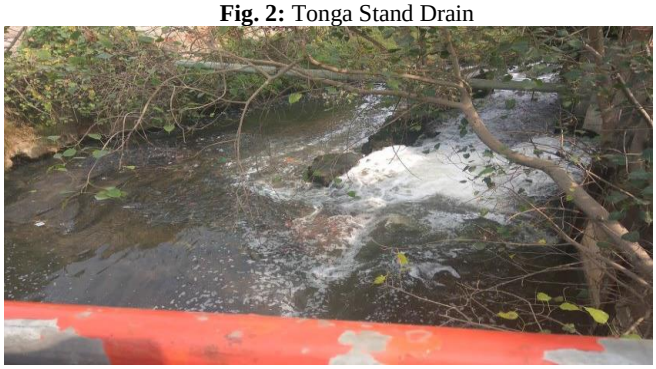


Fig. 3: Delhi Gate Drain



Fig. 4 :Metcalf House Drain(Bela Road)

***Corresponding Author,**
E-mail address: prateeks.dtu@gmail.com; sksinghdce@gmail.com; anunaygour@dtu.ac.in

All rights reserved: http://www.ijari.org

Table 1: Status of Drains out falling into River Yamuna in Delhi

Drain	Observations	Recommendations
Arunanagar Drain / Magazine Road Drain (4.5MLD)	Sewerage system from Arunanagar, old Chandrawal have a sewage outfall. Maintenance of point of interception is poor. Leakage/overflow of sewage into river. Solid waste accumulation at interception point.	Proper maintenance of outfall system of drain so that entire sewage is diverted into Arunanagar Sewage Pumping Station Providing proper outfall of sewage system of Arunanagar Provision for measurement of flow shall be made available.
Colony Drain (4MLD)	village/ J.J. Cluster. Leakage from interceptor sewer (Gate open). Freshwater outfall from Chandrawal WTP (Backwash). Solid waste accumulation at interception point. Poor maintenance of tapping point.	Proper maintenance of tapping point so that entire sewage diverts to Arunanagar SPS. Interception/collection of sewage from Provision for measurement of flow shall be made available.
Khyber Pass Drain (1MLD)	Sewage from Arunanagar, old Chandrawal Village, Chandrawal JJ cluster, and Tibetan Camp unauthorized occupants on Army land at Khyber pass etc. Maintenance of point of interception is poor. Leakage/overflow of sewage. Leakage of freshwater. Silting up of drain. Solid waste accumulation at interception point.	Proper maintenance Providing sewage system and outfall for JJ cluster Setting up constructed wetland system for treatment of sewage at source i.e. Chandrawal water works itself Provision for measurement of flow shall be made available.
Metcalf House Drain (Bela Road) (5MLD)	There is no outfall sewer line to cater to the sewage of Civil lines (Shamnath Marg/ Rajniwas etc.). Hence, sewage outfalls into this S.W. drain. Solid Waste accumulation. Leakage of waste freshwater. Faulty gate	Proper operation and maintenance Providing proper baffle for overflow Provision for measurement of flow shall be made available.
Mori Gate Drain (9.5MLD)	Leakage/overflow of Sewage from sewered area of Kashmiri gate, Old Delhi and Mori Gate area (Leakage from sewered area). Leakage from ISBT and area around. Solid waste accumulation at interception point.	Diversion of sewage to sewer line at source in the respective areas Providing root zone treatment before outfall to treat waste water Provision for measurement of flow shall be made available.
Tonga Stand Drain (1.5MLD)	Solid Waste accumulation. Leakage of sewage from Old Delhi Railway station area. Sewage and solid waste from unauthorised colony immediately before outfall. Open defecation near the point of outfall. No action plan by DJB to trap this drain.	Proper operation and maintenance of existing sewers Interception of sewage before outfall Provision for measurement of flow shall be made available.
Civil Military Drain (2.5MLD)	Leakage and overflow of sewage from sewered area of Chandni Chowk etc. area. Leakage of sewage from the gate of regulator. Open defecation near the point of outfall.	Proper operation and maintenance of existing sewers Providing root zone treatment before outfall or trap this sewage into trunk sewer line near Red fort Provision for measurement of flow shall be made available.
Delhi Gate Drain (15-17MGD)	Sewage from old Delhi i.e. Hauz Khas, Chawri Bazaar, Ballimaran and Darya Ganj area etc. Leakage Overflow of surplus sewage. Accumulation of some solid waste. Proper management of sewage. Entire discharge is being treated at two STPs of 17.2 MGD capacity.	Proper cleaning of drain operation and maintenance of existing treatment systems Provision for measurement of flow shall be made available.
Dr. Sen Nursing Home Drain (2.2MGD)	Sewage from railway colony area. Leakage of sewage from NDMC sewered area. Accumulation of solid waste. Overflow from the regulator gate. Inadequate treatment capacity of STP.	NDMC must divert sewage to sewer lines at source. Railway department must trap sewage into DJB sewer line near Tilak Marg or Deen Dayal Upadhyay Marg. Proper operation and maintenance of existing treatment system at STP Providing root zone treatment before outfall Provision for measurement of flow shall be made available.
Drain No. 14 Drain **(**1MLD)	Accumulation of solid waste Overflow of sub-soil water (freshwater) from construction activity at Pragati Maidan.	Sewage is being trapped into Ring Road Trunk sewer Provision for measurement of flow shall be made available.

Taimoor Nagar Drain (6MGD)	Leakage of sewage from adjoining colonies Sewage from Taimoor Nagar colony (Non-sewered) and sewered area of Defence Colony etc. Accumulation of solid waste.	Taimoor Nagar colony to be sewered Trap out-flowing sewage to CV Raman Trunk sewer should be made functional. Diverting sewage into sewer lines in sewered areas at source. Provision for measurement of flow shall be made available.
Tughlakabad Drain (4-5MGD)	To Sarita Vihar SPS, Flow of drain is partially trapped Operation and Maintenance of drains is improper. Partially trapped 10%, Minimum flow not available. Sewer coming from sewered area. Sewerage system not working properly. Wastewater from Tughlaqabad Govindpur.	O&M to be done properly, sump at zero level Provision for measurement of flow shall be made available. Coffer Dam for diversion of sewage to maintain minimum flow into Sarita Vihar SPS to Okhla STP, Flow meter not present, Weir type dam. Provision for measurement of flow shall be made available.
Tekhand Drain (6MGD)	Sewerage system does not exist. Flow is not trapped. STP of 10 MGD plant proposed. From Tejpur group of unauthorised colonies, Tehkhand Village.	Online or decentralised plant Sewerage system for unauthorised colony to be sent to decentralised plant Provision for measurement of flow shall be made available.
Barapullah Drain (30MGD)	Sewage from unsewered area of Sangam Vihar colony Leakage of sewage from NDMC area and sewered area of South Delhi. Accumulation of some solid waste.	Proposed STP of 20 MGD (as per information from DJB) Trapping of sewage into sewer lines at source to reduce flow of sewage into drains. Proper Maintenance of trapping points at Krishi Vihar, Andrews Ganj and Nizamuddin. Root zone treatment before outfall Provision for measurement of flow shall be made available.



Fig. 5: Khyber Pass Drain



Fig. 6:



Fig. 7: Arunanagar Drain

4. Conclusions

1. It was observed that most of the drains which were reports as plugged, actually, there was sufficient discharge going downstream except at Delhi Gate Drain.
2. It is observed that proper trapping of the sewage is not being done due to improper/poor operation & maintenance.
3. It is observed that most of the places a large amount of the solid waste was found in the drain which was causing obstruction of flow and un-aesthetic view.
4. It is recommended that it will be better to provide decentralized treatment of sewage at the outfall itself.
5. Monitoring of flow has not been done. Discharges given above are approximate and are subjected to verification.

Acknowledgments

The success of this study requires help and contribution from numerous individuals and the organization. Writing the report of this study work gives us an opportunity to express our gratitude to everyone who has helped in shaping up the outcome of the study.

We express our heartfelt gratitude to the Delhi Jal Board for providing all the visits of relevant locations. Their constant support and encouragement has made us realize that it is the process of learning which weighs more than the end result. We are highly indebted to the panel faculties during all the progress evaluations for their guidance, constant supervision and for motivating to complete the work. They helped me throughout by giving new ideas, providing necessary information and pushing me forward to complete the work.

References

[1]. AK Gosain R Khosa. Department of Irrigation and Flood Control, Government of Delhi Report on Drainage Master Plan for NCT of Delhi, 7, 2018.

[2]. MRJ Doorn, S Towprayoon, S Maria, M Vieira, W Irving, C Palmer, R Pipatti, C Wang. Wastewater treatment and discharge. In 2006 IPCC Guidelines for National Greenhouse Gas Inventories 5, 28, 2006, 1-6.

[3]. WMO, UNEP. Storm Water Audit, Sydney Water, 11, 2017

[4]. An Assessment of Small Scale Sewage Treatment Plant Technologies, India, 5, 2017

[5]. F Ghassemi, AJ Jakeman, HA Nix. Salinisation of land and water resources: Human causes, extent, management and case studies. Sydney, Australia, and CAB International, Wallingford: UNSW Press, 1995.

[6]. G Tchobanoglous, FL Burton, HD Stensel. Metcalf & Eddy, Inc. Wastewater Engineering: Treatment and Reuse (4^{th} ed.). McGraw-Hill, 2003. ISBN 978-0-07-112250-4.

[7]. Metcalf & Eddy, Inc. Wastewater Engineering. New York: McGraw-Hill, 1972. ISBN 978-0-07-041675-8.

[8]. SJ Burrian, et al. The Historical Development of Wet-Weather Flow Management. US Environmental Protection Agency (EPA). National Risk Management Research Laboratory, Cincinnati, OH, Document No. EPA/600/JA-99/275, 1999.

Journal of Intelligent Manufacturing
https://doi.org/10.1007/s10845-020-01710-x

Synthetic data augmentation for surface defect detection and classification using deep learning

Saksham Jain[1] · Gautam Seth[1] · Arpit Paruthi[1] · Umang Soni[1] · Girish Kumar[2]

Received: 29 October 2019 / Accepted: 29 October 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract
Deep learning techniques, especially Convolutional Neural Networks (CNN), dominate the benchmarks for most computer vision tasks. These state-of-the-art results are typically obtained through supervised learning, for which large annotated datasets are required. However, acquiring such datasets for manufacturing applications remains a challenging proposition due to the time and costs involved in their collection. To overcome this disadvantage, a novel framework is proposed for data augmentation by creating synthetic images using Generative Adversarial Networks (GANs). The generator synthesizes new surface defect images from random noise which is trained over time to get realistic fakes. These synthetic images can be used further for training of classification algorithms. Three GAN architectures are trained, and the entire data augmentation pipeline is implemented for the Northeastern University (China) Classification (NEU-CLS) dataset for hot-rolled steel strips from NEU Surface Defect Database. The classification accuracy of a simple CNN architecture is measured on synthetic augmented data and further it is compared with similar state-of-the-arts. It is observed that the proposed GANs-based augmentation scheme significantly improves the performance of CNN for classification of surface defects. The classically augmented CNN yields sensitivity and specificity of 90.28% and 98.06% respectively. In contrast, the synthetically augmented CNN yields better results, with sensitivity and specificity of 95.33% and 99.16% respectively. Also, the use of GANs is demonstrated to disentangle the representation space and to add additional domain knowledge through synthetic augmentation that can be difficult to replicate through classic augmentation. The proposed framework demonstrates high generalization capability. It may be applied to other supervised surface inspection tasks, and thus facilitate the development of advanced vision-based inspection instruments for manufacturing applications.

Keywords Surface defects · Classification · Convolutional neural network · Generative adversarial network · Deep learning

Introduction

In today's world, the industry is facing a challenge in the use of supervised learning algorithms for classification. Despite the various advantages of these learning algorithms, their performance decreases when trained on a real-world dataset. One reason is a shortage of training data due to factors that are typical to the manufacturing industry. With the advancements made in modern manufacturing processes and implementation of quality programs including six sigma, the number of good samples greatly exceeds the number of defective samples. The number of defectives are sometimes to the tune of parts per million (ppm). Therefore, a class imbalance exists in manufactured samples. However, if we discard the surplus of non-defective samples in order to keep class balance, the total dataset size is drastically reduced. Thus, to create a large dataset of defects, there is an opportunity cost involved due to having to manufacture in large numbers without addressing the defects and their causes. Therefore, manufacturing defect datasets are typically smaller.

Industrial anomaly detection is a highly challenging task because what constitutes a defect and what does not, can be considered a soft boundary, which can lead to false-negatives or false-positives depending upon the priority of classification. In addition to the shortage of the sample data, the available datasets can contain multiple categories of defects inside their positive (defective) class which can be difficult to model. One method to handle this classification problem is to train the classifier only on the positive (defective) sam-

✉ Umang Soni
umangsoni.1@gmail.com

[1] Netaji Subhas University of Technology, New Delhi, India

[2] Delhi Technological University, New Delhi, India

Published online: 18 November 2020

ples. In such a case, even if the data contains multiple classes of defects, there are no non-defective samples. This leads to a cautious classification boundary, as only when a sample does not correspond to any class of defects with a high enough confidence, is it classified as negative (non-defective) (Lai et al. 2018). However, this approach is still unable to handle the problem of small dataset size, which is particularly important when working with Convolutional Neural Networks (CNNs). The standard solution is the augmentation of the training data by rotating, reflecting, cropping, translating and scaling existing images to increase the dataset size and include some additional domain knowledge such as rotation-invariance, size-invariance, etc.

Deep neural networks have lower performances on small datasets when trained using convolutional techniques in comparison to traditional machine learning methods such as the support vector machine (SVM). But it has been experimentally proven that deep networks outperform such methods when trained on larger datasets (Feng et al. 2019). A significant reason for this parity is that SVMs rely on handcrafted features while deep neural networks are purely data-driven and as such, are better able to extract and utilise salient features from larger datasets for the classification task (Tabernik et al. 2020).

In this paper, authors investigate the use of GANs to sample additional synthetic data by modeling the training data distribution for manufacturing datasets. This synthetic data can be used for augmentation of the real data that is subsequently used to train the CNN for classification. Although GANs cannot produce images with perfect fidelity despite recent advancements, the authors of (Dosovitskiy et al. 2015; Richter et al. 2016) demonstrate marked improvements in results after synthetic augmentation of the data even with poorer fidelity images. Thus, expanding the dataset with these realistic, if slightly degraded, synthetic samples helps in reducing overfitting and improving the accuracy as well as the generalization capability of the CNN classifier (Madani et al. 2018). Authors verify these assertions and evaluate the suitability of the entire pipeline for real-world applications.

In Frid-Adar et al. (2018), the authors demonstrate promising results for medical image classification tasks by employing the use of GANs for data augmentation directly as a pre-processing step without any additional data. In this paper an attempt has been made to further build upon these works for surface defect detection and classification tasks.

Motivation and contribution

Apart from the typically smaller size of manufacturing datasets used in defect detection and classification tasks, two other specific problems occur. Features present in the intra-class defects may be random and highly diverse whereas the features in the inter-class defect may be nearly identical (Song and Yan 2019). Also, the defect images can vary because of the ambient illumination, slight material changes, and the severity of the defect that lead to a change in the grey value. This change can have an effect on the stability of the obtained defect features and may result in a reduction in the recognition rate.

While the application of GANs to expand the dataset with realistic but distinct fakes of the training data is straightforward, it is capable of handling the above two issues. Hand-crafted features are not required for deep learning algorithms. GANs also do not depend upon the user and instead model all the sources of variation (features) in the dataset, by learning the distribution of the training samples. Thus, as long as distinct features exist, even if they are too distributed for intra-class defect samples or too similar for inter-class defect samples, GANs have the potential to capture and model them correctly.

Data pre-processing involves transformation of the discrete distribution of training samples into a continuous distribution. It allows for GANs to model even challenging features present in the dataset and generate examples along the continuum. For example, if samples with different levels of severity of defects are used, the GAN would learn to distinguish between them and generate samples of all levels of severity, though they are not explicitly present in the training data. With recent advances made in GAN architectures (Chen et al. 2016), it has become possible to control the latent space of the generator which captures such information. Thus, if sufficient samples of the features affecting the grey values are generated, a more robust classifier can be trained.

Following the work of Madani et al. (2018) and Frid-Adar et al. (2018), the use of GANs as a pre-processing step for manufacturing tasks is proposed in this study. Further, the use of GANs for disentangling the representation space of the training data is explored. Thus, the objective of the paper is three-fold:

- To evaluate the efficacy of GAN architectures for synthetic augmentation of small manufacturing datasets, and to investigate GAN-based procedure as a standard augmentation technique during pre-processing.
- A simple CNN trained on synthetically augmented data is shown to outperform more complex CNN based feature extractors trained with classic augmentation.
- To explore and disentangle the latent space of the training data for obtaining a degree of control on the additional domain knowledge locked within the dataset that cannot be obtained through traditional classical augmentation techniques.

An illustration is provided to demonstrate the proposed augmentation and CNN classification pipelines and to evaluate their performance on the NEU-CLS dataset of surface

defects in hot-rolled steel strips. The NEU-CLS dataset contains 6 balanced classes of defective samples but no non-defective samples. Further details of the dataset are given in "Surface defects in hot-rolled steel strip" section.

Literature survey

Generative adversarial networks were first introduced in Goodfellow et al. (2014). They implemented discriminator and generator which were based on multi-layer perceptrons and were trained using backpropagation.

GANs have gained popularity in recent times, as numerous variations of GANs have been developed, which further improve the quality of the generated images and also widen the scope for countless applications. Deep convolutional generative adversarial network (DCGAN) constitutes a constrained architecture that reduces the gap between CNNs for supervised and unsupervised learning (Radford et al. 2015). Least square generative adversarial network (LSGAN) uses least square as a loss function for the discriminator which leads to higher quality image generation and more stable training (Mao et al. 2017). In a conditional generative adversarial network (CGAN), the discriminator and generator were conditioned to generate images based on class labels (Mirza and Osindero 2014). Conditional deep convolutional generative adversarial network (C-DCGAN) is a combination of DCGAN and CGAN. C-DCGAN utilizes convolutional neural network for feature extraction and with a conditional extension to augment data (Luo et al. 2020). The InfoGAN is based on an information theoretic formulation that allows for manipulation of the latent space (Chen et al. 2016). Also, InfoGAN uses a sleep–wake algorithm for training (Chen et al. 2016). Wasserstein generative adversarial network (WGAN) improves the training algorithm to avoid problems of mode collapse and provides better learning curves (Arjovsky et al. 2017). Auxiliary classifier generative adversarial network (ACGAN) employed label conditioning (Odena et al. 2017) while energy based generative adversarial network (EBGAN) involves the use of energy functions for discriminator and generator as a producer of low energy samples (Zhao et al. 2017). Boundary seeking generative adversarial network (BGAN) trains the discriminator with a heuristic of difference between target and resulting images (Hjelm et al. 2018) and boundary equilibrium generative adversarial network (BEGAN) uses an equilibrium based method on top of WGAN (Berthelot et al. 2017).

Some of the key applications of GANs include handwriting recognition (Tian 2017), facial age progression (Zhang et al. 2017), super-resolution imaging (Ledig et al. 2017), visual saliency prediction (Pan et al. 2017), object detection (Li et al. 2017), and unsupervised domain adaptation (Bousmalis et al. 2017). Owing to their excellent results in image synthesis, researchers have begun using GANs for data augmentation. These have been used initially to improve image quality and then for further training (Shrivastava et al. 2017), for generating different versions of an image (Antoniou et al. 2018), and even for imposing emotions on bland faces to augment smaller classes (Zhu et al. 2018). GANs have also find application in data augmentation for medical imaging tasks such as denoising CT scans (Wolterink et al. 2017), medical image segmentation, brain MRI segmentation and synthetic augmentation (Moeskops et al. 2017), and image segmentation of the liver (Yang et al. 2017), of skin lesions (Izadi et al. 2018) and of the brain (Alex et al. 2017).

In the manufacturing industry, computer vision is widely used with applications in, magnetic tile surface defects (Huang et al. 2018), additive manufacturing anomaly detection (Scime and Beuth 2018; Davtalab et al. 2020), steel surface defect inspection (Hao et al. 2020; Sun et al. 2016), defect detection in lithium ion battery electrodes (Badmos et al. 2020), roughness prediction (Grzenda and Bustillo 2019) etc. GANs models, especially after improvement of its training algorithm (Gulrajani et al. 2017), have been proposed for anomaly detection in structured and arbitrary textured surfaces (Lai et al. 2018) and unsupervised inspection of surfaces (Zhai et al. 2018).

Surface defects in hot-rolled steel strip

Steel-strip, being one of the most elemental components in the steel industry, has a high impact on the characteristics of the end product. Due to limitations in manufacturing processes, it encounters various defects like inclusion, surface defects, etc. (Song and Yan 2013). These surface defects can be caused due to many factors, which can be broadly divided into the following two categories:

- Defects on the continuous slabs introduced during the rolling process.
- Defects due to inappropriate rolling technique or foreign body inclusions in the strip-roll interface (Devadas et al. 1991; Yu et al. 2013).

Surface defect detection and classification

The main purpose of anomaly detection is to find out deviations or outliers from normal data. It has many industrial applications, from system health monitoring (spotting a malignant tumour in an MRI scan), to fault detection in manufacturing processes (through visual inspection of castings or welds). In the case of steel strips, the deviations can be in the form of surface defects. Many computer vision approaches based on feature extraction utilizing deep learning have been used for anomaly detection and classification (Lai et al. 2018). Owing to the data requirements

of deep learning models, and the typically small size of manufacturing datasets (and sometimes class imbalance), data augmentation is required. Basic data augmentation techniques can be graphical transformations such as translation, rotation, scaling, mirroring, etc. These may not be able to add sufficient information to the dataset after a certain limit. Hence, generative models like generative adversarial networks (Goodfellow et al. 2014), variational autoencoder (Kingma and Welling 2014), restricted Boltzmann machine (Carreira-Perpiñán and Hinton 2005), Helmholtz machine (Hinton et al. 1995) can be employed. Generative Adversarial Networks (GANs) is selected for this work. GAN consists of two convolutional neural networks, namely Generator and Discriminator. These are trained in an adversarial process, in which the generator generates fake images while the discriminator distinguishes between the fake and real images. GANs have recently gained popularity since good quality synthetic images can be generated using them. Detailed discussion on GANs is provided in "Synthetic augmentation" section.

Further, the details of the augmentation process for the surface defect dataset are provided in "Data augmentation" section, while for the detection and classification process we use the "Convolutional neural network" architecture as described in subsection.

Dataset

The dataset used to evaluate the effectiveness of the proposed model was generated by Northeastern University (NEU). This dataset includes six frequent hot rolling defects for steel strips: Rolled-in scale (RS), Patches (Pa), Crazing (Cr), Pitted Surface (PS), Inclusion (In) and Scratches (Sc).

Data set consists of a total of 1800 sample images for the above-mentioned defects, i.e., 300 sample images for each defect, with a resolution of 200×200 pixels and a grayscale value from 0 to 255. In this study, the intra-class defects have been defined as those having the same kind of defect while the inter-class defects are those having the different classes of the defect. Intraclass defects can be easily observed in Fig. 1, for example, different orientations of cracks. The inter-class defects may have identical aspects, for example, in RS, Cr, PS. Also, the defect images can vary due to illumination and material changes, (Song and Yan 2019).

Classification

Classification refers to a machine learning technique used to determine the class or categorization that a given object belongs to. The output of a classification algorithm is a discrete response with the predicted class as the output. Any architecture or algorithm for classification tasks is said to work by observing the labeled samples in the given training set, and then uses a set of criteria (whether rule-based or by learning over time) to attempt drawing conclusions in order to classify the samples in the test set. Provided with one or more inputs, a classification model attempts to predict the value (e.g. class labels) of one or more data samples (e.g. surface defects).

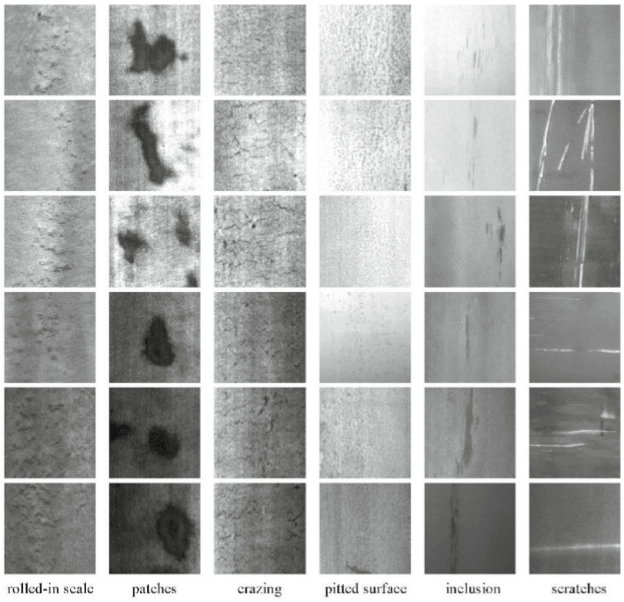


Fig. 1 Inter and Intraclass surface defects (Song and Yan 2019)

In this work, effort has been made to classify surface defects in hot-rolled steel strips using a Convolutional Neural Network-based deep learning approach. This approach is employed as the CNN has powerful architectures for image classification tasks, and the way forward for future applications (Deng et al. 2009). However, CNN typically requires massive amounts of data. Therefore, CNN-based classification becomes feasible for this study as the NEU-CLS dataset can be appreciably enlarged via data augmentation. Refer "Data augmentation" section for detailed discussion.

Convolutional neural network

In fully connected neural networks (FNNs), each neuron or node in all layers (disregarding the input layer), is connected to every neuron in the preceding layer. Matrix multiplication is utilized to obtain the weights and biases for each neuron, and there are thousands and thousands of such neurons in a deep network. This makes the use of FNNs not only very computationally intensive but also requires a large amount of time to successfully train them. To overcome this, CNN was proposed as an extension to standard deep neural networks. CNN uses an operation known as a convolution in their layers rather than matrix multiplication, and as a result, substantially decreases the number of parameters to train in the network, along with achieving a faster forward-propagation process

that makes their training and implementation faster. This is advantageous because the learned convolutional filters are consequently used for the entire image. Another major reason for their widespread use and success in most computer vision tasks is that all optimization techniques (e.g. Backpropagation with stochastic gradient descent) used in standard FNNs are also applicable for CNNs.

Implementation details for baseline

The CNN which was used for classification had 3 convolutional layers. The first and second convolutional layers had 32 filters of size 3×3 while the third convolutional layer had 64 filters of size 3×3. The Rectified linear unit (ReLU) activation function was used for all the layers. The fully connected layer had 256 neurons. The model was trained for images with 3 colour channels, i.e., Red Green Blue (RGB). In the CNN, Adam optimizer is used with softmax cross-entropy with logits as the cost function. The optimizer had a learning rate of 0.0001. For training and validation, a train-test split of 9000 images and 1800 images respectively was used while keeping the classes balanced. During the testing time, the same validation set was utilized for all experiments. The data consists of 6 classes with the mapping: Cr: 0, In 1, PS: 2, Pa: 3, RS: 4, Sc: 5 which was used for classification. The training of the model was done in batches of size 60 for 20 epochs (Fig. 2).

Implementation details for fine tuning

The above architecture was pre trained on a smaller subset[1] of the imagenet dataset (Deng et al. 2009) and consequently fine-tuned on the actual 9000 training images and 1800 validation images, keeping the training conditions the same.

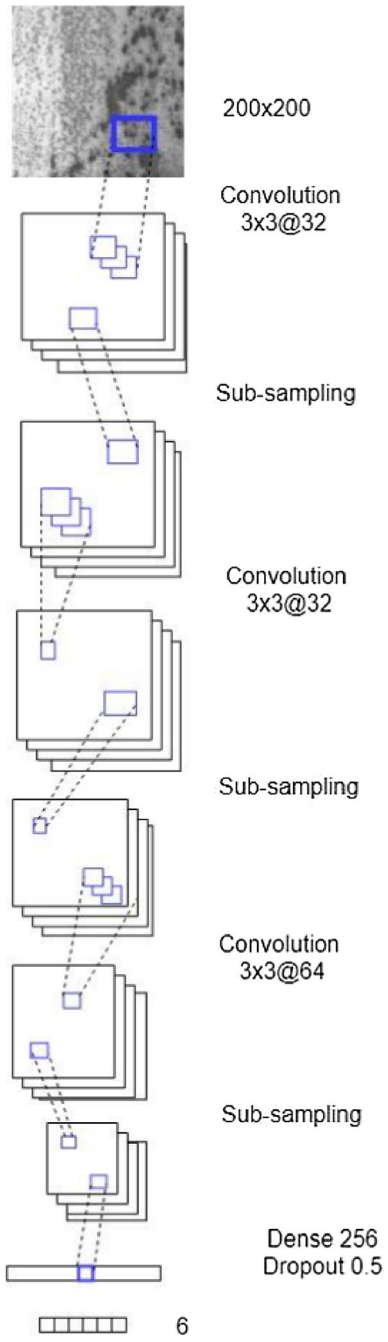


Fig. 2 Architecture of the proposed CNN

Data augmentation

Data augmentation is a procedure carried out to enlarge a small input training set to get more effective results. Most important information can be extracted from the small training set when it is subjected to the augmentation process, because it allows us to include some knowledge about potentially never-before-seen data during training, without having to collect additional data. For example, if the end application is such that image features can appear in any orientation during the testing phase, it would be useful to introduce randomly rotated samples of the existing dataset into the training procedure.

In the data augmentation operation, the additional input training samples can be created by two methods: (i) geometrically transforming (usually, affine transformations are performed) the given training samples to increase input variety, in a process known as classic transformation, or (ii) generating new never-before-seen samples using generative models trained on the existing input training sample, in a process known as synthetic data augmentation.

To be successfully benefitted from data augmentation techniques, it is important to keep in mind the details and specifics of the dataset, the network framework, and the task, because haphazard or unfit augmentation rules and opera-

tions may lead to even worse performance of the model than if the dataset had not been augmented.

The following sections, first the classic augmentation operation along with its implementation has been described and then our novel approach to synthetic augmentation using generative adversarial networks for surface defect detection and recognition task has been illustrated.

Classic augmentation

Thousands of parameters must be trained for even small CNNs. When deep learning architectures with multiple layers are built, or when the training set is limited in terms of the number of images, there is a high likelihood of overfitting. One of the standard solutions to resolve the problem of overfitting is to artificially enlarge the dataset via data augmentation. For greyscale images, classic augmentation refers to data augmentation techniques that involve geometric coordinate transformations, e.g. random translation, random rotation, scaling, random flipping, elastic deformation, and shearing. For our particular application—surface defect detection and recognition –elastic deformation or shearing has not been used so as to protect the useful surface defect features or characteristics that depend upon the shape of the image.

Implementation details

The dataset was expanded using the classic data augmentation method. The following transformations were applied to the dataset:

- The images were randomly rotated in clockwise or anticlockwise direction with an angle ranging from 0° to 180°.
- The images were randomly flipped about the vertical or horizontal axes.
- The images were translated by random sampling of [x,y] pixel values in a set range.
- The images were randomly rescaled to a factor between 0.5 to 1.5 times.

Ultimately, all images were resized to the original size of 200×200 pixels by bicubic interpolation. With the help of these transformations, the size of the training set was increased to 9000 images. Three such training sets were generated, and the mean score is reported in "Results and discussion" section.

A major issue that arises with classic augmentation is that the improvement offered saturates after a point (Richter et al. 2016). This is because no new additional information is actually being added with classic augmentation, and after a while, the model becomes robust to the changes brought on by simple transformations since only a limited set of operations can be performed. Thus, to introduce more diversity into the training samples, synthetic data augmentation is required.

Synthetic Augmentation

Synthetic data augmentation differs from classic augmentation where the additional images are synthetically generated fakes, which closely resemble the actual input images rather than being the transformed input images. Although data augmentation for computer vision tasks has long existed, the advent of advanced GANs has provided a unique opportunity for taking it further.

A correctly trained GAN can augment each source of variance in the dataset, simultaneously obtaining a continuous distribution of training samples by transforming the discrete distribution. If a GAN is trained on sufficient samples at different states of rotation, it will learn to generate samples at any state of rotation. Therefore, a GAN trained on classically augmented data can learn to replicate its effects and include additional domain knowledge such as—translation-invariance (e.g. by horizontal or vertical-shift) or size-invariance (e.g. by random resize), etc. However, the true power of GANs comes to light when considering more challenging features such as the severity of, or size of the image area affected by, the defect. It would be extremely difficult to model the shape, impact, and size of the defect and then use just affine transformations to produce augmentations. However, given sufficient samples of different discrete levels of severity, the GAN will infer the model on its own and learn to generate samples along the continuum of severity. Since CNNs are powerful discriminators, any GAN models will require careful and stable training, not to mention correctly tuned parameters to obtain usable results. While there is no good theoretical basis for how to select hyperparameters for training stable GAN models, which remains an open problem, there exists empirically demonstrated heuristics that work in practice. The authors of DCGAN (Radford et al. 2015) provide recommendations for stably training GANs which are considered as the defect starting point. In this paper, authors use the default values for the hyperparameters, as described in the respective GAN implementations to eliminate the need for optimal hyperparameter search. These values are referred to as out-of-box hyperparameters. This reduces the overall complexity of the pipeline and also demonstrates that even without hyperparameter tuning, obtaining synthetically augmented data that outperforms classical augmentation techniques is possible.

In this section, authors describe three approaches to synthetic data augmentation that utilize the improvements to the original GAN framework, namely Deep Convolutional GAN (DCGAN), Auxiliary Classifier GAN (AC-GAN) and Information-theoretic GAN (InfoGAN) for generating image samples for augmentation.

DCGAN

DCGAN was proposed by Radford et al. (2015), aiming to improve the training stability along with the quality of generated images over Goodfellow's vanilla GAN model. DCGAN better incorporates the recently made advances for training CNNs and uses fully-convolutional neural networks in its architecture.

An important additional attribute of DCGAN is that the discriminator can be used as a powerful tool for feature extraction, as demonstrated in the paper (Radford et al. 2015). Its use, especially for unsupervised image classification tasks, achieves performance that is competitive with the state-of-the-art unsupervised models on benchmark problems such as CIFAR10, etc.

Architecture overview Stride convolutions are used rather than pooling layers, and transposed convolutions are used in place of up sampling so as to make both discriminator and generator networks fully-convolutional. Figure 3 illustrates the DCGAN generator framework as proposed by Radford et al. (2015). Random numbers taken from a uniform distribution (which is learned during the training) are sent as input into the generator network, which encodes this noise into a surface defect image, taken as the generator output. The generator framework contains 4 transposed convolutional layers that up sample the image using a filter of the specified size, and a fully connected layer reshaped to the requisite size. The transposed convolutional layer, termed as the deconvolutional layer can extend the image pixels by pushing zeroes into the intermittent locations. Performing the convolution operation across this extended image leads to the generator outputting an image of a larger size. Batch-normalization is added to all convolutional layers of the framework, with the final generator layer being the exception. The DCGAN learning procedure is stabilized by normalizing the responses to contain variance equal to 1, and mean equal to 0 across the complete mini-batch. This also prevents full or partial mode collapse for the generator framework. Figure 4 shows a batch of images for the Inclusion class generated by the trained DCGAN. It is important to note that the discriminator architecture is simply the generator architecture mirrored.

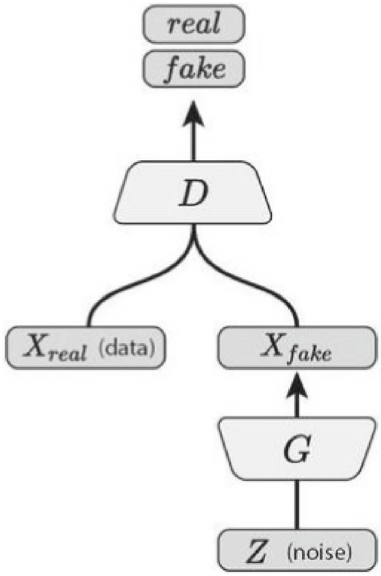


Fig. 3 Diagrammatic representation [figure taken from Frid-Adar et al. (2018)] of the DCGAN Architecture as per the original implementation given in Radford et al. (2015)

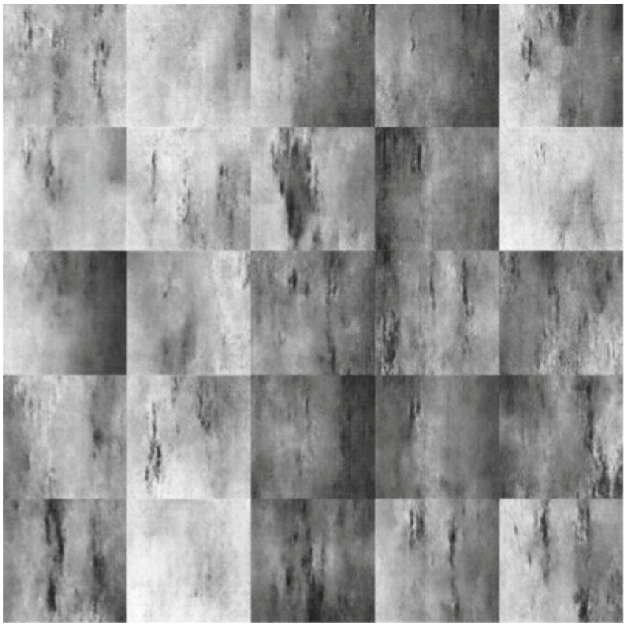


Fig. 4 Images generated by DCGAN

Implementation details The DCGAN was trained separately on the input data in batches of 25 images. The learning rate was set to 0.0002 and momentum (β_1) was set to 0.5. The discriminator of the model had 5 hidden layers, out of which 4 had leaky ReLU as an activation function while the final had a linear activation function. The generator of the network had 3 hidden layers with ReLU as the activation function for all the layers. The DCGAN model was trained for 100 epochs, and 24 synthetic batches of the same batch size were generated for each class. Each batch was generated in the form of a grid of 5×5, which was sliced using the *image_slicer* module available in python. Richter et al. (2016) has been referred to train the DCGAN on a 5400 images-strong randomly selected subset (keeping classes balanced) of the classical augmentation training set, to generate 3600 samples. Three such training sets were generated and the mean score has been reported in "Results and Discussion" section. The size of the resultant training set was kept

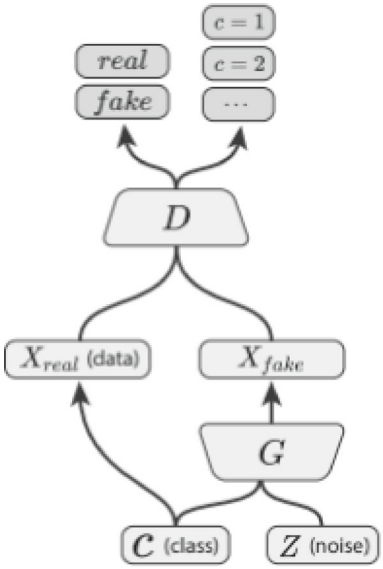


Fig. 5 Diagrammatic representation of the AC-GAN Architecture as per the original implementation given in Odena et al. (2017)

the same so as to fairly compare the results to those of classic data augmentation. DCGAN requires a separate model to be trained in each class, which makes it much more cumbersome to use than classical augmentation. In order to do away with tuning the hyperparameters multiple times, same hyperparameters have been used for all models.

AC-GAN

The second GAN model employed for this study is known as the Auxiliary Classifier GAN (ACGAN) (Odena et al. 2017). It is used for conditional image synthesis and was formulated as an extension to Conditional GANs (Mirza and Osindero 2014). Figure 5 illustrates the typical architecture for this model. ACGAN allows model conditioning on information obtained from an external data source in order to increase the generated image sample quality. Additionally, it is perfectly capable of incorporating the provided class labels from the training set and thus produces labeled image samples. This is useful for synthetic augmentation of datasets for supervised learning tasks, e.g. surface defect detection and recognition tasks.

Architecture Overview The architecture of ACGAN does not appear much different from Goodfellow's GAN. However, ACGAN has major differences in the training procedure and function performed by the discriminator and generator. The ACGAN discriminator is best served by being assigned the task of reconstruction of side information rather than simply inputting the side information into the framework (Richter et al. 2016). To carry out this, the discriminator is so adjusted as to consist of an auxiliary decoder neural architecture. Consequently, it not only provides a decision on whether the image is real or fake but also proffers the class label as an output. The ACGAN is employed in this work for synthetic data augmentation requirements with only slight modifications for the six classes. The auxiliary decoder in the discriminatory framework successfully classifies all six classes of surface defects when the training procedure as described above is followed.

Implementation details The ACGAN was trained on the input data in batches of size 25. The learning rate and β_1 were set at the same level as DCGAN, i.e., 0.0002 and 0.5 respectively. The discriminator of our ACGAN had 4 hidden layers with leaky ReLU activation function for 3 layers and linear activation function for 1 layer. The generator of the network also had 4 hidden layers with the ReLU activation function for all the layers. The ACGAN model was also trained for 100 epochs and 24 synthetic batches of the same batch size were generated for each class. Each batch was generated in the form of a grid of 5×5, which was sliced using the *image_slicer* library available in python. Once again, ACGAN was trained on a 5400 images-strong randomly selected subset (keeping classes balanced) of the classical augmentation training set, to generate 3600 samples. Three such training sets were generated and the mean score is reported in "Results and discussion" section.

InfoGAN

While changing the noise vector in GANs, it has consistent and meaningful effects on the generator output (Radford et al. 2015). There is no systematic way to find these structures because the generator uses the in a highly entangled way. Therefore, individual dimensions of z show no correspondence to semantic features in the data. The main idea of InfoGAN is to disentangle the representation space so as to decompose the domain knowledge existing in the data into a set of meaningful factors of variation (Chen et al. 2016).

The generator input is split into two parts: the noise vector and something referred to as the latent code vector. The latent code targets the structured semantic features in the data, made meaningful by maximizing the mutual information between the code and the generator output.

Since both the noise vector, z and latent code vector, c are provided to the generator, the form of the generator is now G(z,c), and the latent factors can be discovered in an unsupervised manner. Unlike in the original GAN, InfoGAN contains an information-theoretic regularization to handle the problem of the triviality of codes, i.e. the latent codes c and the distribution of the generator G(z,c) should have high mutual information (Fig. 6).

Overall, the image quality may not be as good as obtained from DCGAN. The major benefit of InfoGAN is that the dis-

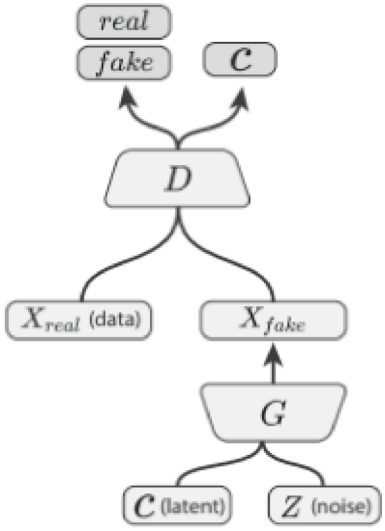


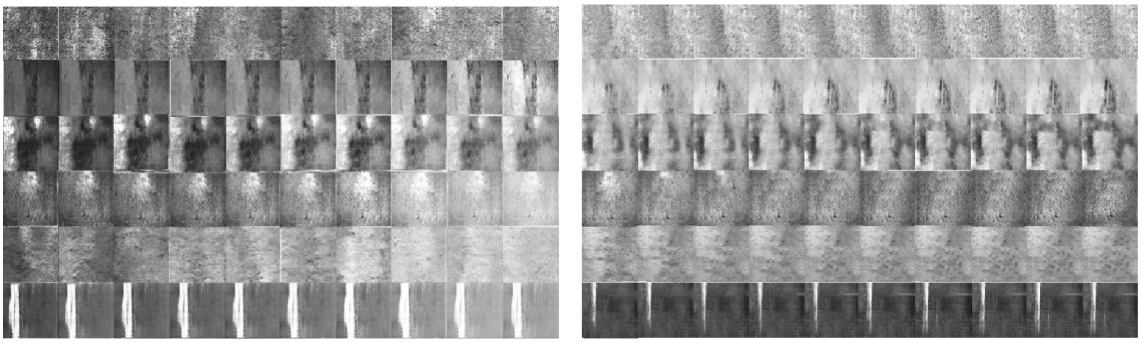
Fig. 6 Diagrammatic representation [figure taken from (Odena et al. (2017)] of the InfoGAN Architecture as per the original implementation given in Chen et al. (2016)

crete latent code allows for the conditional generation, and the continuous latent code allows for disentanglement of the latent space and control over the latent factors, as demonstrated in Fig. 7.

Architecture overview The architecture of the InfoGAN is effectively the same as the original GAN. However, it contains an extra regularization term over the original GAN's objective function. In practice, calculating mutual information is hard so standard variational arguments are used to approximate a lower bound. This involves introducing an auxiliary conditional distribution $Q(c|x)$, modeled by a parameterized neural network, and is meant to approximate the real conditional likelihood $P(c|x)$. Q and D as in the original GAN, share all convolutional layers, and one final fully connected layer output parameter for $Q(c|x)$.

InfoGAN contains a hyperparameter λ in the extra regularization term. It is easily tunable and set to 1 for discrete latent codes, whereas it's smaller is used for continuous latent codes. That ensures the L1 regularized term is on the same scale as the original GAN objectives. Since GANs are difficult to train, other approaches available on DCGAN in the literature are used to stabilize InfoGAN training (Chen et al. 2016).

Implementation details The InfoGAN was trained on the input data in batches of 30 images. The learning rate was set to 0.003 and, β_1 and β_2 were set to 0.5 and 0.999 respectively. The discriminator of the model had 3 hidden layers, with leaky ReLU as the activation function for all the layers. The generator had 5 hidden layers, out of which 4 had ReLU as an activation function while the final layer had a tanh activation function. The InfoGAN model was trained for 120 epochs, and synthetic batches of the same batch size were generated for each class. The grids were sliced using the *image_slicer* module available in python. In this experiment, two uniform continuous variables were used as latent codes. For the discrete latent code, the softmax nonlinearity was used, while for continuous latent codes it was found that simply treating the auxiliary distribution as a factored Gaussian was effective. The InfoGAN was also trained on a 5400 images-strong randomly selected subset (keeping classes balanced) of the classical augmentation training set, to generate 3600 samples, keeping classes balanced and randomly varying the continuous latent codes to synthetically



(a) *Latent code: Lighting and*

(b) *Latent code: Severity*

Fig. 7 **a** Latent code: Lighting and **b** Latent code: Severity. The discrete latent code ranging from 0 to 5 captures the 6 classes in the NEU-CLS dataset. The continuous latent codes vary from −1 to 1 and the effects are shown from left to right. In (**a**), the continuous latent code captures the variation in lighting; In (**b**), the categorical code controls the severity of the defect: for example, in the 2nd row which contains the inclusion class, the size of the affected area seems to grow from left to right. Similarly, in the 6th column, the scratched surface area also seems to grow from left to right. This indicates that the InfoGAN has learned to separate the size of the affected area by looking at a sufficient number of samples in the data distribution. This can help us control and add samples with additional domain knowledge into the augmented dataset

add additional domain knowledge (Fig. 7). Three such training sets were generated, and the mean score is reported in "Results and discussion" section.

Algorithm-GAN-based data augmentation

- Image data $x_1, x_2, x_3, \ldots, x_i$ with i images and their respective labels $y_1, y_2, y_3, \ldots, y_i$ were initially present. The total number of t classes was present. Therefore, $\{y_1, y_2, y_3, \ldots, y_i\} \in \{0,1,2,3,\ldots, i\}$.
- Build GANs for generating samples of known classes (in the case of DCGAN, t separate models for each class).
- Use this data to train a robust CNN classifier.
- Generate classically augmented images using the training data of j images per class.
- A subset of the resultant data with k images per class is used to train the GANs.
- Newly generated images $X_1, X_2, X_3, \ldots, X_k$ along with their corresponding class labels $Y_1, Y_2, Y_3, \ldots, Y_k$ are used to train the CNN.

Results and discussion

In the present study, an attempt has been made to compare the performance of CNNs for classification of surface defect data over the course of multiple experiments, where the CNNs are trained using different augmentation pipelines. An experiment is also performed with a pretrained CNN fine-tuned on the surface defect data, achieving state of the art results. In classification models, the confusion matrix is used for assessing the model performance.

All evaluation metrics for multi-class classification problems can be easily understood in the context of the binary classification problem (where the labels are either 0 or 1) when considering each class one-by-one. Basic terminology used in the confusion matrix is given as:

- *True Positive* (TP) Cases when the actual label is 1 and the predicted label is also 1, i.e., positive label was correctly predicted.
- *True Negative* (TN): Cases when the actual label is 0 and the predicted label is also 0, i.e., the negative label was correctly predicted.
- *False Positive* (FP): Cases when the actual label is 0 and the predicted label is 1. They are also known as 'Type I errors'.
- *False Negative* (FN): Cases when the actual label is 1 and the predicted label is 0. They are also known as 'Type II errors'.

The following terms are used to interpret the confusion matrix

- *Accuracy* It tells us how many labels were correctly predicted out of all the predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- *Sensitivity* It tells us how often, the model predicts the label as 1 when it is actually 1. It is also known as 'True positive rate' or 'Recall'.

$$Sensitivity = \frac{TP}{TP + FN}$$

- *Specificity* It tells us how often, the model predicts the label as 0 when it is actually 0. It is also known as 'True negative rate'

$$Specificity = \frac{TN}{TN + FP}$$

The proposed model performs multiclass classification of labels: RS, Pa, Cr, PS, In and Sc. The synthetically generated datasets were tested using simple CNN for the performance measures defined above. In the first experiment (Fig. 8), CNN was trained on dataset augmented using classic methods. The test accuracy was found out to be 90.28% for 20 epochs with a batch size of 60. Table 1 describes the confusion matrix for the first experiment. It can be observed from the confusion matrix that the model has good sensitivity for Cr, In, PS, and RS, but it is quite low for Pa and Sc. This results in a number of misclassifications for Pa and Sc. The specificity remains in an acceptable range for this model.

In the second experiment (Fig. 9), the CNN was trained for the DCGAN augmented dataset of the same size under similar training conditions, i.e., batch size of 60 for 20 epochs. The test accuracy was found out to be 95.78%, which is better for this dataset using a simple CNN without transfer learning. The sensitivity was within the acceptable range for all the classes, though it reduced insignificantly. However, the overall misclassifications were reduced by a significant amount. The specificity remained approximately the same for both the experiments.

In the third experiment (Fig. 10), CNN was trained for the ACGAN augmented dataset of the same size under the same training conditions and the test accuracy was found out to be 92.78%.

In the fourth experiment (Fig. 11), CNN was trained on the InfoGAN augmented dataset of the same size under the same training conditions. This experiment showed the second-highest test accuracy (lower than only the pipeline with DCGAN augmentation) of 94.86%. Despite the InfoGAN

Actual/Predicted	Cr	In	Pa	PS	RS	Sc	Sensitivity	Specificity
Cr	276	0	9	8	7	0	92.00%	99.07%
In	7	261	17	7	8	0	87.00%	96.07%
Pa	4	26	253	0	0	17	84.33%	97.27%
PS	3	0	1	296	0	0	98.67%	99.00%
RS	0	0	5	0	281	14	93.67%	99.00%
Sc	0	33	9	0	0	258	86.00%	97.93%

Fig. 8 Confusion matrix for experiment-1 (Classic Augmentation)

Table 1 Comparison of classic augmentation, DCGAN, ACGAN, and InfoGAN for CNN baseline

Augmentation method	Sensitivity		Specificity		Accuracy (%)
	Mean (%)	Std. Dev (%)	Mean (%)	Std. Dev (%)	
Classic augmentation	90.28	4.99	98.06	1.11	90.28
DC-GAN augmentation	95.33	3.03	99.16	0.37	95.78
AC-GAN augmentation	92.28	1.76	98.56	0.38	92.78
InfoGAN augmentation	94.06	2.78	98.81	0.36	94.86

Actual/Predicted	Cr	In	Pa	PS	RS	Sc	Sensitivity	Specificity
Cr	295	0	0	5	0	0	98.33%	99.20%
In	0	269	14	0	0	17	87.00%	99.20%
Pa	4	5	284	0	7	0	94.67%	98.87%
PS	3	0	3	289	1	4	96.33%	99.67%
RS	4	4	0	0	291	1	97.00%	99.47%
Sc	1	3	0	0	0	296	98.67%	98.53%

Fig. 9 Confusion matrix for experiment-2 (DCGAN Augmentation)

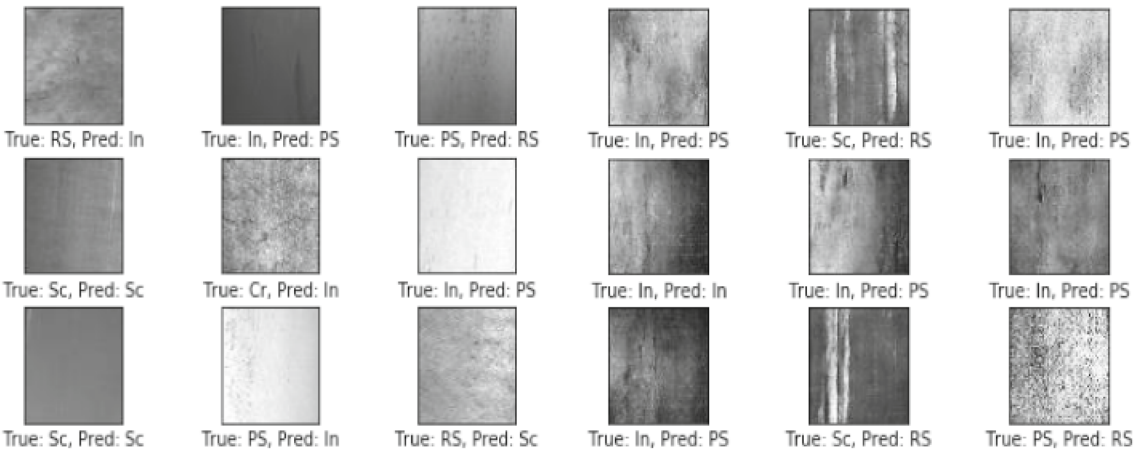
Actual/Predicted	Cr	In	Pa	PS	RS	Sc	Sensitivity	Specificity
Cr	280	0	0	11	9	0	93.33%	98.33%
In	0	270	14	0	0	16	87.00%	98.33%
Pa	7	9	273	0	11	0	91.00%	98.80%
PS	3	0	4	283	3	7	94.33%	99.27%
RS	9	7	0	0	279	5	93.00%	98.47%
Sc	6	9	0	0	0	285	95.00%	98.13%

Fig. 10 Confusion matrix for experiment-3 (ACGAN Augmentation)

Actual/Predicted	Cr	In	Pa	PS	RS	Sc	Sensitivity	Specificity
Cr	289	0	0	8	3	0	96.33%	98.53%
In	0	271	10	0	0	19	90.33%	98.67%
Pa	2	10	279	0	9	0	93.00%	99.13%
PS	11	0	3	273	9	4	91.00%	99.47%
RS	9	3	0	0	288	0	96.00%	98.60%
Sc	0	7	0	0	0	293	97.67%	98.47%

Fig. 11 Confusion matrix for experiment-4 (InfoGAN Aug.)

augmented dataset containing additional domain information, the accuracy was lesser than that of the DCGAN augmented pipeline. This is due to the image fidelity being lesser.



(a) *Classic augmentation* **(b)** *DC-GAN augmentation*

Fig. 12 Predicted examples

Table 2 Comparison of the feature extraction approaches for classification, namely AECLBP, SCN, Overfeat, DCGAN augmented CNN baseline, and DCGAN augmented finetuned CNN

Classification method	Accuracy (%)
AECLBP	97.22
SCN	98.60
Overfeat	98.70
CNN Baseline	95.78
Finetuned CNN	99.11

Figure 12 shows predicted examples from classic augmented model and DC-GAN augmented model.

Table 2 analyses the results of different augmentation methods used in our study. Training the dataset using the DC-GAN augmentation method, yielded better results compared to the classic augmentation method as well as AC-GAN and InfoGAN augmentation methods. The CNN trained using DC-GAN augmentation has a better sensitivity than the other two methods, which signifies that it can predict positive labels more accurately. It also has higher specificity which implies that the CNN trained using the DC-GAN augmentation method has a higher negative label prediction ability. It also has higher overall accuracy as a result of its higher overall classification ability.

From Table 1 it can be inferred that DCGAN Augmentation gives the best results. Therefore, in the fifth experiment (Fig. 13), the pretrained CNN, described in "Convolutional neural network" section, was fine tuned on the training data. The training conditions were kept the same and the test accuracy achieved was 99.11%.

In the final experiment (Fig. 13) the classification result for feature extraction from our proposed augmentation scheme was compared with the several other methods, surpassing them with the fine tuned CNN. Song and Yan's approach (Song and Yan 2013) introduces and uses the Adjacent Evaluation Completed Local Binary Pattern (AECLBP) feature descriptor for the SVM. Feature extraction through Scattered Convolutional Network (SCN) (Song et al. 2014), and Overfeat Network (Sermanet et al. 2014) for classification are also compared. Our approach simply utilizes the features extracted from CNN. The impact of an information-rich training set generated from our GAN based augmentation scheme is that those features are enough for a higher classification performance. So, our approach has the added advantage of not requiring any feature-engineering. That is why a simple CNN trained on synthetically augmented data outperforms more complex CNN based feature extractors trained with classic augmentation.

Conclusion

The quality of the product plays an important role in determining the effectiveness of a manufacturing operation. In the manufacturing industry, inspection not only plays an important role in quality control, but it also works as a feedback system for the process. Hence, defect classification is as important as defect detection.

In this study, GANs are trained to augment the data for surface defects on hot-rolled steel strips. The framework has been developed in order to investigate the use of synthetic augmentation for classification of manufacturing datasets. Thus, a synthetic data augmentation scheme is developed and compared with classic data augmentation over the course of five experiments. Initially, the training set was enlarged to a size of 9000 images using graphical transformation methods such as flipping, rotation, and scaling. The validation accuracy of the CNN model was found out

Actual/Predicted	Cr	In	Pa	PS	RS	Sc	Sensitivity	Specificity
Cr	300	0	0	0	0	0	100%	99.60%
In	1	296	3	0	0	0	98.67%	99.60%
Pa	1	3	295	0	1	0	98.33%	99.80%
PS	0	3	0	297	0	0	99.00%	100%
RS	1	0	0	0	299	0	99.67%	99.93%
Sc	3	0	0	0	0	297	99.00%	100%

Fig. 13 Confusion Matrix for Experiment-5

to be 90.28%. In the second experiment, the CNN is trained using DCGAN-augmented data having an identical size and under the same training parameters. The models predicted the defect classes with an accuracy of 95.78%, which was a significant improvement over the model proposed in the first experiment. In the third experiment, DC-GAN is substituted with AC-GAN, keeping the remainder of the data augmentation pipeline identical. The accuracy of this model was found out to be 92.78%. The accuracy of AC-GAN was observed to be less than that of DC-GAN by 3%. In the fourth experiment, InfoGAN is employed, keeping the remainder of the data augmentation pipeline identical and the accuracy of this model was found to be 94.86%, less than that of DC-GAN by 0.92%.

Overall, for hot-rolled steel strips, an improvement in the performance of 5.5% is achieved using DCGAN-based augmentation over Classic augmentation. The observed experimental results show the efficacy and robustness of the proposed model and can be used as a benchmark. In our fifth experiment, using a pretrained CNN for transfer learning, state of the art results on NEU CLS were achieved obtaining a performance gain of 3.33% over the baseline.

The main finding of this work is that the proposed method shows great potential for being adopted as a standard augmentation technique for deep learning for industrial applications, which generally have small-sized datasets. The suggested GAN-based data augmentation procedure offers a great advantage by significantly cutting down the time taken for, and the opportunity cost of, collection of real-world data. Furthermore, our pipeline provides another benefit—it may be used to train successful models for other applications (such as for structural defects using X-ray images) through domain adaptation, reducing the time even further. It is also easy to run (once trained) and thus can be physically employed at a manufacturing plant using cameras and single-board computers.

This study can be extended to other models. As the classification is performed using a basic CNN, any development on the base architecture will also be reflected in the final model. This has been demonstrated using transfer learning. Similarly, more advanced GAN models may also positively influence the augmentation process.

Future research could scrutinize the segmentation problem for manufacturing defect datasets. This is an inspiring problem as it is required to generate both synthetic surface defect images and their ground truth masks. The applied methodology may be extended for the purpose to improve industry performance in different sectors.

References

Antoniou, A., Storkey, A., & Edwards, H. (2018). Augmenting image classifiers using data augmentation generative adversarial networks. In *Artificial neural networks and machine learning—ICANN 2018*.

Arjovsky, M., Chintala, S., & Bottou, L. (2017). *Wasserstein GAN*.

Badmos, O., Kopp, A., Bernthaler, T., & Schneider, G. (2020). Image-based defect detection in lithium-ion battery electrode using convolutional neural networks. *Journal of Intelligent Manufacturing, 31,* 885–897.

Berthelot, D., Schumm, T., & Metz, L. (2017). *BEGAN: Boundary equilibrium generative adversarial networks*.

Bousmalis, K., Silberman, N., Dohan, D., Erhan, D., & Krishnan, D. (2017). Unsupervised pixel-level domain adaptation with generative adversarial networks. In *IEEE conference on computer vision and pattern recognition (CVPR)*.

Carreira-Perpiñán, M. Á., & Hinton, G. E. (2005). On contrastive divergence learning. In *AISTATS*.

Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., & Abbeel, P. (2016). Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *NIPS*.

Davtalab, O., Kazemian, A., Yuan, X., & Khoshnevis, B. (2020). Automated inspection in robotic additive manufacturing using deep learning for layer deformation detection. *Journal of Intelligent Manufacturing*.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, Miami, FL, USA.

Devadas, C., Samarasekera, I. V., & Hawbolt, E. B. (1991). The thermal and metallurgical state of steel strip during hot rolling: Part III. Microstructural evolution. *Metallurgical Transactions A, 22*(2), 335–349.

Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., et al. (2015). Flownet: Learning optical flow with convolutional networks. In *ICCV*.

Feng, S., Zhou, H., & Dong, H. (2019). Using deep neural network with small dataset to predict material defects. *Materials and Design, 162,* 300–310.

Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). Synthetic data augmentation using GAN for improved

liver lesion classification. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Washington, DC.
Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative Adversarial Nets. In *Advances in neural information processing systems 27*.
Grzenda, M., & Bustillo, A. (2019). Semi-supervised roughness prediction with partly unlabeled vibration data streams. *Journal of Intelligent Manufacturing, 30,* 933–945.
Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). *Improved training of Wasserstein GANs*.
Hao, R., Lu, B., Cheng, Y., Li, X., & Huang, B. (2020). A steel surface defect inspection approach towards smart industrial monitoring. *Journal of Intelligent Manufacturing*.
Hinton, G. E., Dayan, P., & Frey, B. J. (1995). The "Wake-Sleep" Algorithm for Unsupervised Neural Networks. *Science, 268,* 1158–1161.
Hjelm, R. D., Jacob, A. P., Che, T., Trischler, A., Cho, K., & Bengio, Y. (2018). Boundary-seeking generative adversarial networks. In *ICLR*.
Huang, Y., Qiu, C., Guo, Y., Wang, X., & Yuan, K. (2018). Surface defect saliency of magnetic tile. In *IEEE international conference on automation and engineering*, At Munich, Germany.
Izadi, S., Mirikharaji, Z., Kawahara, J., & Hamarneh, G. (2018). Generative adversarial networks to segment skin lesions. In *IEEE 15th international symposium on biomedical imaging*, Washington, DC, USA.
Kingma, D. P., & Welling, M. (2014). Auto-Encoding Variational Bayes. In *ICLR*.
Lai, Y. T. K., Hu, J. S., Tsai, Y. H., & Chiu, W. Y. (2018). Industrial anomaly detection and one-class classification using generative adversarial networks. In *IEEE/ASME international conference on advanced intelligent mechatronics (AIM)*.
Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *IEEE conference on computer vision and pattern recognition (CVPR)*.
Li, J., Liang, X., Wei, Y., Xu, T., Feng, J., & Yan, S. (2017). Perceptual generative adversarial networks for small object detection. In *IEEE conference on computer vision and pattern recognition (CVPR)*, Honolulu, HI, USA.
Luo, J., Huang, J., & Li, H. (2020). A case study of conditional deep convolutional generative adversarial networks in machine fault diagnosis. *Journal of Intelligent Manufacturing*.
Madani, A., Moradi, M., Karargyris, A., & Syeda-Mahmood, T. (2018). Chest x-ray generation and data augmentation for cardiovascular abnormality classification. In *Medical imaging 2018: Image processing, vol. 10574*.
Mao, X., Li, Q., Xie, H., Lau, R. Y., Wang, Z., & Smolley, S. P. (2017). Least squares generative adversarial networks. In *IEEE international conference on computer vision,* Venice, Italy.
Mirza, M., & Osindero, S. (2014). *Conditional generative adversarial nets*.
Moeskops, P., Veta, M., Lafarge, M. W., Eppenhof, K. A., & Pluim, J. P. (2017). Adversarial training and dilated convolutions for brain MRI segmentation. In *Workshop on deep learning in medical image analysis*.
, V., Mohammed Safwan, K. P., Chennamsetty, S. S., & Krishnamurthi, G. (2017). Generative adversarial networks for brain lesion detection. In *SPIE medical imaging*, Orlando, Florida, United States.
Odena, A., Olah, C., & Shlens, J. (2017). *Conditional image synthesis with auxiliary classifier GANs*.
Pan, J., Canton, C., McGuinness, K., O'Connor, N., Torres, J., Sayrol, E., et al. (2017). SalGAN: Visual saliency prediction with adversarial networks. *Computer Vision and Image Understanding*.
Radford, A., Metz, L., & Chintala, S. (2015). *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*.
Richter, S.R., Vineet, V., Roth, S., & Koltun, V. (2016). Playing for data: Ground truth from computer games. In *ECCV*.
Scime, L., & Beuth, J. (2018). Anomaly detection and classification in a laser powder bed additive manufacturing process using a trained computer vision algorithm. *Additive Manufacturing, 19,* 114–126.
Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. (2014). Overfeat: Integrated recognition, localization and detection using convolutional networks. In *International conference on learning representations*, Banff.
Shrivastava, A., Pfister, T., Tuzel, O., Susskind, J., Wang, W., & Webb, R. (2017). Learning from simulated and unsupervised images through adversarial training. In *CVPR*.
Song, K., Hu, S., & Yan, Y. (2014). Automatic recognition of surface defects on hot-rolled. *Journal of Computational Information Systems, 10*(7), 3049–3055.
Song, K., & Yan, Y. (2013). A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. *Applied Surface Science, 285,* 858–864.
Song, K., & Yan, Y. (2019). NEU surface defect database. Northeastern University. http://faculty.neu.edu.cn/yunhyan/NEU_surface_defect_database.html. [Accessed 5 4 2019].
Sun, T., Tien, F., Tien, F., Tien, F. C., & Kuo, R. J. (2016). Automated thermal fuse inspection using machine vision and artificial neural networks. *Journal of Intelligent Manufacturing, 27,* 639–651.
Tabernik, D., Šela, S., Skvarč, J., & Skočaj, D. (2020). Segmentation-based deep-learning approach for surface-defect detection. *Journal of Intelligent Manufacturing, 31,* 759–776.
Tian, Y. (2017, April 16). Master Chinese calligraphy with conditional adversarial networks. https://kaonashi-tyc.github.io/2017/04/06/zi2zi.html. [Accessed 10 5 2019].
Wolterink, J. M., Leiner, T., Viergever, M. A., & Išgum, I. (2017). Generative Adversarial Networks for Noise Reduction in Low-Dose CT. *IEEE Transactions on Medical Imaging, 36*(12), 2536–2545.
Yang, D., Xu, D., Zhou, S. K., Georgescu, B., Chen, M., Grbic, S., et al. (2017) Automatic liver segmentation using an adversarial image-to-image network. In *International conference on medical image computing and computer-assisted intervention*, Quebec City, QC, Canada.
Yu, H., Tieu, K., Lu, C., Deng, G., & Liu, X. (2013). Occurrence of surface defects on strips during hot rolling. *International Journal of Advanced Manufacturing Technology, 67,* 1161–1170.
Zhai, W., Zhu, J., Cao, Y., & Wang, Z. (2018). A generative adversarial network based framework for unsupervised visual surface inspection. In *IEEE international conference on acoustics, speech and signal processing (ICASSP)*, Calgary, AB, Canada.
Zhang, Z., Song, Y., & Qi, H. (2017). Age progression/regression by conditional adversarial autoencoder. In *IEEE conference on computer vision and pattern recognition (CVPR)*.
Zhao, J., Mathieu, M., & LeCun, Y. (2017). Energy-based generative adversarial networks. In *ICLR*.
Zhu, X., Liu, Y., Li, J., Wan, T., & Qin, Z. (2018). Emotion classification with data augmentation using generative adversarial networks. In *Advances in knowledge discovery and data mining. PAKDD 2018*, 2017.

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

Kragujevac Journal of Mathematics
Volume 47(3) (2023), Pages 431–443.

THE FAMILY OF SZÁSZ-DURRMEYER TYPE OPERATORS INVOLVING CHARLIER POLYNOMIALS

NAOKANT DEO[1] AND RAM PRATAP[2]

Abstract. In this paper, we consider Szász-Durrmeyer type operators based on Charlier polynomials associated with Srivastava-Gupta operators [17]. For the considered operators, we discuss error of estimation by using first and second order modulus of continuity, Lipchtiz-type space, Ditzian-Totik modulus of smoothness, Voronovskaya type asymptotic formula and weighted modulus of continuity.

1. Introduction

For the Charlier polynomials [8], the generating functions are as follows:

$$e^u\left(1-\frac{u}{a}\right)^t=\sum_{j=0}^{\infty}C_j^{[a]}(t)\frac{u^j}{j!}, \tag{1.1}$$

where $C_j^{[a]}(t)=\sum\limits_{r=0}^{j}\binom{j}{r}(-t)_r\frac{1}{a^r}$ and $(j)_0=1$, $(j)_i=j(j+1)(j+2)\cdots(j+i-1)$ for $i\geq 1$.

Suppose $\gamma>0$, the space $C_\gamma[0,\infty):=\{g\in C[0,\infty):|g(t)|\leq Me^{\gamma t}\}$ for some $M>0$.

In view of Charlier polynomials, Varma and Tasdelen [19] proposed a sequence of linear positive operators for $g\in C_\gamma[0,\infty)$ as follows:

$$L_n(g;x,a)=e^{-1}\left(1-\frac{1}{a}\right)^{(a-1)nx}\sum_{j=0}^{\infty}\frac{C_j^{[a]}(-(a-1)nx)}{j!}g\left(\frac{j}{n}\right), \tag{1.2}$$

Key words and phrases. Charlier polynomials, Srivastava-Gupta operators, modulus of continuity, Ditzian-Totik modulus of smoothness, weighted modulus of continuity.

2010 *Mathematics Subject Classification.* Primary: 41A25, 41A36

DOI

Received: November 03, 2019.

Accepted: September 10, 2020.

where $a > 1$ and $x \in [0, \infty)$. For sufficiently large a, if we replace x by $x - \frac{1}{n}$ the above operators reduce to well-known Szász-Mirakyan operators [18].

In [17], Srivastava and Gupta introduced a new family of linear positive operators as follows:

$$G_n^c(g;x) = (n-c)\sum_{j=0}^{\infty} p_{n,j}(x;c) \int_0^{\infty} p_{n+c,j-1}(u;c)du + p_{n,0}(x;c)g(0), \tag{1.3}$$

where $p_{n,j}(x;c) = \frac{(-x)^j}{j!}\phi_{n,c}^{(j)}(x)$ and

$$\phi_{n,c}(x) = \begin{cases} e^{-nx}, & c = 0, \\ (1+cx)^{\frac{-n}{c}}, & c = 1, 2, 3, \ldots \end{cases}$$

For the operators (1.3), they also studied the rate of convergence for the functions of bounded variation. Ispir and Yüksel [10] defined the Bézier varient of the Srivastava-Gupta operators and discussed rate of convergence for the functions of bounded variation. Srivastava-Gupta [17] contains several well-known operators for different values of c. Many authors have proposed various forms and modifications of the above operators and studied several local and global approximation results. For more (see [1, 3, 7, 12, 14, 16, 20, 21]).

Motivated from the above stated work, we define a linear positive operators for $g \in C_B[0, \infty)$ as follows:

$$\begin{aligned} G_{n,c}^{[a]}(g;x) =& (n-c)e^{-1}\Big(1-\frac{1}{a}\Big)^{(a-1)nx}\Bigg[\sum_{j=1}^{\infty}\frac{C_j^{[a]}(-(a-1)nx)}{j!}\int_0^{\infty} p_{n+c,j-1}(u;c)g(u)du \\ &+ C_0^{[a]}g(0)\Bigg]. \end{aligned} \tag{1.4}$$

In above operators, it can easily be seen that if we take $c = 1$, we obtain Szász-Durrmeyer type operators involving Charlier Polynomials which were proposed by Kajla and Agrawal [11] and studied several approximation results like Vorovskaya type asymptotic theorem, local approximation, statistical rate of convergence and functions of bounded variation. For more articles based on Charlier polynomials (see [2, 4]).

The main purpose of this article is to define the operators (1.4) and discuss the approximation results using the first and second order modulus of continuity, Lipschitz-type space, Ditzian-Totik modulus of smoothness, Voronovskaya-type formula and weighted approximation.

2. Auxiliary Results

Lemma 2.1 ([19]). *For the operators $L_n(\cdot; x, a)$, we have*

(i) $L_n(1; x, a) = 1$;
(ii) $L_n(u; x, a) = x + \frac{1}{n}$;

(iii) $L_n(u^2;x,a) = x^2 + \frac{x}{n}\left(3 + \frac{1}{a-1}\right) + \frac{2}{n^2}$;

(iv) $L_n(u^3;x,a) = x^3 + \frac{x^2}{n}\left(6 + \frac{3}{a-1}\right) + \frac{x}{n^2}\left(5 + \frac{3}{a-1} + \frac{1}{(a-1)^2}\right) + \frac{5}{n^3}$;

(v)

$$\begin{aligned} L_n(u^4;x,a) =& x^4 + \frac{x^3}{n}\left(10 + \frac{6}{a-1}\right) + \frac{x^2}{n^2}\left(31 + \frac{30}{a-1} + \frac{11}{(a-1)^2}\right) \\ &+ \frac{x}{n^3}\left(37 + \frac{31}{a-1} + \frac{20}{(a-1)^2} + \frac{6}{(a-1)^3}\right) + \frac{15}{n^4}. \end{aligned}$$

Lemma 2.2. *The moments of the operators $G_{n,c}^{[a]}(u^i;x)$, $i = 0,1,2,3,4$, are as follows:*

(i) $G_{n,c}^{[a]}(1;x) = 1$;

(ii) $G_{n,c}^{[a]}(u;x) = \frac{1}{(n-2c)}(nx+2)$;

(iii) $G_{n,c}^{[a]}(u^2;x) = \frac{1}{(n-2c)(n-3c)}\left(n^2x^2 + n\left(6 + \frac{1}{a-1}\right)x + 7\right)$;

(iv)

$$\begin{aligned} G_{n,c}^{[a]}(u^3;x) =& \frac{1}{(n-2c)(n-3c)(n-4c)}\left(n^3\,x^3 + 3n^2\left(4 + \frac{1}{a-1}\right)x^2\right. \\ &\left. + n\left(28 + \frac{12}{a-1} + \frac{2}{(a-1)^2}\right)x + 34\right); \end{aligned}$$

(v)

$$\begin{aligned} G_{n,c}^{[a]}(u^4;x) =& \frac{1}{(n-2c)(n-3c)(n-4c)(n-5c)}\left(n^4x^4 + 2n^3\left(10 + \frac{3}{a-1}\right)x^3\right. \\ &+ n^2\left(126 + \frac{60}{a-1} + \frac{11}{(a-1)^2}\right)x^2 \\ &\left. + n\left(292 + \frac{126}{a-1} + \frac{40}{(a-1)^2} + \frac{6}{(a-1)^3}\right)x + 209\right). \end{aligned}$$

Lemma 2.3. *The central moments for the defined operators:*

(i) $G_{n,c}^{[a]}(u-x;x) = \frac{2}{(n-2c)}(1+cx)$;

(ii) $G_{n,c}^{[a]}((u-x)^2;x) = \frac{1}{(n-2c)(n-3c)}\left(c(n+6c)x^2 + \left(n\left(2 + \frac{1}{a-1}\right) + 12c\right)x + 7\right)$;

(iii)

$$\begin{aligned} G_{n,c}^{[a]}((u-x)^4;x) =& \frac{1}{(n-2c)(n-3c)(n-4c)(n-5c)}\left((3n^2 + 86cn + 126c^2)c^2x^4\right. \\ &+ \frac{2c(3(2a-1)n^2 + 4c(43a-28)n + 240(a-1)c^2)}{(a-1)}x^3 \\ &+ \frac{(56a^2 - 100a + 47)n^2 + 2c(91a^2 - 62a - 9)n + 840(a-1)^2c^2}{(a-1)^2}x^2 \\ &\left. + \frac{2(78a^3 - 171a^2 + 128a - 32)n + 680c(a-1)^3}{(a-1)^3}x + 209\right). \end{aligned}$$

Lemma 2.4. *For sufficiently large n, we have*

(i) $\lim_{n\to\infty} nG_{n,c}^{[a]}((u-x);x) = 2(1+cx)$;

(ii) $\lim_{n\to\infty} nG_{n,c}^{[a]}((u-x)^2;x) = x\left(cx+\frac{1}{a-1}+2\right)$;

(iii) $\lim_{n\to\infty} n^2G_{n,c}^{[a]}((u-x)^4;x) = x^2\left(3c^2x^2+\frac{3(2a-1)}{(a-1)}x+\frac{56a^2-100a+47}{(a-1)^2}\right)$.

3. Main Result

Theorem 3.1. *Let $g \in C_\gamma[0,\infty)$ and for sufficiently large n the operators $G_{n,c}^{[a]}(g(u);x)$ converges to $g(x)$ uniformly in each compact subset of $[0,\infty)$.*

Proof. From Lemma 2.2, $\lim_{n\to\infty} G_{n,c}^{[a]}(1;x) = 1$, $\lim_{n\to\infty} G_{n,c}^{[a]}(u;x) = x$ and $\lim_{n\to\infty} G_{n,c}^{[a]}(u^2;x) = x^2$. Then by Bohman-Korovokin theorem, $G_{n,c}^{[a]}(g(u);x)$ converges to $g(x)$ uniformly in each compact subset of $[0,\infty)$. □

Theorem 3.2. *For $g \in C_\gamma[0,\infty)$ and $g'(x)$, $g''(x)$ exist in $[0,\infty)$, we have*

$$\left[G_{n,c}^{[a]}(g(u);x)-g(x)\right] = 2(1+cx)g'(x)+\frac{x}{2!}\left(cx+\frac{1}{a-1}+2\right)g''(x).$$

Proof. From Taylor's expansion, we have

$$g(u) = g(x)+(u-x)g'(x)+\frac{(u-x)^2g''(x)}{2!}+r(u,x)(u-x)^2,$$

where $r(u,x)$ converges to 0 when $u \to x$.

Applying $G_{n,c}^{[a]}(\cdot;x)$ in above expression, we have

$$\begin{aligned} n\left[G_{n,c}^{[a]}(g(u);x)-g(x)\right] =& nG_{n,c}^{[a]}((u-x);x)g'(x)+\frac{nG_{n,c}^{[a]}((u-x)^2;x)g''(x)}{2!} \\ &+ nG_{n,c}^{[a]}(r(u,x)(u-x)^2;x). \end{aligned} \tag{3.1}$$

Using Cauchy-Schwarz inequality and Lemma 2.4 in last term of above equality, we obtain

$$\lim_{n\to\infty} nG_{n,c}^{[a]}(r(u,x)(u-x)^2;x) = 0. \tag{3.2}$$

From (3.1), using (3.2) and Lemma 2.4, we get the required result. □

Let $C_B[0,\infty)$ be the space of real valued continuous and bounded functions g on $[0,\infty)$, provided with norm

$$\|g\| = \sup_{x\in[0,\infty)} |g(x)|,$$

and Peetre's K-functional for $g \in C_B[0,\infty)$ is given as:

$$K_2(g;\delta) = \inf_{x\in W_\infty^2}\left\{\|g-h\|+\delta\|h''\|\right\}, \quad \delta > 0,$$

where $W_\infty^2[0,\infty) = \{h \in C_B[0,\infty) : h', h'' \in C_B[0,\infty)\}$. Devore and Lorentz [5, Theorem 2.4, page 177], provided relation between Peetre's K functional and second order modulus of continuity as follows:

$$K_2(g;\delta) \leq C\omega_2(g;\sqrt{\delta}), \tag{3.3}$$

and the second order modulus of continuity $\omega_2(g;\sqrt{\delta})$ is given as

$$\omega_2(g;\sqrt{\delta}) = \sup_{0<i\leq\delta} \sup_{x\in[0,\infty)} \left|g(x+2i) - 2g(x+i) + g(x)\right|.$$

The usual modulus of continuity $\omega(g;\delta)$ for $g \in C_B[0,\infty)$

$$\omega(g;\delta) = \sup_{0<i\leq\delta} \sup_{x\in[0,\infty)} \left|g(x+i) - g(x)\right|.$$

Theorem 3.3. *For $g \in C_B[0,\infty)$ and $a > 1$, we have*

$$\left|G_{n,c}^{[a]}(g(u);x) - g(x)\right| \leq C\omega_2\left(g\sqrt{\delta_{n,c}^a(x)}\right) + \omega\left(g;\left|\frac{2(1+cx)}{(n-2c)}\right|\right),$$

where C is positive constant and $\delta_{n,c}^a(x) = \left[G_{n,c}^{[a]}((u-x)^2;x) + \frac{2(1+cx)^2}{(n-2c)^2}\right]$.

Proof. We consider an auxiliary operators:

$$\tilde{G}_{n,c}^{[a]}(g(u);x) = G_{n,c}^{[a]}(g(u);x) - g\left(x + \frac{2(1+cx)}{n-2c}\right) + g(x).$$

The Taylor's expansion for the function $h \in W_\infty^2[0,\infty)$ is given as

$$h(u) = h(x) + (u-x)h'(x) + \int_x^u (u-x)h''(u)du.$$

Applying $\tilde{G}_{n,c}^{[a]}(\cdot;x)$ in above expression

$$\tilde{G}_{n,c}^{[a]}(h(u);x) - h(x) = \tilde{G}_{n,c}^{[a]}((u-x);x)h'(x) + \tilde{G}_{n,c}^{[a]}\left(\int_x^u (u-x)h''(u)du;x\right).$$

Since $\tilde{G}_{n,c}^{[a]}(1;x) = 1$, $\tilde{G}_{n,c}^{[a]}(u;x) = x$ and $\tilde{G}_{n,c}^{[a]}(u-x;x) = 0$, we get

$$\begin{aligned}
\left|\tilde{G}_{n,c}^{[a]}(h(u);x) - h(x)\right| &= \left|\tilde{G}_{n,c}^{[a]}\left(\int_x^u (u-x)h''(u)du;x\right)\right| \\
&\leq \left|G_{n,c}^{[a]}\left(\int_x^u (u-x)h''(u)du;x\right)\right| \\
&\quad + \left|\int_x^{x+\frac{2(1+cx)}{n-2c}} \left(x + \frac{2(1+cx)}{n-2c} - u\right)h''du\right|
\end{aligned}$$

$$\leq \left[G_{n,c}^{[a]}((u-x)^2;x) + \frac{2(1+cx)^2}{(n-2c)^2} \right] \|h''\|$$

$$\leq \delta_{n,c}^{a}(x) \|h''\| . \tag{3.4}$$

Using auxiliary operators we can write

$$\begin{aligned} \left| G_{n,c}^{[a]}(g(u);x) - g(x) \right| \leq & \left| \tilde{G}_{n,c}^{[a]}(g-h;x) - (g-h)(x) \right| + \left| \tilde{G}_{n,c}^{[a]}(h(u);x) - h(x) \right| \\ & + \left| g\left(x + \frac{2(1+cx)}{n-2c} \right) - g(x) \right| \\ \leq & 2\left| g-h \right| + \delta_{n,c}^{a}(x) \|h''\| + \omega\left(g; \frac{2(1+cx)}{n-2c} \right). \end{aligned}$$

Taking infimum on the right hand side of the above inequality for $g \in W_{\infty}^{2}[0,\infty)$, we have

$$\left| G_{n,c}^{[a]}(g(u);x) - g(x) \right| = 2K_2\left(g; \delta_{n,c}^{a}(x) \right) + \omega\left(g, \left| \frac{2(1+cx)}{n-2c} \right| \right).$$

From (3.3), we obtain

$$\left| G_{n,c}^{[a]}(g(u);x) - g(x) \right| = C\omega_2\left(g; \sqrt{\delta_{n,c}^{a}(x)} \right) + \omega\left(g, \left| \frac{2(1+cx)}{n-2c} \right| \right).$$

Hence, the proof. □

In the next theorem, we estimate global rate of convergence by using Ditzian-Totik modulus of smoothness $\omega_{\phi^\alpha}(g;\delta)$ for $g \in C_B[0,\infty)$, $0 < \alpha \leq 1$ and $\phi(x) = \sqrt{x(1+cx)}$ which is defined as:

$$\omega_{\phi^\alpha}(g;\delta) = \sup_{0\leq s\leq \delta} \sup_{x \pm \frac{s\phi^\alpha(x)}{2} \in [0,\infty)} \left| g\left(x + \frac{s\phi^\alpha(x)}{2} \right) - g\left(x - \frac{s\phi^\alpha(x)}{2} \right) \right|,$$

and the Peetre K-functional is defined as:

$$K_{\phi^\alpha}(g;\delta) = \inf_{g\in W_\alpha} \left\{ \|g-h\| - \delta \|\phi^\alpha g'\| \right\},$$

where W_α is subspaces of those functions which are locally absolutely continuous on $g \in [0,\infty)$ with the normed $\|\phi^\alpha g'\| \leq \infty$. In [6], there exists a constant $C > 0$ such that

$$C^{-1}\omega_{\phi^\alpha}(g;\delta) \leq K_{\phi^\alpha}(g;\delta) \leq C\omega_{\phi^\alpha}(g;\delta).$$

Theorem 3.4. *Suppose $g \in C_B[0,\infty)$ and for sufficiently large n, we have*

$$\left| G_{n,c}^{[a]}(g;x) - g(x) \right| \leq C\omega_{\phi^\alpha}\left(g; \frac{\phi^{1-\alpha}(x)}{\sqrt{n}} \right).$$

Proof. For $h \in W_\alpha$, we have

$$h(u) = h(x) + \int_x^u h'(t)dt.$$

Applying $G_{n,c}^{[a]}(\cdot;x)$ in the above equality and using Hölder's inequality, we obtain

$$\begin{aligned}\left|G_{n,c}^{[a]}(h;x)-h(x)\right| \leq & G_{n,c}^{[a]}\left(\int_x^u h'(t)dt;x\right)\\ & \leq \|\phi^\alpha h'\|\, G_{n,c}^{[a]}\left(\int_x^u \frac{dt}{\phi^\alpha(t)};x\right)\\ & \leq \|\phi^\alpha h'\|\, G_{n,c}^{[a]}\left(|u-x|^{1-\alpha}\left|\int_x^u \frac{dt}{\phi(t)}\right|^\alpha;x\right). \qquad (3.5)\end{aligned}$$

Let $p(u,x)=\left|\int_x^u \frac{dt}{\phi(t)}\right|$, we have

$$\begin{aligned}p(u,x) \leq & \left|\int_x^u \frac{dt}{\phi(t)}\right|\left(\frac{1}{\sqrt{1+cx}}+\frac{1}{\sqrt{1+cu}}\right)\\ \leq & \frac{2\,|u-x|}{\sqrt{x}+\sqrt{u}}\left(\frac{1}{\sqrt{1+cx}}+\frac{1}{\sqrt{1+cu}}\right)\\ \leq & \frac{2\,|u-x|}{\sqrt{x}}\left(\frac{1}{\sqrt{1+cx}}+\frac{1}{\sqrt{1+cu}}\right).\end{aligned}$$

Since $|a+b|^\alpha \leq |a|^\alpha+|b|^\alpha$, $0\leq\alpha\leq 1$, and from the above inequality, we obtain

$$\left|\int_x^u \frac{dt}{\phi(t)}\right|^\alpha \leq \frac{2^\alpha|u-x|^\alpha}{x^{\frac{\alpha}{2}}}\left(\frac{1}{(1+cx)^{\frac{\alpha}{2}}}+\frac{1}{(1+cu)^{\frac{\alpha}{2}}}\right). \qquad (3.6)$$

From (3.5), (3.6) and using Cauchy-Schwarz inequality, we get

$$\begin{aligned}\left|G_{n,c}^{[a]}(h;x)-h(x)\right| \leq & \frac{2^\alpha\,\|\phi^\alpha h'\|}{x^{\frac{\alpha}{2}}}G_{n,c}^{[a]}\left(|u-x|\left(\frac{1}{(1+cx)^{\frac{\alpha}{2}}}+\frac{1}{(1+cu)^{\frac{\alpha}{2}}}\right);x\right)\\ \leq & \frac{2^\alpha\,\|\phi^\alpha h'\|}{x^{\frac{\alpha}{2}}}\left(\frac{1}{(1+cx)^{\frac{\alpha}{2}}}\left(G_{n,c}^{[a]}((u-x)^2;x)\right)^{\frac{1}{2}}\right.\\ & \left.+\left(G_{n,c}^{[a]}((u-x)^2;x)\right)^{\frac{1}{2}}\times\left(G_{n,c}^{[a]}((1+cu)^{-\alpha};x)\right)^{\frac{1}{2}}\right). \qquad (3.7)\end{aligned}$$

From Theorem 3.1, $G_{n,c}^{[a]}((1+cu)^{-\alpha}$ converges to $(1+cx)^{-\alpha}$ for sufficiently large n. Thus, for $\epsilon>0$, there exist $n_0\in N$ such that $G_{n,c}^{[a]}((1+cu)^{-\alpha};x)\leq(1+cx)^{-\alpha}+\epsilon$ for all $n\geq n_0$.

Choosing $\varepsilon=(1+cx)^{-\alpha}$, we get

$$G_{n,c}^{[a]}((1+cu)^{-\alpha};x)\leq 2(1+cx)^{-\alpha}, \quad \text{for all } n\geq n_0. \qquad (3.8)$$

For sufficiently large n there exists a constant $C>0$, such that

$$G_{n,c}^{[a]}((u-x)^2;x)\leq C\frac{\phi^2(x)}{n}. \qquad (3.9)$$

From (3.7) to (3.9), we obtain

$$\left|G_{n,c}^{[a]}(h;x)-h(x)\right| \leq 2^{\alpha+1}C\left\|\phi^{\alpha}h'\right\|\frac{\phi^{1-\alpha}(x)}{\sqrt{n}}. \tag{3.10}$$

We can write

$$\begin{aligned}\left|G_{n,c}^{[a]}(g;x)-g(x)\right| \leq & \left|G_{n,c}^{[a]}(g-h;x)\right|+\left|G_{n,c}^{[a]}(h;x)-h(x)\right|+|h(x)-g(x)| \\ \leq & 2\|g-h\|+\left|G_{n,c}^{[a]}(h;x)-h(x)\right|.\end{aligned}$$

From (3.10), we get

$$\begin{aligned}\left|G_{n,c}^{[a]}(g;x)-g(x)\right| \leq & 2\|g-h\|+2^{\alpha+1}C\left\|\phi^{\alpha}h'\right\|\frac{\phi^{1-\alpha}(x)}{\sqrt{n}} \\ \leq & C\left\{\|g-h\|+\frac{\phi^{1-\alpha}(x)}{\sqrt{n}}\left\|\phi^{\alpha}h'\right\|\right\} \\ \leq & CK_{\phi}^{\alpha}\left(g;\frac{\phi^{1-\alpha}(x)}{\sqrt{n}}\right) \\ \leq & C{\omega_{\phi}}^{\alpha}\left(g;\frac{\phi^{1-\alpha}(x)}{\sqrt{n}}\right).\end{aligned}$$

Hence, the proof. □

In [15], the Lipschitz-type space for positive real numbers β_1, β_2 is defined as:

$$Lip_M^{\beta_1,\beta_2}(\lambda)=\left\{g\in C_{\mathbf{B}}[0,\infty):|g(u)-g(x)|\leq M_g\frac{|u-x|^{\lambda}}{(u+\beta_1x^2+\beta_2x)^{\frac{\lambda}{2}}};x,u\in[0,\infty)\right\},$$

where $M_g>0$ and $0<\lambda\leq 1$.

Theorem 3.5. *Let $g\in Lip_M^{\beta_1,\beta_2}(\lambda)$ and $0<\lambda\leq 1$, then for $x\geq 0$ we have*

$$\left|G_{n,c}^{[a]}(g;x)-g(x)\right|\leq M_g\left(\frac{\mu_{n,2}^{a,c}(x)}{(\beta_1x^2+\beta_2x)}\right)^{\frac{\lambda}{2}},$$

where $\mu_{n,2}^{a,c}(x)=G_{n,c}^{[a]}((u-x)^2;x)$.

Proof. First, we discuss the result for $\lambda=1$. For $g\in Lip_M^{\beta_1,\beta_2}(\lambda)$, we have

$$\begin{aligned}\left|G_{n,c}^{[a]}(g;x)-g(x)\right| \leq & (n-c)e^{-1}\left(1-\frac{1}{a}\right)^{(a-1)nx}\sum_{j=0}^{\infty}\frac{C_j^{[a]}(-(a-1)nx)}{j!} \\ & \times\int_0^{\infty}p_{n,j}(u;c)\,|g(u)-g(x)|\,du \\ \leq & (n-c)e^{-1}\left(1-\frac{1}{a}\right)^{(a-1)nx}\sum_{j=0}^{\infty}\frac{C_j^{[a]}(-(a-1)nx)}{j!}\end{aligned}$$

$$\times \int_0^\infty p_{n,j}(u;c) M_g \frac{|u-x|}{(u+\beta_1 x^2+\beta_2 x)^{\frac{1}{2}}} du.$$

Since $\frac{1}{\sqrt{u+\beta_1 x^2+\beta_2 x}} < \frac{1}{\sqrt{\beta_1 x^2+\beta_2 x}}$, applying Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
\left|G_{n,c}^{[a]}(g;x)-g(x)\right| \leq & \frac{M_g}{\sqrt{\beta_1 x^2+\beta_2 x}}(n-c)e^{-1}\left(1-\frac{1}{a}\right)^{(a-1)nx}\sum_{j=0}^{\infty}\frac{C_j^{[a]}(-(a-1)nx)}{j!} \\
& \times \int_0^\infty p_{n,j}(x;c)\,|u-x|\,du \\
\leq & \frac{M_g}{\sqrt{\beta_1 x^2+\beta_2 x}}\sqrt{G_{n,c}^{[a]}((u-x)^2;x)} \\
\leq & M_g\sqrt{\frac{\mu_{n,2}^{a,c}(x)}{\beta_1 x^2+\beta_2 x}}.
\end{aligned}$$

The result is true for $\lambda = 1$. Now we prove for $0 < \lambda < 1$. Using Hölder's inequality with $p = \frac{2}{\lambda}$ and $q = \frac{2}{2-\lambda}$, we have

$$\begin{aligned}
& \left|G_{n,c}^{[a]}(g;x)-g(x)\right| \\
\leq & \left\{n-c)e^{-1}\left(1-\frac{1}{a}\right)^{(a-1)nx}\sum_{j=0}^{\infty}\frac{C_j^{[a]}(-(a-1)nx)}{j!}\right. \\
\leq & M_g\left\{(n-c)e^{-1}\left(1-\frac{1}{a}\right)^{(a-1)nx}\sum_{j=0}^{\infty}\frac{C_j^{[a]}(-(a-1)nx)}{j!}\right. \\
& \left.\times \int_0^\infty p_{n,j}(x;c)\frac{(u-x)^2}{(u+\beta_1 x^2+\beta_2 x)}du\right\}^{\frac{\lambda}{2}} \\
\leq & \frac{M_g}{(\beta_1 x^2+\beta_2 x)^{\frac{\lambda}{2}}}\left\{(n-c)e^{-1}\left(1-\frac{1}{a}\right)^{(a-1)nx}\sum_{j=0}^{\infty}\frac{C_j^{[a]}(-(a-1)nx)}{j!}\right. \\
& \left.\times \int_0^\infty p_{n,j}(x;c)(u-x)^2 du\right\}^{\frac{\lambda}{2}} \\
\leq & M_g\left(\frac{\mu_{n,2}^{a,c}(x)}{(\beta_1 x^2+\beta_2 x)}\right)^{\frac{\lambda}{2}}.
\end{aligned}$$

Hence, the proof. □

Theorem 3.6. *If $g(x)$ is continuously differentiable function on $[0,\infty)$ and $|g'(x)| \leq D$ for some $D > 0$, then we have*

$$\left|G_{n,c}^{[a]}(g;x) - g(x)\right| \leq D\left|\frac{2(1+cx)}{n-2c}\right| + 2\sqrt{\mu_{n,2}^{a,c}(x)}\omega_b\left(g';\sqrt{\mu_{n,2}^{a,c}(x)}\right),$$

where $\omega_b(g;\delta)$, $\delta > 0$, is usual modulus of continuity on $[0,b]$ and

$$\mu_{n,2}^{a,c}(x) = G_{n,c}^{[a]}((u-x)^2;x).$$

Proof. From Lagrange's mean value theorem, we get

$$g(u) - g(x) = (u-x)g'(\eta) = (u-x)g'(x) + (u-x)(g'(\eta) - g'(x)),$$

where η lies between x and u.

now, we apply $G_{n,c}^{[a]}(\cdot;x)$ on both side of the above equation. Since $x < \eta < u$ we have $|\eta - x| < |u-x|$ and

$$\left|G_{n,c}^{[a]}(g;x) - g(x)\right| \leq |g'(x)|\left|G_{n,c}^{[a]}((u-x);x)\right| + \omega_b(g';\delta)\left(|u-x| + \frac{(u-x)^2}{\delta}\right).$$

Applying Cauchy-Schwarz inequality, we obtain

$$\left|G_{n,c}^{[a]}(g;x) - g(x)\right| \leq D\left|\frac{2(1+cx)}{n-2c}\right| + \sqrt{\mu_{n,2}^{a,c}(x)}\omega_b\left(g';\delta\right)\left(1 + \frac{\sqrt{\mu_{n,2}^{a,c}(x)}}{\delta}\right).$$

Taking $\delta = \sqrt{\mu_{n,2}^{a,c}(x)}$, we get required result. □

In our next theorem, we study the rate of convergence for the operators (1.4) based on Lipscitz maximal function of order r given by Lenze [13] as

$$\varpi_r(g;x) = \sup_{u\neq x,\, x,u\in[0,\infty)} \frac{|g(u) - g(x)|}{|u-x|^r}, \tag{3.11}$$

where $0 < r \leq 1$.

Theorem 3.7. *For $g \in C_B[0,\infty)$, we have*

$$\left|G_{n,c}^{[a]}(g;x) - g(x)\right| \leq \varpi_r(g;x)\left(\mu_{n,2}^{a,c}(x)\right)^r.$$

Proof. From (3.11), we obtain

$$\left|G_{n,c}^{[a]}(g;x) - g(x)\right| \leq \varpi_r(g;x)G_{n,c}^{[a]}\left(|u-x|^r;x\right).$$

Using Hölder's inequality with $p = \frac{2}{r}$ and $q = \frac{2}{2-r}$, we obtain

$$\left|G_{n,c}^{[a]}(g;x) - g(x)\right| \leq \varpi_r(g;x)\left(G_{n,c}^{[a]}\left((u-x)^2;x\right)\right)^r \leq \varpi_r(g;x)\left(\mu_{n,2}^{a,c}(x)\right)^r.$$

Hence, the proof. □

Let $C_2[0,\infty)$ be the space of all continuous functions on $[0,\infty)$ and defined as:

$$C_2[0,\infty) := \left\{ g : |g| \leq M_g(1+x^2) \right\},$$

where M_g is positive constant which may depends on g with the norm

$$\|g\|_2 = \sup_{x>0} \frac{|g(x)|}{1+x^2}.$$

Let $C_2^*[0,\infty) := \left\{ g \in C_2[0,\infty) : \lim_{x\to\infty} \frac{g(x)}{1+x^2} \text{ exists and finite} \right\}$. The weighted modulus of continuity [9] $\Omega(g;\delta)$ is given as

$$\Omega(g;\delta) = \sup_{0\leq\beta<\delta} \frac{|g(x+\beta)-g(x)|}{(1+\beta^2)(1+x^2)}.$$

Lemma 3.1. *For every $g \in C_2^*[0,\infty)$, $\Omega(g;\delta)$ has the properties:*

(i) *$\Omega(g;\delta)$ is a monotonically increasing function of δ;*
(ii) *$\lim_{\delta\to 0^+} \Omega(g;\delta) = 0$;*
(iii) *$\Omega(g;k\delta) \leq 2(1+k)(1+\delta^2)\Omega(g;\delta)$, $k>0$ and $\delta>0$.*

Theorem 3.8. *For $g \in C_2^*[0,\infty)$, we have*

$$\sup_{x\in[0,\infty)} \frac{\left|G_{n,c}^{[a]}(g;x)-g(x)\right|}{(1+x^2)^{\frac{5}{2}}} \leq C\Omega\left(g;\frac{1}{\sqrt{n}}\right),$$

where C is positive constant depends on a and c.

Proof. For $x,u \in [0,\infty)$ and from (3.11), we can write

$$\begin{aligned}
|g(u)-g(x)| \leq& (1+(u-x)^2)(1+x^2)\Omega\left(g;\frac{|u-x|\,\delta}{\delta}\right)\\
\leq& 2(1+\delta^2)(1+x^2)\left(1+\frac{|u-x|}{\delta}\right)(1+(u-x)^2)\Omega(g;\delta).
\end{aligned}$$

Applying $G_{n,c}^{[a]}(\cdot;x)$ in the above inequality, we have

$$\begin{aligned}
\left|G_{n,c}^{[a]}(g;x)-g(x)\right| \leq& 2(1+\delta^2)(1+x^2)\Omega(g;\delta)G_{n,c}^{[a]}\left(\left(1+\frac{|u-x|}{\delta}\right)(1+(u-x)^2);x\right)\\
\leq& 2(1+\delta^2)(1+x^2)\Omega(g;\delta)\left\{G_{n,c}^{[a]}(1;x)+G_{n,c}^{[a]}((u-x)^2;x)\right.\\
&+\frac{1}{\delta}G_{n,c}^{[a]}\left(|u-x|;x\right)+\frac{1}{\delta}G_{n,c}^{[a]}\left(|u-x|\,(u-x)^2;x\right)\Big\}\\
\leq& 2(1+\delta^2)(1+x^2)\Omega(g;\delta)\left\{1+G_{n,c}^{[a]}((u-x)^2;x)\right.\\
&+\frac{1}{\delta}\Big(G_{n,c}^{[a]}\left((u-x)^2;x\right)\Big)^{\frac{1}{2}}\\
&+\frac{1}{\delta}\Big(G_{n,c}^{[a]}\left((u-x)^2;x\right)\Big)^{\frac{1}{2}}\Big(G_{n,c}^{[a]}\left((u-x)^4;x\right)\Big)^{\frac{1}{2}}\Big\}. \qquad (3.12)
\end{aligned}$$

From Lemma 2.3, we have

$$G_{n,c}^{[a]}((u-x)^2;x) \leq \frac{C_1(1+x^2)}{n}$$

and

$$G_{n,c}^{[a]}((u-x)^4;x) \leq \frac{C_2(1+x^2)^2}{n^2},$$

where C_1 and C_2 are positive constants depend on a and c.

Using the above inequality in (3.12) and taking $\delta = \frac{1}{\sqrt{n}}$, we get

$$\begin{aligned}\left|G_{n,c}^{[a]}(g;x) - g(x)\right| \leq & 2\left(1+\frac{1}{n}\right)\Omega\left(g;\frac{1}{\sqrt{n}}\right)(1+x^2)\left\{1+C_1\,(1+x^2)\right. \\ & \left. +\sqrt{C_1(1+x^2)}+\sqrt{C_1C_2}(1+x^2)^{\frac{3}{2}}\right\}.\end{aligned}$$

Taking $C = 4\left(1+C_1+\sqrt{C_1}+\sqrt{C_1C_2}\right)$, we obtain the result. □

Acknowledgements. The authors are thankful to the Delhi Technological University, Delhi, India, for providing financial support to complete this research work. They are also indebted to the anonymous reviewers and the handling editor for their comments and suggestions to leads the improvements in the original manuscript.

References

[1] A. M. Acu and C. Muraru, *Certain approximation properties of Srivastava-Gupta operators*, J. Math. Inequal. **12**(2) (2018), 583–595. https://dx.doi.org/10.7153/jmi-2018-12-44

[2] R. Chauhan, B. Baxhaku and P. N. Agrawal, *Szász type operators involving Charlier polynomials of blending type*, Complex Anal. Oper. Theory **13** (2019), 1197–1226. https://doi.org/10.1007/s11785-018-0854-x

[3] N. Deo, *Faster rate of convergence on Srivastava-Gupta operators*, Appl. Math. Comput. **218**(21) (2012), 10486–10491. https://doi.org/10.1016/j.amc.2012.04.012

[4] N. Deo and M. Dhamija, *Charlier-Szász-Durrmeyer type linear positive operators*, Afr. Mat. **19**(1) (2018), 223–232, https://doi.org/10.1007/s13370-017-0537-1

[5] R. A. Devore and G. G. Lorentz, *Constructive Approximation*, Springer-Verlag, Berlin, Heidelberg, 1993.

[6] Z. Ditzian and V. Totik, *Moduli of Smoothness*, Springer series in Computational Mathematics **9**, Springer, New York, 1987.

[7] V. Gupta and H. M. Srivastava, *A General family of the Srivastava-Gupta operators preserving linear functions*, Eur. J. Pure Appl. Math. **11**(3) (2018), 575–579, https://doi.org/10.29020/nybg.ejpam.v11i3.3314

[8] M. E. H. Ismail, *Classical and Quantum Orthogonal Polynomials in One Variable*, Cambridge University Press, Cambridge, New York, 2005.

[9] N. Ispir, *On modified Baskakov operators on weighted spaces*, Turkish J. Math. **26**(3) (2001), 355–365.

[10] N. Ispir and I. Yüksel, *On the Bézier variant of Srivastava-Gupta operators*, Appl. Math. E-Notes **5** (2005), 129–137.

[11] A. Kajla and P. N. Agrawal, *Szász-Durrmeyer type operators based on Charlier polynomials*, Appl. Math. Comput. **268** (2015), 1001–1014, https://doi.org/10.1016/j.amc.2015.06.126

[12] A. Kumar, V. N. Mishra and D. Tapiawala, *Stancu type generalization of modified Srivastava-Gupta operators*, Eur. J. Pure Appl. Math. **10**(3) (2017), 890–897.
[13] B. Lenze, *On Lipschitz maximal functions and their smoothness spaces*, Indag. Math. **50**(1) (1988), 53–63. https://doi.org/10.1016/1385-7258(88)90007-8
[14] P. Maheswari (Sharma), *On modified Srivastava-Gupta operators*, Filomat **29**(6) (2015), 1173–1177. https://doi.org/10.2307/24898197
[15] M. A. Özarslan and O. Duman, *Local approximation properties for certain king type operators*, Filomat **27**(1) (2013), 173–181. https://doi.org/10.2307/24896345
[16] R. Pratap and N. Deo, *Approximation by Genuine Gupta-Srivastava Operators* Rev. R. Acad. Cienc. Exactas Fis. Nat. Ser. A Mat. RACSAM **113** (2019), 2495–2505. https://doi.org/10.1007/s13398-019-00633-4
[17] H. M. Srivastava and V. Gupta, *A certain family of summation-integral type operators*, Math. Comput. Modelling **37** (2003), 1307–1315. https://doi.org/10.1016/S0895-7177(03)90042-2
[18] O. Szász, *Generalisation of S. Bernstein polynomials to the infinite interval*, Journal of Research of the National Bureau of Standards **56** (1950), 239–245.
[19] S. Varma and F. Tasdelen, *Szász type operators involving Charlier polynomials*, Math. Comput. Modelling **56** (2012), 118–122.
[20] D. K. Verma and P. N. Agrawal, *Convergence in simultaneous approximation for Srivastava-Gupta operators*, Math. Sci. (Springer) **6** (2012), Article ID 22, 8 pages. https://doi.org/10.1186/2251-7456-6-22
[21] R. Yadav, *Approximation by modified Srivastava-Gupta operators*, Appl. Math. Comput. **226** (2014), 61–66. https://doi.org/10.1016/j.amc.2013.10.039

[1,2]Department of Applied Mathematics,
Delhi Technological University,
Delhi-110042
Email address: naokantdeo@dce.ac.in
Email address: rampratapiitr@gmail.com

International Journal of Management (IJM)
Volume 11, Issue 11, November 2020, pp. 1489-1499. Article ID: IJM_11_11_142
Available online at http://www.iaeme.com/ijm/issues.asp?JType=IJM&VType=11&IType=11
Journal Impact Factor (2020): 10.1471 (Calculated by GISI) www.jifactor.com
ISSN Print: 0976-6502 and ISSN Online: 0976-6510
DOI: 10.34218/IJM.11.11.2020.142

© **IAEME** Publication **Scopus** Indexed

THE STATUS OF WOMEN PARTICIPATION IN ENGINEERING AND TECHNOLOGY PROFESSION AND THEIR CHALLENGES IN WORKING INSTITUTIONS

Addissie Melak
Delhi Technological University, PhD Scholar in Economics,
Department of Humanities, Delhi, India
Debre Tabor University, Department of Economics, Ethiopia

Seema Singh
Delhi Technological University, Professor of Economics,
Department of Humanities, Delhi, India

ABSTRACT

The purpose of this study is to make a discussion about the trends of women involvement in engineering & technology profession along with evidence from Ethiopia and challenges of women working in this professions. Data has been collected on the ground of the variable; (1) women enrollment in engineering & technology undergraduate and postgraduate degree programs, and (2) the share of graduated women in engineering & technology from first degree to Ph.D level. Primary data has also gathered to investigate the challenges of women working in engineering and technology professions. The result of the study shows there is under- representation of women in engineering & technology enrollment in all degree programs. In the regular undergraduate degree program of engineering & technology, the average percentage of women from the total enrollment is 23.34% which is below one fourth of the total. Moreover, the average percentage of women enrollment out of the total regular post graduate degree is 13.4%. The average number of women graduates in regular and evening from first to third degree programs of the total graduate is only 14.5%. The result also indicates 35% of respondents suffering various challenges while they are working in engineering and technology professions. The finding implies low participation of women in engineering & technology education which needs great emphasis by the prospective body of parents, students themselves and government bodies to improve the role of women in the route of economic development.

Key words: Women, Engineering & Technology Professions, Ethiopia.

Cite this Article: Addissie Melak and Seema Singh, The Status of Women Participation in Engineering and Technology Profession and their Challenges in Working Institutions, *International Journal of Management,* 11(11), 2020, pp 1489-1499.
http://www.iaeme.com/IJM/issues.asp?JType=IJM&VType=11&IType=11

1. INTRODUCTION

Studies by (Budapest, Hungary, 2013; Academy, 2016) confirms that engineering and technology plays a central role in economic development. Engineers can design and build infrastructure, such as transport, bridges, dams, communication, waste management, water supply and sanitation as well as energy and digital infrastructure which can raise productivity and enhance other economic variables. By having a well-developed transport and communications infrastructure, countries can be better able to get goods and services to market and move workers to jobs. A strong communications network allows a rapid and free flow of information, and helps to ensure businesses easily communicate and make timely decisions.

Unfortunately, there is a big imbalance in the distribution between the two genders in Africa as it happens in other parts of the world. The women engineers are few and unable to fully participate in national development. Among the problems are high isolation of women engineers in the profession and the absence of many prominent women who can serve as role models. In order to overcome this, the UNESCO regional office for science and technology in Africa embarked on a pilot project to prepare a directory of female engineers in a few African countries including Ethiopia (UNESCO, 1995). In Ethiopia, the average enrolment of women in engineering and technology in regular undergraduate degree is 23.34% since the academic year 2007 to 2016.

Studies by (UNESCO, 1995; Catherine Hill et al., 2010; Dimitriadi, 2013; C. García Villa, G, 2014) also confirms that, the number of women in engineering and technology is near to the ground, even though studies by (**S.** Barnard et al., 2012; Singh .S, 2013) indicates, there is some progress in the numbers of women entering engineering and technology programmes, as a result of general trends towards increasing numbers of women entering higher education as a whole. However, men continue to outnumber women, especially at the upper levels of these professions. Despite significant improvements, women are still under-represented in the science and technology academic sector. This is due to a variety of reasons, mostly related to the role allocated to women in modern society as well as pre-existing biases and women's perceptions of commitment to engineering and technology careers.

1.1. Challenges of women working in engineering and technology professions

Although a number of national and international policies and programs have been implemented to promote and increase the inclusion of gender equity in the field of science and technology there is a significant gender gap about the participation of women at higher levels of science academic careers. According to Sharma et al., (2019), Minimum wage, progression, security, unconscious biases, stereotype, societal, cultural and institutional issues are barriers to women reaching the higher levels of science and technology professions. Rosser (2004) found the most considerable challenges that women engineering and technology professionals facing are balancing carriers and their family. Bastalich et al. (2007) also founds from the interviewed women engineers and stated that, cultural stereotypes about women as under-confident or unsuited to engineering work starkly is the major challenge.

On the basis of data MOE (2016), the status of women employment participation in engineering and technology as academic teaching in government universities and other sectors

in Ethiopia is dismal. The average number of employed women in engineering and technology professions as teaching staff in Ethiopian universities during 2007 to 2016 is 7.6%. This confirms the finding of (Jote, 2017; Addissie M. and Seema. S. 2020) that there are substantial variations observed between men and women, and the employment status of women in this profession is consistently lower in Ethiopia.

1.2. Objective of the Study

Based on the above backgrounds, the focus of this study is to investigate:

- the status of women participation in engineering and technology education on the bases of available data during 2007 to 2016,
- the challenges of women working in engineering and technology professions in Ethiopia.

2. MATERIALS AND METHODS

2.1. Study area

This study is focused on Ethiopia, which is found in eastern part of Africa. Ethiopian education system is structured by different stages, which is kindergarten, alternative basic education, first cycle primary education (grade 1-4), second cycle primary level from grade 5-8, first cycle secondary level (grade 9 and 10), second cycle secondary education level (preparatory school grade 11 and 12), technical and vocational education and training (TVET) undergraduate degree, Master degree and Ph.D level. Higher education institutions in Ethiopia have been started in 50 years ago (Asalf et al., 2020) and providing education as undergraduate degree, Master's degree and Ph.D degree in regular, extension, summer and distance educational programs.

2.2. Data and method of analysis

In this study, secondary data has collected from ten successive academic years of 2007 to 2016. Data has been taken from the ministry of education statistics annual abstract of Ethiopia to analyze the trends and share of women involvement in engineering & technology education. Data has collected based on variables such as the share of women enrollment and graduation. The primary data also has been conducted in Ethiopia, Amhara region which is one of among nine states in a country to investigate challenges of women working in engineering & technology professions. Method of gathering information has been conducted through some steps. First we have selected one state via convenience sampling. The state has ten public higher institutions categorized into four-generation on the basis of year of establishment. Then we select three universities from each generation: These are Bahir Dar University from 1st generation, Wollo University from 2nd generation, and Debre Tabor University from 3rd generation. Data not collected from 4th generation universities since they are new and started working in the 2017/2018 academic year, so that they may not have sufficient data for this study. The selection of universities from each generation has been done purposely. Because universities established at different times can have different experiences, but the choice of universities within generations is random. Then from these universities we have received the full address of total 1840 women graduates in engineering and technology professions in these three universities during 2016 and 2017 from the registrar office of universities since we have legal letters from concerned organizations. Then we have interviewed 244 sample respondents through telephone through random sampling. Since our objective is investigating challenges of women working in the engineering and technology profession, our questionnaire was concerned only employed sample respondents. Hence, the total respondents are 244 women respondents working in the STEM profession out of 1840

total women graduates during 2016 and 2017 in these three sample universities. These collected data have been organized and analyzed with the help of descriptive statistical methods. These process of data collection framework has been presented in the following diagram.

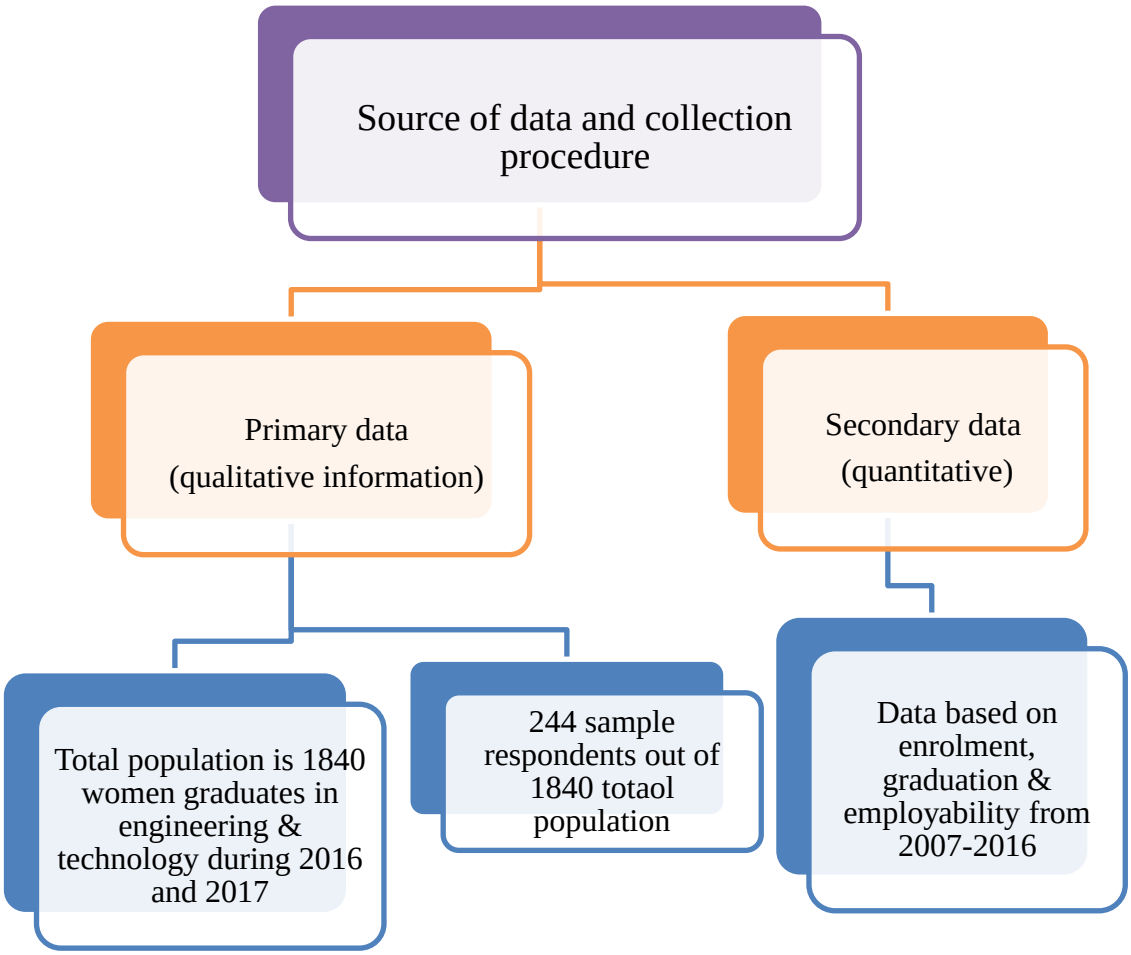


Figure 1

3. RESULTS AND DISCUSSION

3.1. Quantitative results

Table 1 shows, women enrolment in engineering and technology in regular undergraduate educational programs is improving based on data from the academic year 2007 to 2016. In 2007, women enrollment in engineering and technology was 18% out of the total enrolled students of 13070. In the year 2008, there was better enrolment than in 2007. Out of the total enrolled 9570 students, women accounted for almost 21% in 2008. However, this number failed again in 2009 which is 18.16%. This happened in 2010 also. But starting from 2011, the percentage of women enrolling in engineering and technology fields of study is increasing from 22.24% to 28.4%. In 2014, their enrollment was 31% which is better from 2007 to 2016. But in 2015 and 2016 their enrolment declined to 28%. In general, when we see the average enrollment of women for ten consecutive academic years, the figure shows there is significant underrepresentation in the area. This is due to various social, cultural and economic factors. Besides, women undermine themselves as they cannot perform science subjects.

Table 1 Trends of women students' enrolment in engineering & technology education in regular under graduate program in government universities

Year	Man	Women	Total	%women from the Total
2007	10676	2394	13070	18.087%
2008	7561	2009	9570	20.99%
2009	25437	5647	31084	18.1669%
2010	43958	9679	53637	18.045%
2011	68276	19523	87799	22.236%
2012	91742	30145	121887	24.7%
2013	91192	29794	120986	24.23%
2014	118078	52881	170959	30.93%
2015	127837	50025	177862	28.125%
2016	125171	49657	174828	28.4%

Source: MOE, 2016

As a statistical summary of table 1, that on average, 25175 women are enrolled within the last ten years, while 70993 men are enrolled. This indicates women enrollment in engineering & technology in regular under graduate is one third of man enrolment. The maximum number of women was enrolled in 2014, which is 52881, whereas the minimum has happened in 2008 which was only 2009 women students. In the same statistical summary, the maximum number of man enrolment is 127837 which was recorded in 2015, and the minimum is 7561 students which was seen in 2008 academic year of study. The average difference between man and women enrollment in engineering and technology regular undergraduate degree within ten years is 45817. Similarly, the enrollment variation of them within the year is increasing from year to year. This again implies there is a non-proportional gender balance in engineering & technology education regular undergraduate programs.

Table 2 women enrolment in engineering & technology education in evening under graduate program in government universities

Year	Man	Women	Total	%women from the total
2007	4157	551	4708	11.7%
2008	1297	375	1672	22.42%
2009	5740	717	6457	11.1%
2010	6145	905	7050	12.83%
2011	9945	1987	11932	16.65%
2012	12797	2725	15522	17.6%
2013	13186	3032	16218	18.7%
2014	22606	8467	31073	27.25%
2015	18211	4500	22711	19.8%
2016	16917	4478	21395	20.93%
Average women% from the total = 17.898%				

Source: MOE, 2016

In Ethiopia, evening undergraduate educational programs are provided in the time outside formal working hours, and mostly at the weekend. Hence, evening or extension educational programs are important, which civil servants and others can improve their educational level side by side from their daily formal and informal job. The problem is women have low participation. Table 2 represents women enrollment in engineering and technology education in the evening undergraduate programs in government universities. In 2007, the percentage of women enrollment was 11.7% from the total of 4708 students. In 2008, it was increasing to 22.4% but immediately it declined to 11.1% in 2009. After 2010 women enrollment has increased slightly, and reached 27.25% out of the total of 31073 students in 2014, which is the

year that better participation has been seen. However, it failed down again to 20% during 2015 and 2016. The average women enrolment engineering and technology in evening undergraduate program for the last ten years is around 18%, which shows their under-representation in this area of study.

Table 3 Trends of women enrollment in engineering & technology education in summer and distance undergraduate program in government universities

Year	Man	Women	Total	%women from the total
2007	903	58	961	6.305%
2008	272	29	301	9.634%
2009	1524	98	1622	6.04%
2010	1745	288	2033	14.166%
2011	2847	359	3206	11.198%
2012	2867	674	3541	19.034%
2013	2867	674	3541	19.034%
2014	7388	2476	9864	25.10%
2015	10598	828	13605	6.085%
2016	11476	3157	14633	21.57%

Source MOE, 2016

Summer and distance undergraduate educational program is an essential method which is given at the time of summer. In this program, mostly teachers are involved to improve their educational level, since teachers are free during summer for at least two months on the basis of the low number of workers in Ethiopia. Here, the problem is that the quality of education will be under question mark. Because one semester course has to be completed within two months only. In table 3, women enrolment in engineering and technology in summer and distance undergraduate educational programs in government universities are presented. As shown in table 3, women enrolment in this program was 58(6.3%) from a total of 961 students in 2007. Similarly in 2008, only 29(9.634%) women were enrolled out of the total 301 students. Women enrolment has increased to 14.17% in 2010 but it failed again to 11% in 2011. In 2012 and 2013, it increased to 19% and reached to 25.1% in 2014. This is the year where a better record has shown. However, women enrolment declined to 6% (828) in 2015 out of the total of 13605 students, and again it increased to 21.5% in 2016. The average women enrolment in this program is 11.9% from 2007 to 2016. Generally in this program, there has been an up and down trend in the last ten years.

Table 4 Trends of women enrollment in engineering & technology education in regular postgraduate programme (M.Sc. & Ph.D) in government universities

Year	Man	Women	Total	% women from the total
2007	590	22	613	3.5%
2008	596	39	635	6.14%
2009	1328	69	1397	4.93%
2010	1075	79	1154	6.84%
2011	2555	441	2996	14.7%
2012	2516	447	2963	15%
2013	2516	447	2963	15%
2014	2938	1945	4883	39.8%
2015	4327	674	5001	13.5%
2016	5635	988	6623	14.9%
Average women% 13.431%				

Source MOE, 2016

Table 4 represents trends of women enrollment in engineering and technology education in regular post graduate programs of Master and Ph.D degree in government universities. In 2007 there were only 22 women enrolled which is 3.5% out of the total 613 students. It slightly increased in 2008, which is 6.14% but declined to 4.93% from the total of 1397 enrolled students in 2009. Again it increased to 6.84% in 2010 and to 14% in 2011. There was the same percentage of women enrolment during 2012 and 2013 which is 15% and it goes up dramatically to 39.8% in 2014 that is considered as the best improvement has been seen. However, it immediately failed to 13.5% in 2015, and slightly goes up to 14.9% in 2016. The average women enrollment in engineering and technology education in masters and Ph.D degree level is 13.4% during the last ten years. To sum up, women enrolment in Masters and Ph.D degree is very low proportion as we have seen in the table, and has an up and down trend. In fact in Ethiopia, Masters and Ph.D education programs are not expanded in domestic universities due to shortage of supervisors. The government takes action to solve this problem that students are sent to Asian and European countries to study their post graduate degree. In some cases, Ethiopia government also imports professors from other countries to provide these programs inside the country. However, even in this process also the number of women participation is understated. The average enrollment of women is 515 which is one fourth (1/4) of man's average enrollment of 2408. The maximum number of women enrollment has occurred in 2014, which accounts 1945 while the minimum is only 29 women students. However, man enrolment is better than women, their maximum number accounts 5635 students, and the minimum number is 590. This indicates there serious under-representation of women in engineering & technology education in all degree levels.

Table 5 Trends of women enrollment in engineering & technology education in evening and summer postgraduate programme (M.Sc. / Ph.D) in government universities

Year	Man	Women	Total	%women
2010	247	21	268	7.84%
2011	338	25	363	6.887%
2012	993	114	1107	10.298%
2013	993	124	1117	11.1%
2014	1106	338	1444	23.407%
2015	897	1199	2096	57.2%
2016	2575	388	2963	13.09%
Average women percentage 18.5%				

Women enrollment in engineering and technology education in evening and summer post graduate programs in Government University was started in 2010. In table 5, the number of women enrolled in engineering and technology is only 21(7.84%) out of 268 students in 2010, even if this failed to 6.88% in 2011. It has shown an increasing trend in 2012, 2013, and 2014, which is 10.3%, 11.1% and 23.4% respectively. In 2015 it reached a peak level and rose to 57.2% which is a very high proportion and beyond 50%. However, this improvement unexpectedly declined to 13%. The average women enrolment in this program is 18.5 percent which is very low as it happens in other programs as we have seen above.

Table 6 Trends of women graduates in engineering and technology education in regular and evening undergraduate and post graduate program in government universities

Year	Under graduate		MSc & Ph.D		Total		Total	%women
	Man	Women	Man	Women	Man	Women		
2007	2686	401	482	43	3168	444	3612	12.29%
2008	1704	267	144	14	1848	281	2129	13.2%
2009	2664	477	570	58	3234	535	3769	14.1%
2010	2463	379	450	57	2913	436	3349	13%
2011	3123	615	565	39	3668	654	4322	15.13%
2012	7157	1547	572	27	7729	1574	9303	16.9%
2013	7157	1547	572	27	7729	1574	9303	16.9%
Average women percentage 14.5%								

Source: MOE, 2016

Table 6 shows women graduates in engineering and technology education in regular and evening undergraduate and postgraduate programs in government universities. As we have seen, what seems in the proportion of women enrolment situations, the same thing is happening in the number of women graduates in the same field of study. In 2007, the number of women graduates in engineering and technology was 12% which is 444 from the total graduates of 3612 students. There is a very low proportion in the graduation year of 2007 to 2013. The average number of women graduates from the year 2007 to 2013 is 14.5%. We couldn't get data on women graduates in engineering and technology after 2014 separately.



Figure 1 Graduated women and man in under graduate and post graduate program in engineering & technology field of study

Source: MOE, 2016

Women and man graduates in undergraduate and postgraduate degree programs in engineering & technology education have presented in figure 1 which is constructed on the bases of data provided in table 6. As the figure easily illustrates, there is a big imbalance between women and man graduates in the field, for instance, when the maximum of graduated men is 7729, only 1574 women are graduated in this area of study. This also contributes to their low level of participation in employment in engineering and technology which is a pillar economic development. This again may lead to low levels of women participation in income generation and reeducating of poverty especially, underdeveloped countries like Ethiopia.

3.2. Qualitative results

Table 7 Statistical Summary of challenges of women working in engineering and technology professions

Challenges of Women Working in STEM Professions	Number of Respondents	Percentage (%)
Sometimes I have problem of conflict with costumers, staff members and leaders.	19	7.79%
There was problem of communication with costumers, students and staff members at the beginning of joining my working organization.	8	3.28%
There is discrimination, undermining and discouraging on the bases of gender, race and religion	18	7.38%
Problem of accessibility of infrastructure like water sanitation, electricity and materials and problem of adaptation of whether condition when we change residence place of work.	6	2.46%
Work load and conflict with leaders	10	4.10%
Lack of good governance of managers and shortage of materials	6	2.46%
Scarcity of capital, working place, and market accessibility to run own business and others	19	7.79%
I didn't suffer any challenges in my work place	158	64.75%

Source: Authors survey data, 2018

The results from table 7 shows that out of 244 total respondents 35.25 % have mentioned major problems and challenges that women in engineering and technology professionals are suffering. Such obstacles' are

- Problem of conflict with customers, staff members and leaders,
- Problem of communication with costumers and students, discrimination, undermining and discouraging on the bases of gender, race and religion,
- problem of accessibility of infrastructure like water sanitation and electricity
- Problem of work load and conflict with leaders, lack of good governance of managers.
- Shortage of materials, scarcity of capital, working place, and market accessibility to run their own business and others.
- Minimum and unequal wage, progression, security, unconscious biases, stereotype, and societal, cultural and institutional issues.

These findings are in line with the previous studies done by Botella, C. et al., (2019) entitled as "Gender Diversity in STEM Disciplines: A Multiple Factor Problem" founds that 42% professional women worker thought as there was a lack of female role models, 39% experienced gender biasing the workplace, 36% pointed out that they had unequal growth opportunities compared to men, and 35% indicated that there was a gender pay gap for the same skills. However, fortunately, the result of this current study indicates, the majority of women professionals in engineering and technology have better working environments in Ethiopia.

4. CONCLUSION

Despite significant improvements, women are still under-represented in science and technology. This is due to a variety of reasons, mostly related to the role allocated to women in modern society as well as pre-existing biases and women's perceptions of commitment to engineering and technology careers. The purpose of this study is to make a discussion about the trends of women involvement in engineering & technology profession along with evidence from Ethiopia. The study has also made an attempt to discuss the challenges of women working in engineering and technology professions in Ethiopia.

In this study, ten year secondary data was collected from 2007 to 2016 in Ethiopian Ministry of Education Statistics Annual Abstract. Data has collected on the ground of the variable; (1) women enrollment in engineering & technology in regular, evening, summer and distance undergraduate degree program, (2) women enrollment in engineering & technology in regular, evening and summer post graduate degree program in government higher education institution and (3) the share of graduated women in engineering & technology from first degree to Ph.D level. Primary information has also been collected from 244 from women who work in engineering and technology professions. This data has been computed by the help of a method of descriptive analysis.

The result of the study shows that there is under- representation of women in engineering & technology enrollment in all degree programs. In the regular undergraduate degree program of engineering & technology, the average percentage of women from the total enrollment is 23.34% which is below one fourth of the total. Similarly, their participation is very small in evening, summer and distance programs. Moreover, the average percentage share of women enrolment out of the total regular post graduate degree is 13.4% and the same phenomena has been observed in other programs. The average number of women graduates in regular and evening from first to third degree programs of the total graduate is only 14.5%. Besides the low proportion of participation of women in engineering and technology professions, they have many challenges in the working place of these professions as discussed in the result section of this study.

The finding implies, low participation of women in engineering & technology education and employment needs great emphasis by the prospective body of parents, women students and professionals, government bodies and other stakeholders to improve the role of women in the route of economic development. Such as women labourers must try to solve conflict with others through negotiation. The government must work on: creating awareness about gender equality for the concerned body, establishing good governance of managers, must have rules and regulation to reduce discrimination and violation, must enhance basic infrastructure like electricity, water, road, etc., by using endogenous knowledge and materials, must have financial credit availability, and must have strong security of citizens.

REFERENCES

[1] Abebayehu Tora (2013) Assessment of Sexual Violence against Female Students in Wolaita Sodo University, Southern Ethiopia, Journal of Intertamporal violence vol.8, issue 11

[2] Academy, R. (2016). Engineering and economic growth : a global view. (September), 1–80.

[3] Addissie Melak, S. S. (2020) Participation of Women in Engineering & Technology Education and Employment. International Journal of Management and Humanities, vol.4, issue 7, 26-32.

[4] Asalf Habtegeorgis, G.C Maheshwari and Rajan Yadav, Adoption of Quality management: The Case of Ethiopian Universities, *International JournalofManagement*,11(7),2020,pp.1610-1621. http://www.iaeme.com/IJM/issues.asp?JType=IJM&VType=11&IType=7

[5] Bastalich,W., Franzway, S., Gill, J., Mills, J., & Sharp, R. (2007). Disrupting masculinities: Women engineers and engineering workplace culture. *Australian Feminist Studies*, *22*(54), 385–400.

[6] Botella, C., Rueda, S., López-Iñesta, E., & Marzal, P. (2019). Gender diversity in STEM disciplines: A multiple factor problem. *Entropy*, *21*(1), 30.

[7] Catherine Hill et al. (2010) Women in Science, Technology, and Engineering, and Mathematics, page xiv, Library of Congress Control Number: 2010901076 ISBN: 978-1-879922-40- 2 Published by AAUW Washington, DC 20036 Web: www.aauw.org

[8] Carmen García Villa & Elsa M. González y González (2014) Women students in engineering in Mexico: exploring responses to gender differences, International Journal of Qualitative Studies in Education, 27:8, 1044 1061, DOI: 10.1080/09518398.2014.924636

[9] Dimitriadi (2013) Young women in science and technology: the importance of choice, Journal of Innovation and Entrepreneurship vol. 2:No.5 pp1-14:http://www.innovation-entrepreneurship.com/content/2/1/5

[10] IEEE AFRICON (2015), IEEE Women in Engineering (WiE) Workshop, Addis Ababa Institute of technology University, Ethiopia.

[11] Jote T. (2017) Exploring employment status and education–job match among engineering graduates in Ethiopia: Policy implications. *International Journal of African Higher Education* 4(1).

[12] MOE (2016) Fedral Democratic Republic of Ethiopia, ministry of education, educational statistics annual abstract report. URL http: //www.moe.gov.et/statistics.

[13] Robsan Margo Egne (2014) Gender Equality in Public Higher Education Institutions of Ethiopia: The Case of Science, Technology, Engineering, and Mathematics, *Discourse and Communication for Sustainable Education, vol.5, pp. 3ñ21, 2014:* DOI: 10.2478/dcse-2014 0001

[14] Rosser, S. V. (2004). Using powre to advance: Institutional barriers identified by women scientists and engineers. *NWSA journal*, 50–78.

[15] S. Barnard, T. Hassan, B. Bagilhole & A. Dainty (2012) 'They're not girly girls': an exploration of quantitative and qualitative data on engineering and gender in higher education, European Journal of Engineering Education, vol.37:No.2, pp193-204, DOI: 10.1080/03043797.2012.661702

[16] Singh.S (2013) Women in Engineering Education in India http://www.youblisher.com/p/109258-Women-in-Engineering-Education-in-India/

[17] Sharma, J., Yarlagadda, T., Sharma, S., & Yarlagadda, P. K. (2019). Vertical segregation: issues and challenges of women engineers in Australia. *Procedia Manufacturing*, 30, 671–676.

[18] UNISCO (1995) directory of female engineers in four African countries, Ethiopia, Ghana, Nigeria and Tanzania

Portuguese Economic Journal
https://doi.org/10.1007/s10258-020-00190-4

ORIGINAL ARTICLE

Time varying dynamics of globalization effect in India

Shikha Gupta[1] · Nand Kumar[1]

Received: 4 August 2020 / Accepted: 18 November 2020/Published online: 26 November 2020
© ISEG – Instituto Superior de Economia e Gestão 2020

Abstract

The link between globalization and economic growth is getting complex as the propagators of globalization are opting protectionism. The paper attempts to identify the time-varying dimension of globalization in India from Q2 1996 to Q3 2019. The aim is to capture the portion of growth explained by domestic and foreign factors suggesting the pace of globalization. The factor-induced domestic and foreign loadings are used in Time-Varying Parameter Regression and Time-Varying Parameter Autoregressive approach to indicate the evidence for slowbalization in India. The models combined with stochastic volatility, aid in capturing the structural changes in the economy in a robust manner.

Keywords Time-varying parameter estimations · Globalization · Economic growth · Slowbalization

JEL classification C110 · C520 · C820 · F620 · F690 · F630

1 Introduction

Whether globalization brings cost or benefits to an economy, has always been a hot topic among researchers. Globalization, in simple words, can be explained as the integration of economies across the world. However, in the recent past, we have seen a reversal of the phenomenon. The economies which were once the propagators of the globalization are opting for protectionism and nationalism.

A dearth of literature is available reflecting the impact of globalization on economic growth using various time series tools and techniques. However, the dynamics of globalization cannot be studied without considering structural changes. Vector Autoregression models have been widely used in literature to address modelling and forecasting.

✉ Shikha Gupta
shikha.dtu@gmail.com

[1] Discipline of Economics, Department of Humanities, Delhi Technological University, Delhi, India

Our study aims to identify the time-varying dynamics of globalization in India. We estimate it by analyzing the portion of growth explained by foreign factors and domestic factors. We observe that the time-varying coefficients of the foreign factors have been increasing but at a marginal rate since the Global Financial Crisis (GFC) 2008. The results provide evidence for a slower pace of globalization, and, hence, indicate 'slowbalization'. The term Slowbalization, was proposed by Bakas, a Dutch blogger, in 2015, who defined it as a slower pace of globalization with international linkages turning regional rather than global.

The challenge is to model the data series to produce timely and robust forecasts using macroeconomic indicators, as these indicators are subject to rapid structural change exhibiting possibility of the structural break and policy changes (Liu et al. 2012; Maier 2011). To examine this, we estimate two models; one, using Time-Varying Parameter Regression (TVP-R model) and the other using Time-Varying Parameter Vector Autoregressive model (TVP-VAR model). The models are estimated based on Bayesian inferences using Markov Chain Monte Carlo (MCMC) estimations. Both approaches validate evidence of slowbalization.

We estimate India's economic growth with the factor-induced domestic and foreign indicators.

Factor analysis approach enables us to extract information from various macroeconomic leading indicators. Further, the performance of the TVP-VAR model is compared to the TVP-Regression model. The rest of the paper is structured in four sections. The next section discusses the literature review followed by methodologies used in section 3. Section 4 discusses and compares the results of both the approaches, and finally, section 5 presents the conclusion.

2 Literature review

India has been benefitted by economic globalization. However, deep integration has also made India vulnerable to shocks across the world. The researchers assessed the relationship between globalization and economic growth, and have suggesting various measures of international integration, ranging across trade, trade restrictions, social, cultural and political factors. The inclusion of international indicators in macroeconomic models boosts accuracy, and hence the inclusion of supranational elements in the model is required (Schumacher 2010). We first find out Leading Economic Indicators (LEIs) exhibiting a similar pattern of fluctuations as that of Gross Domestic Product (GDP), so that the indicators reflect the true fluctuations in business cycles. Factor models help in extracting a common pattern from multiple indicators. Globalization effect can be measured by examining the magnitude of the foreign effect on the growth of the economy (Reklaite 2015). Going ahead, the study examines the dynamics of this effect and evaluate the pace of globalization over time.

The dynamics of the growth and globalization are of particular interest to both the researchers and policymakers. Stock and Watson (1989) suggested that reference cycles can be measured by observing the co-movements of various indicators, given by one common unobserved variable. They introduced a model to gauge the unobserved variable that represents the "current state of the economy". Engle and Watson (1981), Geweke (1977) and Sargent and Sims (1977) suggest Dynamic Factor Models

(DFM) to estimate the unobserved variables. Researchers like Mendez et al. (2001) have employed |DFM to capture and forecast the GDP growth rate of European countries based on Automatic Leading Indicator model.

We use the time-varying estimations to gauge the time-varying coefficients of domestic and foreign factors. Bhattacharya et al. (2019) in forecasting GDP growth use TVPR approach and suggest that the TVPR model outperforms the conventional constant parameter models. Gupta and Kumar (2020) use Dynamic OLS to suggest the relationship of globalization, trade and economic growth in India. They use a DOLS estimate with one lead and one lag suggesting that a better measure can be observed using time-varying regressions.

TVP-VAR model was introduced by Primiceri (2005). The model since then has been used in investigating the dynamics of various macroeconomics phenomenon. The author suggested that the model can apprehend the time-varying characteristics of the economy in a more robust manner using stochastic volatility. Black (1976) proposed the phenomenon of Stochastic volatility which is very crucial for TVP- VAR models. Models using stochastic volatility are proved superior over those with constant volatility. Constant volatility ignores the possible variation of volatility in disturbances. TVP-VAR estimations are widely used after Cogley and Sargent (2001) proposed the analysis of the time-varying structure of growth using unemployment, inflation and nominal short-term interest rates in a three-variable VAR model. Later in 2005, Cogley and Sargent introduced stochastic volatility in the TVP-VAR model with non-varying structural shock. Primiceri (2005) enriched the model, allowing all the parameters to vary over time. The model is used for examining the dynamics of inflation and output growth in the US by Canova et al. (2018). Nakajima et al. (2011) studied the macroeconomic data of Japan using TVP-VAR model. Baumeister et al. (2008) examined the impact of excess liquidity shocks on macroeconomic dynamics in the euro area. Prieto et al. (2016) used Bayesian time-varying parameter estimation to suggest that during the crisis period, the transmission of the effects of the shock from the financial time series transmit to the real economy. We address a similar phenomenon with respect to shocks in the foreign sector and the domestic sector.

Our analysis is, however, based on Primiceri (2005)'s TVP-VAR model using estimation package as suggested by Nakajima (2011). The coefficients so obtained can explain the movements in growth rates and are also in line with the dynamics of the domestic and foreign leading indicators.

Inoue et al. (2017) and Karakatsani and Bunn (2008) suggested that time-varying parameter estimations outperform other constant parameter models to gauge the unobserved structural changes in an economy using rolling windows and Kalman filtering technique respectively. Eickmeier et al. (2015) and Su and Wang (2017) augmented the time-varying parameter models with principal components to capture the estimations from a large pool of indicators than selecting a limited set of indicators. Our study tries to examine the turning points and GDP growth rate using time-varying parameter estimations, using the factor-induced domestic and foreign factor loadings obtained from the factor model. Then we compare the performance of each model.

3 Methodology

Initially, a set of 11 domestic and 75 foreign LEIs are selected from the top 21 trading partner of the Indian economy (Appendix Table 3). The trade with these 21 trading partners

accounts for more than 66% of the total trade volume. Using, Dynamic Hierarchical Factor Model, factor-induced domestic and foreign loadings are estimated (Gupta and Kumar, 2020). The evaluated factor-induced domestic and foreign leading indicator series (obtained from DHFM- Appendix Table 4) are interpreted using GDP growth rates of India. This also helps in understanding the time-varying dynamics between them and the role of stochastic volatility. In TVP-R model, GDP growth rate of the next period is taken as the dependent variable. MCMC algorithm is used to calculate posterior estimates, based on 10,000 iterations, out of which 1000 are discarded as burn-in for both the methodologies.

3.1 Time-varying parameter regression

The sources of time variation so obtained with TVP-R model include both the coefficients.

The regression model following Nakajima (2011) is

$$y_t = x_t^{'}\,\beta_t + z_t^{'}\,\alpha_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}\left(0, \sigma_t^2\right), \qquad t = 1, \ldots . n, \tag{1}$$

Where, the time-varying coefficient is given as follows

$$\alpha_{t+1} = \alpha_t + \mu_t, \qquad \mu_t \sim \mathcal{N}(0, \Sigma), \qquad t = 0, ..n{-}1 \tag{2}$$

and the stochastic volatility is given by

$$\sigma_t^2 = \gamma \exp(h_t\,), \qquad h_{t+1} = \phi\, h_t + \eta_t\ , \qquad \eta_t \sim N\left(0, \sigma_n^2\right),\ \ t = 0, \ldots . . n{-}1$$

y_t represents the scalar of response. x_t and z_t represent the vector of covariates of dimension $(k \times 1)$ and $(p \times 1)$ *respectively.* β_t *is* a $(k \times 1)$ vector of constant coefficients, and α_t is a $(p \times 1)$ vector of time-varying coefficients, and h_{0t} is the stochastic volatility. The model assumes $\alpha_t = 0,\ \ \mu_0 \sim \mathcal{N}(0,)\ \Sigma, \gamma > 0$ *and* $h_0 = 0$.

Equation (1) has two portions of covariates, one suggesting the constant coefficients β, and the other suggesting the time-varying coefficients α. Since we have to observe the time-varying relationship of foreign and domestic factor loadings, the reduced form of the model is formulated as:

$$y_t = z_{1t}\ \alpha_{1t}\ \ + z_{2t}\ \alpha_{2t} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}\left(0, \sigma_t^2\right), \qquad t = \ \ 1, \ldots . . n, \tag{3}$$

where z_1 represents foreign factors, and z_2 represents the domestic factors. α_{1t} and α_{2t} are the time-varying coefficients of the foreign factors and domestic factors, respectively. The disturbance in the regression model, denoted by ε_t follows the normal distribution with the time-varying variance σ_t^2. The log–volatility, $h_t = \log \sigma_t^2/\gamma$ follows AR (1) process. The TVP-R model forms the state-space model, which is estimated using the Bayesian approach using MCMC following Nakajima (2011).

3.2 Time-varying parameter -vector autoregressive model

The sources of time variation so obtained with TVP-VAR model include both the coefficients and the variance of nature. Unlike the TVP-regression model, in this

model, the parameters are assumed to follow a random walk process. We use the same specifications, as suggested by Nakajima (2011). The reduced form model is:

$$y_t = X_t\,\beta_t + A_t^{-1}\,\Sigma_t e_t, \qquad e_t \sim \mathcal{N}(0, I_n), \qquad t = s+1, \ldots\ldots T \tag{4}$$

Where y_t is a vector of economic variables (GDP growth rate in the currents study) of size $(n \times 1)$, $X_t = I_s \otimes (y'_{t-1} \ldots\ldots y'_{t-s})$ is the matrix of the time-varying coefficients, A_t is a lower triangular matrix with elements $a_t = (a_{21,t}, a_{31,t}, a_{32,t} \ldots a_{nm-1,t})'$, wherein the diagonal elements are equal to one. Σ_t is the diagonal matrix, and the natural log for the diagonal elements of t are $\sigma_t = (\sigma_{11,t}, \ldots. \sigma_{mn,t})'$. The dynamics of the parameters hence are

$$\begin{aligned} \beta_{t+1} &= \beta_t + \mu_t^{\beta} \\ \alpha_{t+1} &= \alpha_t + \mu_t^{\alpha} \\ \sigma_{t+1} &= \sigma_t + \mu_t^{\sigma} \end{aligned}$$

The error term vector for each variable is

$$\begin{pmatrix} u_t^{\beta} \\ \hline u_t^{a} \\ u_t^{\sigma} \end{pmatrix} \sim N\left(0, \begin{pmatrix} v_{\beta} & 0 & 0 \\ 0 & v_a & 0 \\ 0 & 0 & v_{\sigma} \end{pmatrix}\right) \tag{5}$$

Where $(v_{\beta}, v_a, v_{\sigma})$ are diagonal matrices.

4 Results

4.1 TVP regression stochastic volatility

Estimation results of 10,000 draws (Fig. 1) suggests that the sample path is reliable and have low sample autocorrelations. It indicates the strength of the sampling method to produce samples with low autocorrelation. Table 1 suggests the results of posterior means, standard deviation and the 95% credible interval. Diagnostic tests for convergence and efficiency are performed following Geweke (1992), suggesting convergence to the posterior distribution at 5 % level of significance. The low inefficiency factors indicate efficient sampling for parameters.

Figure 2 plots the time-varying coefficients of the domestic factor and foreign factors, which reflect that the proportion of growth as explained by the foreign factors is increasing but at a slower rate after witnessing a fall in 2008–09, signifying the pieces of evidence of the presence of slowbalization. This trend is also supported by reflections on the regionalization of countries rather than globalization. Indicating that globalization has led to an era of sluggishness. This proves that while there is the continued integration of the global economy, it is at a much slower pace. The stochastic volatility (Fig. 3) shows moderate movements except during the global financial crisis period 2008 and reflect a slow recovery; hence, a slightly upward trend after 2010, where the world economy was recovering. In the post-recession period, the foreign

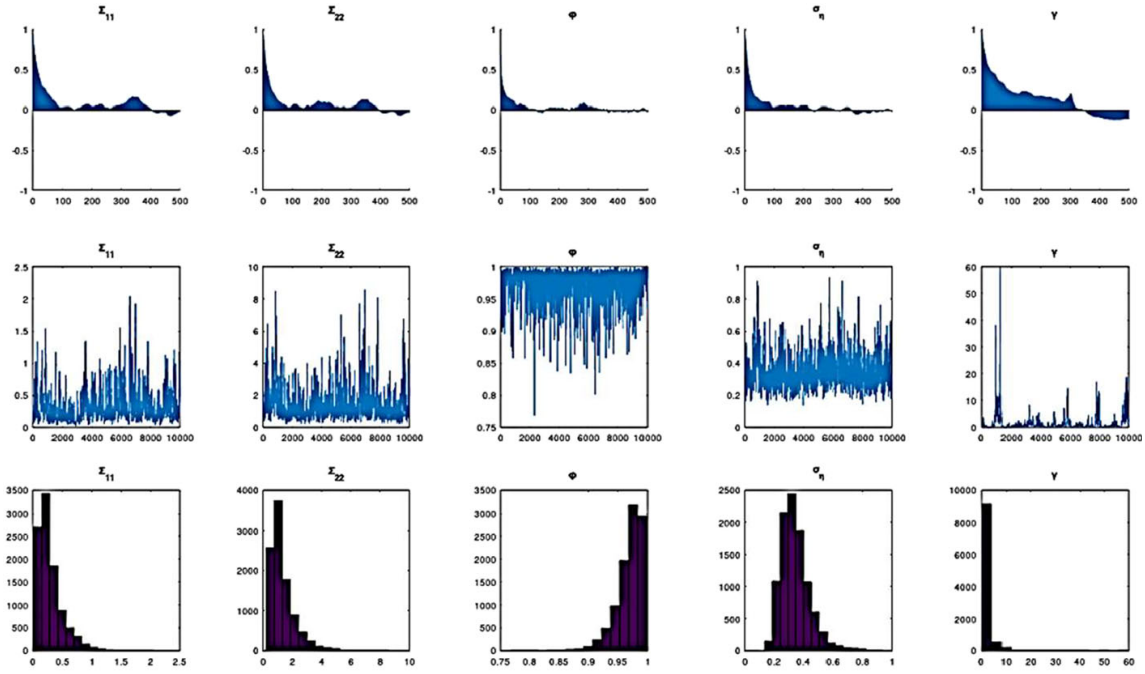


Fig. 1 Estimation Results of the TVP Regression Model (with stochastic volatility)

factors exhibited a downward trend. The posterior estimates follow the true values, which are included in the credible intervals. The growth of globalization hence is positive, but the rate of growth is slow. This is so because most of the linkages are deepening regionally. The value chains in North America, Europe and Asia are sourcing more from nearby countries. Asia and Europe have exhibited trends of spurge in intra-regional trade since 2011. Asian organizations are reported to have more exports within the continent in 2017 (The Economist, 2019).

4.2 TVP regression constant volatility

Further, we estimate the constant volatility model to examine the role of stochastic volatility. Fig. 4 suggests that the posterior means are not close to the true states for the

Table 1 Posterior estimate results of TVPR

Parameter	True	Mean	St.dev	95% Interval	Geweke	Inefficiency
a With Stochastic Volatility						
α_1	1	0.2970	0.2190	[0.0570, 0.8709]	0.043	62.01
α_2	−1	1.2905	0.8419	[0.3672, 3.8139]	0.067	20.07
phi	0.95	0.9699	0.0225	[0.9140, 0.9969]	0.000	23.31
siget	0.5	0.3442	0.1002	[0.1972, 0.5838]	0.008	35.54
gamma	0.1	1.3804	2.5879	[0.0499, 8.4900]	0.333	71.38
b With Constant Volatility						
Parameter	True	Mean	St.dev	95% Interval	Geweke	Inefficiency
α_1	1	0.9959	1.2574	[0.0095, 4.3000]	0.000	22.01
α_2	−1	1.7542	1.7757	[0.2193, 6.8982]	0.000	21.55
sigma		1.2479	0.1768	[0.9189, 1.5780]	0.000	74.28

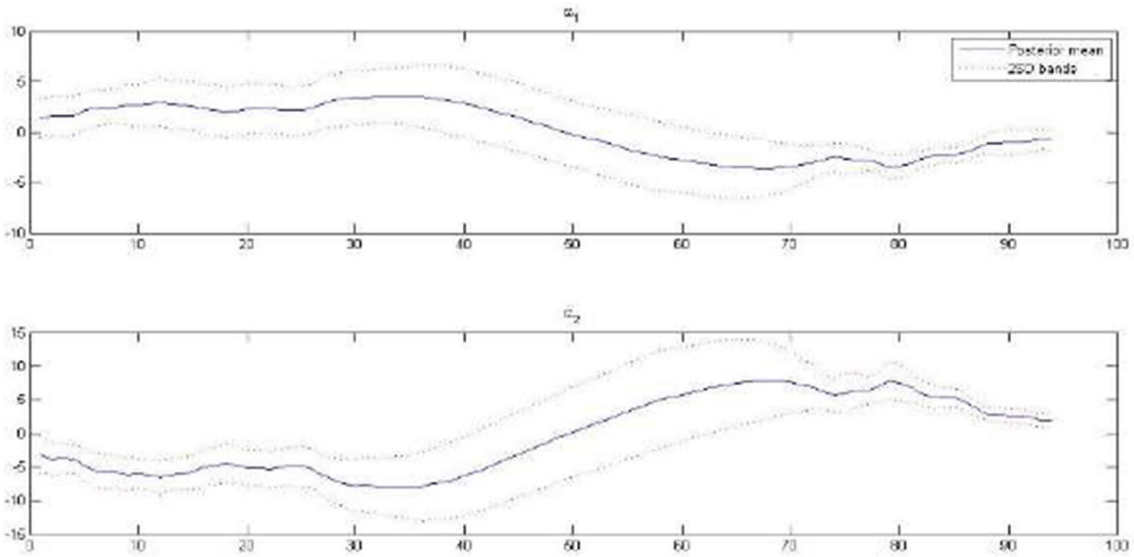


Fig. 2 Estimation results of αt -TVP Regression Model (with stochastic volatility)

time-varying coefficients. Also, the credible intervals are too wide to capture the movements of the true values (Fig. 5). Over-estimation of the constant volatility can be the possible reason for this, as the constant volatility model ignores the behavior of volatility change.

4.3 TVP VAR stochastic volatility

We further extend the estimation from a univariate TVP estimation to a multivariate TVP-VAR model. The factor-induced domestic and foreign loadings (obtained from

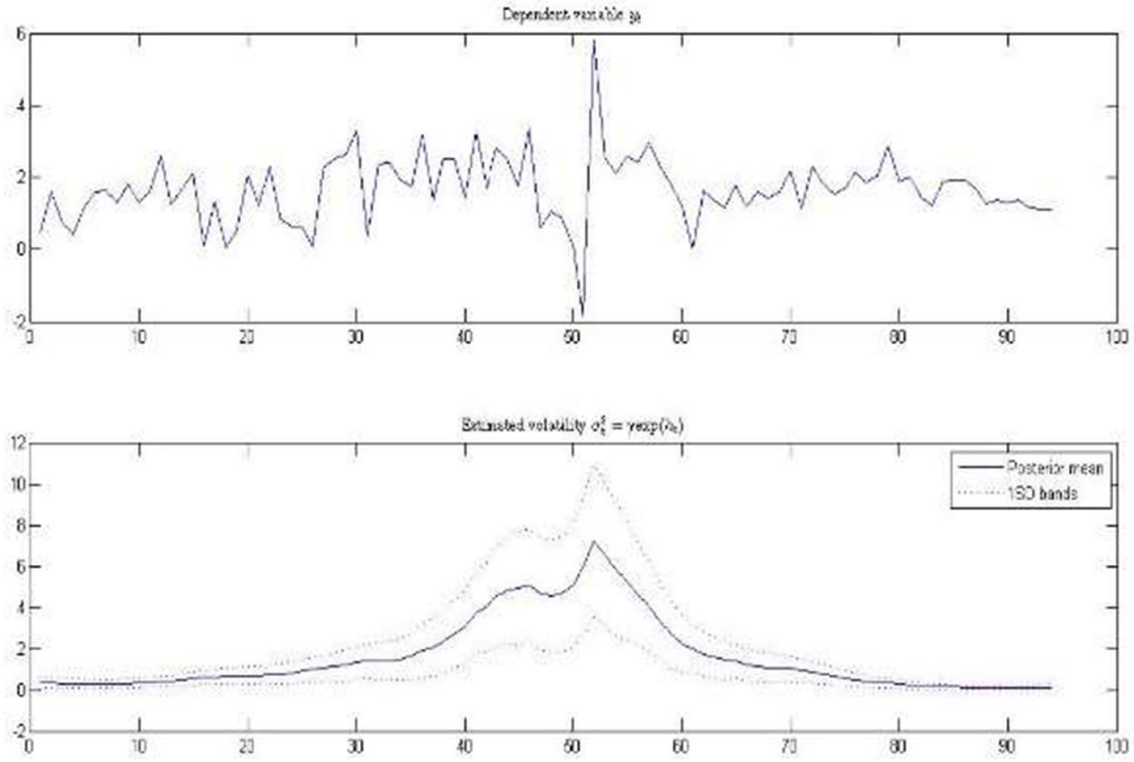


Fig. 3 Estimates of volatility - TVP Regression Model (with stochastic volatility)

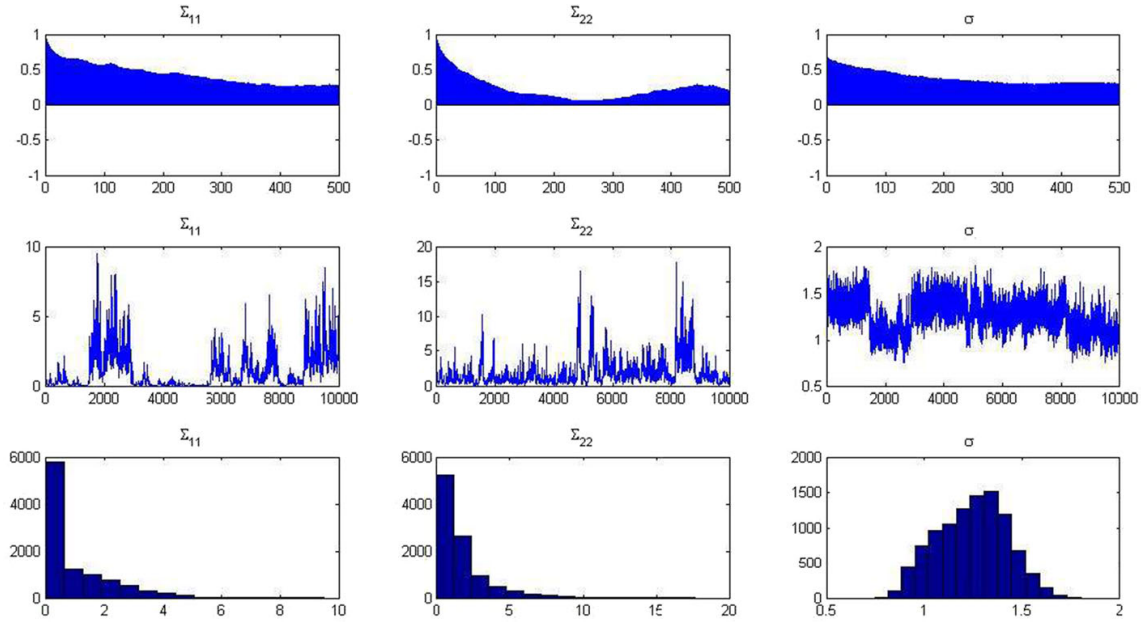


Fig. 4 Estimation Results of the TVP Regression Model (with constant volatility)

DHFM) are interpreted using GDP growth rates of India, to understand the time-varying dynamics between them using the TVP-VAR estimates (Table 2).

The sample path is reliable, and the sample autocorrelations are low, indicating the strength of the sampling method to produce samples with less autocorrelation (Fig. 6).

The low inefficiency factors indicate efficient sampling for parameters. Presence of structural breaks is confirmed with the presence of mild volatility during the periods of the Asian crisis and high volatility during the GFC. The sources of time variation in TVP-VAR model includes the coefficient as well as the variance of the shocks.

Fig. 7 plots the simultaneous relation of the domestic factor and foreign factors with the GDP growth rate, which reflects that the proportion of growth, as explained by the foreign factors, has increased in recent years. It signifies the evidence of the presence of globalization, suggested by the amount of growth explained by foreign factors.

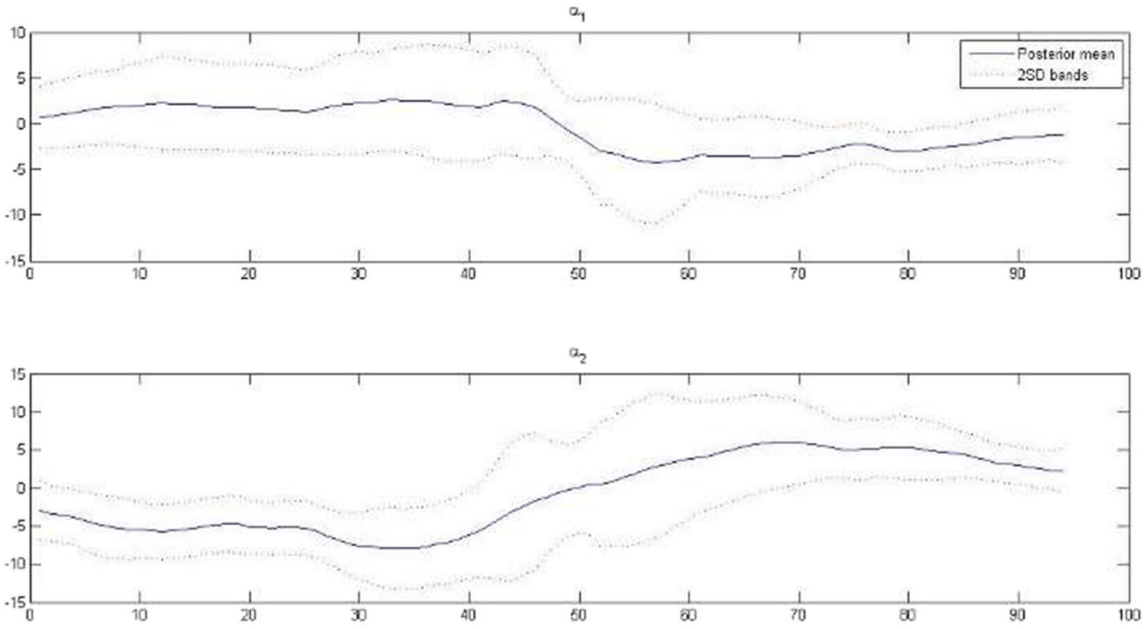


Fig. 5 Estimation Results of αt -TVP Regression Model (with constant volatility)

Table 2 Posterior estimate results from TVP-VAR model

Parameter	Mean	St.dev	95% Interval	Geweke	Inefficiency
a With Stochastic volatility					
sb1	0.0023	0.0003	[0.0018,0.0029]	0.885	7.55
sb2	0.0023	0.0003	[0.0018,0.0029]	0.011	6.04
sa1	0.0056	0.0017	[0.0034,0.0102]	0.641	40.75
sa2	0.0054	0.0015	[0.0034,0.0091]	0.951	20.64
sh1	0.1859	0.0677	[0.0934,0.3535]	0.989	37.8
sh2	0.0057	0.0021	[0.0034,0.0105]	0.113	47.97
b with constant volatility					
sb1	0.0029	0.0006	[0.0020,0.0044]	0.407	20.92
sb2	0.0028	0.0006	[0.0020,0.0042]	0.982	17.01
sa1	0.0060	0.0023	[0.0034,0.0119]	0.047	58.99
sa2	0.0056	0.0017	[0.0033,0.0097]	0.755	22.65
sh1	0.1810	0.0710	[0.0798,0.3481]	0.159	74.71
sh2	0.0059	0.0022	[0.0034,0.0113]	0.311	33.64

However, this trend is also supported by reflections on the regionalization of countries rather than globalization. It is indicating that globalization has led to an era of sluggishness. This proves that while there is the continued integration of the global economy, it is at a much slower pace. The simultaneous relations plot the impact of other variables on one unit of structural shock based on recursive identification. The simultaneous relations of foreign factor to domestic shock $(p \rightarrow x)$ stays positive with little fluctuations till 2014–15 and then tends to fall afterwards. Whereas, the simultaneous relations of the domestic factor to the growth rate $(x \rightarrow i)$ are smooth and foreign factors to growth $(p \rightarrow i)$ vary over time with some fluctuations till 2014 and are negative after that. This is supported by the decreasing trend in the global volume of

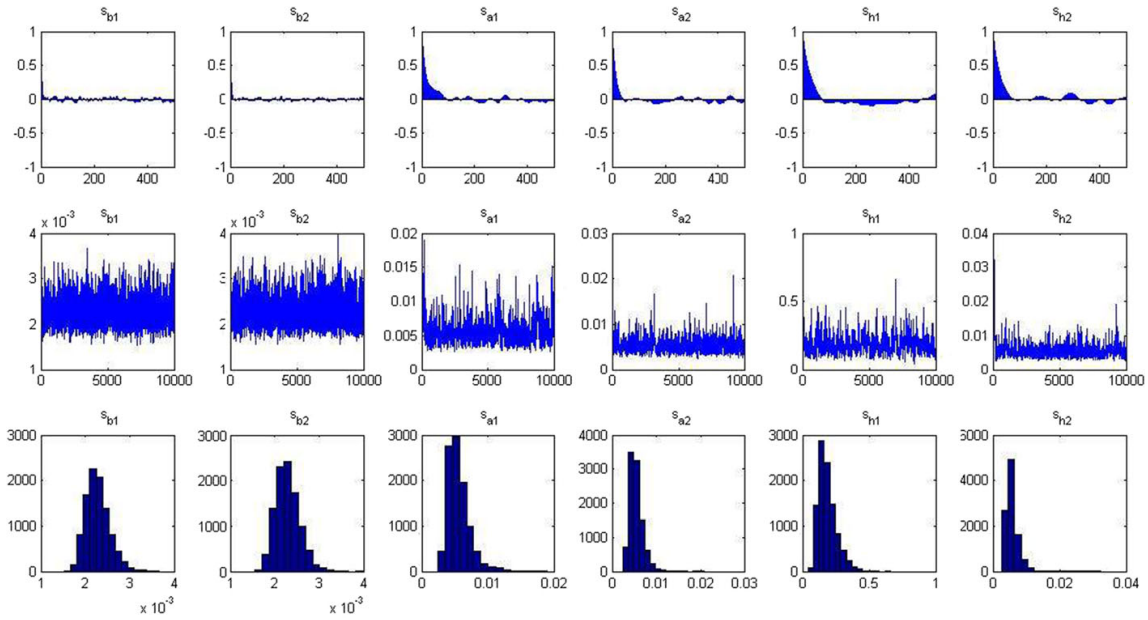


Fig. 6 Estimation Results of the TVP-VAR Model (with stochastic volatility)

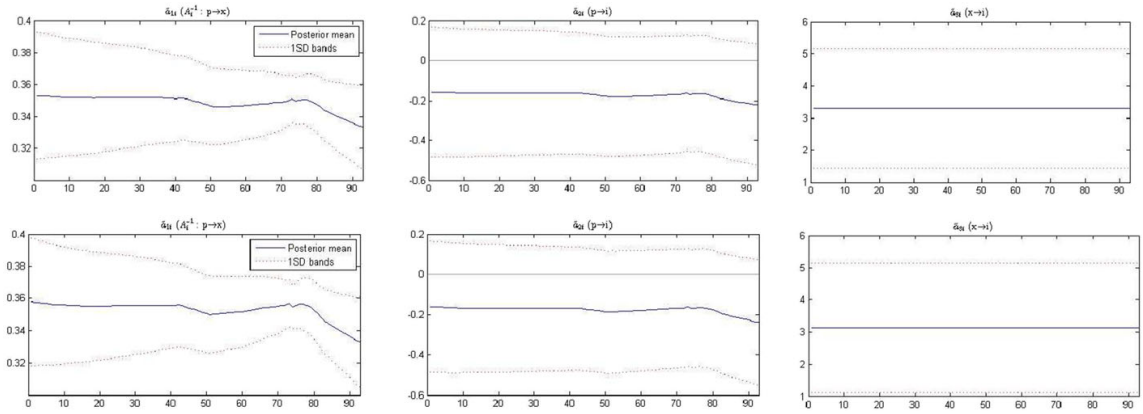


Fig. 7 Simultaneous Relations of the variables: (Row 1 with stochastic volatility and Row 2 with constant volatility) for TVP-VAR Model (with stochastic volatility)

merchandise trade. The results of TVP-VAR estimations with constant volatility suggest similar results.

Fig. 8 represents the posterior stochastic volatility and the simultaneous relation of the three variables (p, x, i) where *p* represents the foreign factor loadings, x represents domestic factor loadings, and *i* represents the ΔGDP_{t+1}. Stochastic volatility of foreign factors shows a slight upsurge during global financial crisis around the year 2008. The stochastic volatility of GDP growth shows moderate movements in the volatility in domestic factors and reflects smoother yet increasing trend. In contrast, the foreign factors reflect a slow recovery and hence a slightly upward trend after 2010. In contrast, a significant upward trend is seen in the year 2014, where the world economy was recovering. However, after 2015, again, the foreign factors exhibited a downward trend.

To gauge the evolution of the globalization effect from foreign factors, the time-varying Impulse Response functions (IRF) (Fig. 9) and Forecast Error Variance Decomposition (FEVD) (Fig. 10) of a few innovations in domestic and foreign factors are estimated. IRFs in the TVP-VAR model are estimated at all points in time. The time-varying coefficients are used to estimate the recursive innovation of the variables.

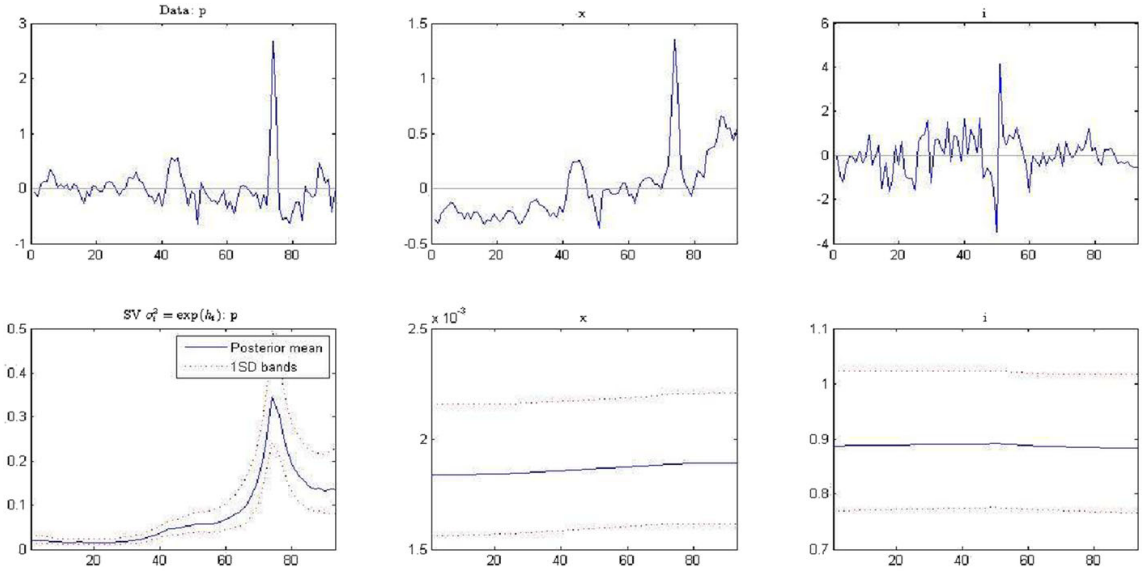


Fig. 8 Stochastic Volatility, h, for TVP-VAR Model (with stochastic volatility)

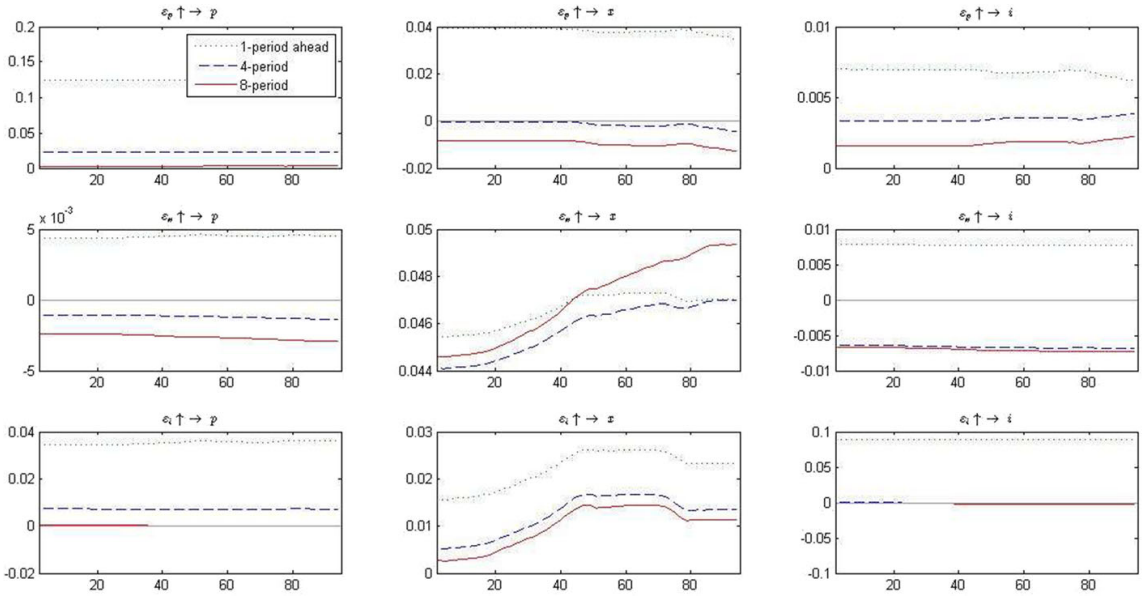


Fig. 9 Impulse Response Functions

The responses show the size of the impulses for one -quarter, one-year and two-year horizon over time. The impact of the global financial crisis was high during 2008–10 and reduced in 2012. The IRFs show the responses for the selected horizon at all points in time. The IRFs reflect that with shock in the foreign factors, the impulse responses of GDP growth rate stay positive in the next quarter and reduces in a year but still is not close zero. In the long term, the impact of the foreign shock on growth decays.

To identify the shocks that explain the growth fluctuations, we use FEVD from the three identified shocks (Fig. 10). During the Asian crisis1998–99, the global financial crisis 2006–08 and the Russian crisis, 2014, the foreign factors impacted the Indian growth negatively in the range of 0–8%, whereas, the domestic factors contributed to fluctuations in growth dynamics in the range of 0–14%. The impact of these foreign shocks on the growth of Indian economy usually are severe in the year 2008–10, and later, in the long-run, die off.

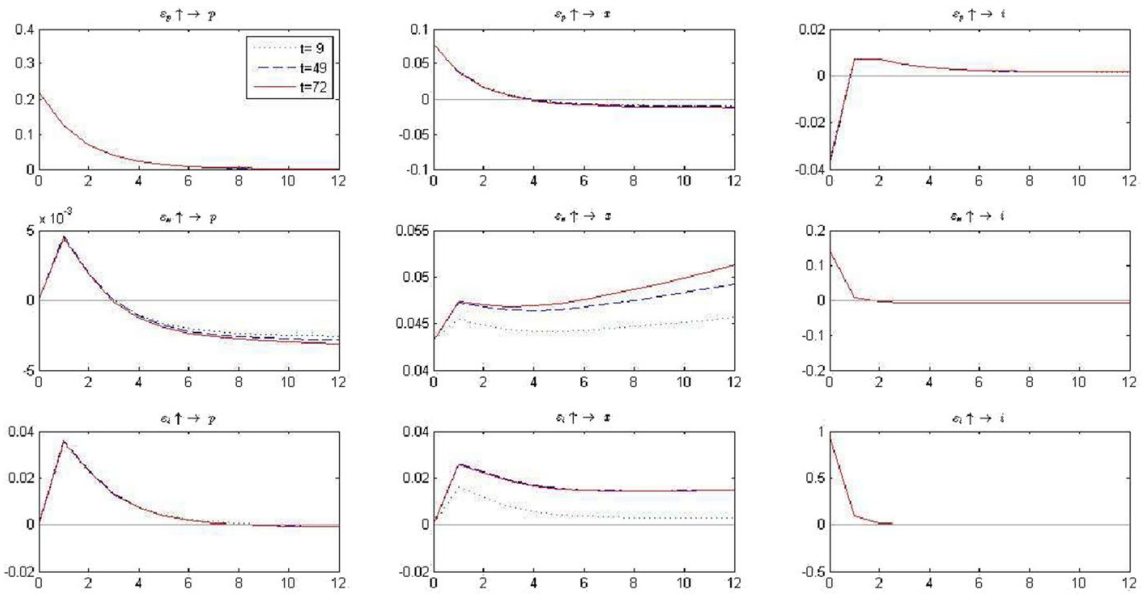


Fig. 10 FEVD

However, the fluctuations in the foreign factors bring about significant fluctuation in the domestic factors also (0–8%), which do not wither quickly. This suggests that the contagion in the foreign sector does not impact the growth directly but slows down the movement of the domestic sector. The results also suggest that while the contribution of the foreign indicators to domestic growth reduces, the contribution of domestic factors to domestic growth increases.

5 Conclusion

The two decades of globalization can indeed be called the golden period, new economies like India opened up sectors in a flurry to integrate with the global economy. However, even since the financial crisis, the pace of growth has slowed. Integration has seen a marked shift from global to 'glocal', which combines the effects of local with global. Many factors have contributed to this trend, like decreased trade, no further reduction in trade costs and the increased competitiveness of the domestic firms.

It is crucial to establish a persistent empirical validation of the growth dynamics, in the presence of profound structural changes. Unlike the conventional regression analysis and the standard SVAR model, we used TVP-R, and TVP-VAR estimates to elucidate the economic growth dynamics with its time-varying properties. In line with the suggestion proposed by Nakajima (2011), the study found TVP-VAR model to be an efficient toolkit to explore the dynamics of the economy. We found the importance of both domestic and foreign factors in the growth, but the results suggest the role of domestic factors becomes inevitable to keep the pace of the growth up.

The study indicates a slower rate of growth of globalization in India. Such periods of decreasing globalization have been traced earlier also. However, the current situation might be a different one as the economies which were once the propagators of globalization are taking up protectionist policies. The study hence provides empirical evidence to the notion of slowbalization trend in India as suggested by D'urbino (2019).

Appendix 1

Table 3 List of indicators included in the model

Region	Country	Indicator
East Asia and Pacific	Australia	Manufacturing: Industrial Confidence Indicator
		Index: Share Price: S&P/ASX 200 Index
		Consumer Confidence Indicator
	China	Monetary Aggregate M2
		Consumer Expectation Index
		Production of Manufactured Crude Steel
	Hongkong	Index of Consumer Confidence

Table 3 (continued)

Region	Country	Indicator
		Industrial Production Index: Manufacturing: 2015 = 100: QoQ%
	Indonesia	Construction Building Permit: Jakarta: Issued
		Imports
		Cargo Unloaded at 4 Main Ports: International
		Production Index: Manufacturing
		Consumer Confidence Indicator
	Japan	Index of Producer's Inventory Ratio of Finished Good: Producer Goods
		Nikkei Commodity Price Index: 42 Items
		Real Exports: 2015 = 100
	Korea	Index: Share Price: KSE KOSPI Index
		Monetary Aggregate: M1
		Early Estimates: Labour Productivity: PoP: Total
	Malaysia	Real Effective Exchange Rate Index: Based on CPI
		Government Bond Yield: Long Term
		Cargo Loaded: Klang Port
	Singapore	Interest Rate Spread
		Total Trade
		Equity Market Index: Month End: SGX Strait Times
		Industrial Production Index
	Vietnam	Imports: % of GDP
		Residential Housing (RH): Hanoi: Newly Built Area (NBA)
Europe and Central Asia	Belgium	Exports: fob
		Production: PoP: Total Industry
	Germany	Index: Share Price: CDAX Index
		Orders Inflow or Demand: Tendency: Manufacturing
		Exports: fob
		Orders Index: Total Manufactured Goods: Volume
		Finished Goods Stocks: Level: Manufacturing
	Netherlands	Manufacturing: Rate of Capacity Utilization
		Imports: cif
		Exports: fob
		Hourly Wage Rate Index: Manufacturing
	Switzerland	Purchasing Manager Index: Manufacturing
		Consumer: Expected Economic Situation
		Manufacturing: Business Situation: Future
		Index: Share Price: UBS 100 Index
		Real Effective Exchange Rate Index: CPI Based
		Monetary Aggregate: M3
		Exports: fob
		Consumer Expected Economic Situation
		GDP: NOGA 2002: FI: Real Estate, Rental, Informatics and R&D

Table 3 (continued)

Region	Country	Indicator
	United Kingdom	Index: Share Price: FT Ordinary industrial Share Price Index
		Monetary Aggregate: M2 (Cash and Retail Deposits in M4)
		Monetary Aggregate: M3
		Exports: fob
		Early Estimates: Labour Productivity: PoP: Total
		Permits Issued: Dwellings: PoP
		GDP: swda: Gross Operating Surplus & Gross Mixed-Income
		Labour Productivity
Sub-Sahara Africa	Nigeria	Risk Premium on Lending: Lending Rate Minus Treasury Bill Rate
	South Africa	Yield 10 + -Year Government Bonds
		Permits Issued: Buildings
		Sales: Total Manufactured Goods: Value
Middle East	Saudi Arabia	No of Building Permits Issued: Housing and Commercial
	United Arab Emirates	Index: Abu Dhabi Securities Exchange (ADX)
North America	USA	House Price Index: FHFA: Purchase Only
		Money Supply: Broad Money
		Consumer Confidence Index
		Manufacturing, Wholesale, & Retail: Sales
		Manufacturing, Wholesale, & Retail: Inventories
		Manufacturing, Wholesale, & Retail: Inventories to Sale Ratio
		Manufacturing Industries: New Orders
		Mfg. Ind: Shipment: Capital Goods: ow: Nondefense: Excl Aircraft
		Risk Premium on Lending: Lending Rate Minus Treasury Bill Rate
		US: Real Effective Exchange Rate Index: Based on CPI
South Asia	Bangladesh	Official Rate: End of Period: National Currency per SDR
		Industrial Production Index
		Vessel Traffic: Cargo: Imports
Domestic Indicators	India	Index: Share Price
		Real Effective Exchange Rate Index: CPI Based
		Monetary Aggregate: M3
		Exports: fob
		Imports: cif
		Industrial Production Index
		Cargo Traffic
		Registered Motor Vehicles: Cars, Jeeps and Taxies
		M3: Bank Credit to Commercial Sector
		Government Securities Yield: 10 Years
		Production Index: Total Manufacturing: Investment Goods

Appendix 2

Table 4 Factors obtained from DHFM

Time	Foreign Factor	Domestic Factor	Common Factor	Time	Foreign Factor	Domestic Factor	Common Factor
Q3–1996	0.195499993	−0.131394118	0.082750532	Q2–2008	0.000948275	0.069004946	−0.00732184
Q4–1996	−0.074472218	−0.273392866	−0.038429889	Q3–2008	−0.3244002	−0.08653635	−0.16542774
Q1–1997	−0.150452367	−0.308143473	−0.058381087	Q4–2008	−0.06054918	−0.06334438	−0.07001478
Q2–1997	0.069231536	−0.224301855	0.042506835	Q1–2009	−0.10254552	−0.19044145	−0.1119876
Q3–1997	0.121097866	−0.185264838	0.089068793	Q2–2009	−0.64763071	−0.35733014	−0.30639621
Q4–1997	0.121497592	−0.155441738	0.118276365	Q3–2009	0.105058996	−0.0280051	0.018937001
Q1–1998	0.334614986	−0.121099524	0.1775369	Q4–2009	−0.01762442	−0.001615	−0.03011513
Q2–1998	0.260583959	−0.139085557	0.150246603	Q1–2010	−0.12924984	−0.0083577	−0.08530288
Q3–1998	0.023144048	−0.219616957	0.047208415	Q2–2010	−0.23559305	−0.05107626	−0.15418447
Q4–1998	0.079993328	−0.225357542	0.055330637	Q3–2010	−0.21753508	−0.04745312	−0.15276106
Q1–1999	0.009774293	−0.265478705	0.008788486	Q4–2010	−0.29231227	−0.04390465	−0.1843433
Q2–1999	0.079829295	−0.226659864	0.04769834	Q1–2011	−0.13965005	0.04676734	−0.11298762
Q3–1999	−0.032371033	−0.267538432	−0.018978911	Q2–2011	−0.03344665	0.047613702	−0.06696516
Q4–1999	0.076889048	−0.211892126	0.032236833	Q3–2011	−0.37450808	−0.06095074	−0.21174729
Q1–2000	0.054104081	−0.210505131	0.026594974	Q4–2011	−0.22820957	−0.05032015	−0.15324161
Q2–2000	−0.088538687	−0.266473682	−0.069756805	Q1–2012	−0.47260493	−0.13032176	−0.22776438
Q3–2000	−0.278251335	−0.320870086	−0.12298634	Q2–2012	−0.19027708	−0.01196918	−0.10088089
Q4–2000	−0.053303885	−0.273229384	−0.047860104	Q3–2012	0.035908437	0.056595017	0.00538815
Q1–2001	−0.149086734	−0.297899454	−0.065009171	Q4–2012	0.045156948	0.091854008	0.027878609
Q2–2001	0.055510914	−0.229921985	0.02066144	Q1–2013	0.022824775	0.105227706	0.007413512
Q3–2001	0.016680181	−0.266482811	−0.013331285	Q2–2013	−0.04983868	0.090897132	−0.03383869
Q4–2001	−0.05080567	−0.290241945	−0.032288475	Q3–2013	−0.15235123	0.036951321	−0.09893045
Q1–2002	−0.050150618	−0.262704515	0.005199972	Q4–2013	−0.08001248	0.035668501	−0.09088938
Q2–2002	0.086587397	−0.208300021	0.066951372	Q1–2014	−0.36456081	−0.0070048	−0.19579962
Q3–2002	0.120063449	−0.205387171	0.077615861	Q2–2014	−0.13270726	0.101665394	−0.07036937

Table 4 (continued)

Time	Foreign Factor	Domestic Factor	Common Factor	Time	Foreign Factor	Domestic Factor	Common Factor
Q4–2002	0.092762083	−0.225835081	0.056420661	Q3–2014	−0.30672437	0.213173845	−0.00488021
Q1–2003	−0.013075554	−0.274878454	−0.010483487	Q4–2014	0.760595482	0.749316302	0.640707414
Q2–2003	−0.1238097	−0.318638141	−0.071077181	Q1–2015	2.665980663	1.357039136	1.416156883
Q3–2003	−0.041433348	−0.28379235	−0.023772162	Q2–2015	1.45172746	0.887669381	0.744475143
Q4–2003	−4.51E-05	−0.225993099	0.027493705	Q3–2015	−0.38321782	0.188863608	−0.1518867
Q1–2004	0.205173451	−0.130068001	0.127325768	Q4–2015	−0.55862911	0.064088063	−0.30664166
Q2–2004	0.174269407	−0.118286292	0.118018814	Q1–2016	−0.53663327	0.012980909	−0.35268632
Q3–2004	0.280353035	−0.099155152	0.153512716	Q2–2016	−0.62912322	−0.07169283	−0.41048942
Q4–2004	0.139695597	−0.146997702	0.096127865	Q3–2016	−0.43402097	0.052056766	−0.28859579
Q1–2005	0.107809883	−0.161378581	0.051739688	Q4–2016	−0.24527941	0.162308314	−0.19559506
Q2–2005	−0.01148075	−0.189713349	−0.007726008	Q1–2017	−0.2610536	0.134621317	−0.21361099
Q3–2005	−0.140710387	−0.253709718	−0.091514707	Q2–2017	−0.57599234	0.103912585	−0.29762261
Q4–2005	−0.145357448	−0.224353219	−0.089841052	Q3–2017	0.027010262	0.348856859	−0.02922895
Q1–2006	−0.277774677	−0.259452394	−0.15816138	Q4–2017	−0.07609842	0.366118504	−0.04976532
Q2–2006	−0.085518591	−0.171333171	−0.093456044	Q1–2018	−0.16138692	0.38289691	−0.06548862
Q3–2006	−0.328659837	−0.211608438	−0.168299144	Q2–2018	−0.09407898	0.469927969	0.015033454
Q4–2006	−0.059009366	−0.087963191	−0.014740965	Q3–2018	0.465948854	0.659866377	0.252717945
Q1–2007	0.319899824	0.130929906	0.216419373	Q4–2018	0.324473616	0.65588329	0.176056485
Q2–2007	0.560554846	0.250664371	0.356770189	Q1–2019	0.088534763	0.544918687	0.080147704
Q3–2007	0.498611488	0.24899012	0.336902498	Q2–2019	0.163057668	0.556049916	0.049964277
Q4–2007	0.557070557	0.257733366	0.31142322	Q3–2019	−0.4348129	0.437371004	−0.17189753
Q1–2008	0.237598585	0.192424511	0.119452516	Q4–2019	0.022174631	0.548459766	0.003734223

References

Baumeister C, Durinck EJ, Peersman G (2008) Liquidity, inflation and asset prices in a time-varying framework for the euro area. National Bank of Belgium Working Paper, (142). http://sci-hub.tw/10.2139/ssrn.1685159

Bhattacharya R, Chakravartti P, Mundle S (2019) Forecasting India's economic growth: a time-varying parameter regression approach. Macroeconomics and Finance in Emerging Market Economies 12(3): 205–228. https://doi.org/10.2139/ssrn.1685159

Black, F. (1976). Studies of stock market volatility changes. *1976 Proceedings of the American Statistical Association Business and Economic Statistics Section*

Camba-Mendez G, Kapetanios G, Smith RJ, Weale MR (2001) An automatic leading indicator of economic activity: forecasting GDP growth for European countries. Econ J 4(1):S56–S90. https://doi.org/10.1111/1368-423X.00053

Canova F, Ferroni F, Matthes C (2018) Detecting and analyzing the effects of time-varying parameters in dsge models. Int Econ Rev. https://doi.org/10.1111/iere.12418

Cogley T, Sargent TJ (2001) Evolving post-world war II US inflation dynamics. NBER Macroecon Annu 16: 331–373. https://doi.org/10.1086/654451

D'urbino, L. (2019). Slowbalisation: The steam has gone out of globalization. Retrieved from https://www.economist.com/leaders/2019/01/24/the-steam-has-gone-out-of-globalisation

Eickmeier S, Lemke W, Marcellino M (2015) Classical time-varying factor-augmented vector auto-regressive models—estimation, forecasting and structural analysis. Journal of the Royal Statistical Society: Series A (Statistics in Society) 178(3):493–533 www.jstor.org/stable/43965750

Engle R, Watson M (1981) A one-factor multivariate time series model of metropolitan wage rates. J Am Stat Assoc, 76(376), 774–781. https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1981.10477720

Geweke, J. (1977). The dynamic factor analysis of economic time series. *Latent variables in socio-economic models*

Geweke J (1992) Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. Bayesian statistics 4:641–649

Gupta, S., & Kumar, N. (2020) Has globalization reaped the rewards? A fresh perspective from India. *International Journal of Information and decision sciences*

Inoue A, Jin L, Rossi B (2017) Rolling window selection for out-of-sample forecasting with time-varying parameters. J Econ 196(1):55–67. https://doi.org/10.1016/j.jeconom.2016.03.006

Karakatsani NV, Bunn DW (2008) Forecasting electricity prices: the impact of fundamentals and time-varying coefficients. Int J Forecast 24(4):764–785. https://doi.org/10.1016/j.ijforecast.2008.09.008

Liu P, Matheson T, Romeu R (2012) Real-time forecasts of economic activity for Latin American economies. Econ Model 29(4):1090–1098. https://doi.org/10.1016/j.econmod.2012.03.020

Maier, P. (2011). *Mixed frequency forecasts for Chinese GDP* (No. 2011, 11). Bank of Canada Working Paper

Nakajima J (2011) *Time-varying parameter VAR model with stochastic volatility: An overview of methodology and empirical applications* (no. 11-e-09). Institute for Monetary and Economic Studies, Bank of Japan. https://econpapers.repec.org/RePEc:ime:imemes:v:29:y:2011:p:107-142

Prieto E, Eickmeier S, Marcellino M (2016) Time variation in macro-financial linkages. J Appl Econ 31(7): 1215–1233. https://doi.org/10.1002/jae.2499

Primiceri GE (2005) Time-varying structural vector autoregressions and monetary policy. Rev Econ Stud 72(3):821–852. https://doi.org/10.1111/j.1467-937X.2005.00353.x

Reklaite A (2015) Globalization effect measure via hierarchical dynamic factor modelling. Equilibrium Quarterly Journal of Economics and Economic Policy 10, no. 3: 139–149. http://sci-hub.tw/10.12775/EQUIL.2015.029

Sargent TJ, Sims CA (1977) Business cycle modelling without pretending to have too much a priori economic theory. New methods in business cycle research 1:145–168

Schumacher C (2010) Factor forecasting using international targeted predictors: the case of German GDP. Econ Lett 107(2):95–98. https://doi.org/10.1016/j.econlet.2009.12.036

Stock JH, Watson MW (1989) New indexes of coincident and leading economic indicators. NBER Macroecon Annu 4:351–394. https://doi.org/10.1086/654119

Su L, Wang X (2017) On time-varying factor models: estimation and testing. J Econ 198(1):84–101. https://doi.org/10.1016/j.jeconom.2016.12.004

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

Transfer Learning by Deep Tuning of Pre-trained Networks for Pulmonary Nodule Detection

Dhaarna Sethi, Kriti Arora, and Seba Susan
Department of Information Technology,
Delhi Technological University,
Shabad Daulatpur, Bawana Road, Delhi, India-110042
seba_406@yahoo.in

***Abstract*—Lung Cancer is one of the most common forms of cancer found worldwide and the accurate detection of lung nodules from computed tomography (CT) scans is a crucial preliminary step in the diagnosis procedure. In this paper, we address the problem of pulmonary nodule detection with an aim to design a robust model using deep convolutional neural networks (CNN) with transfer learning and fine-tuning of network weights. We use the benchmark LIDC/IDRI dataset which is derived from a set of CT scans that are cropped using annotations provided by a radiologist. Our study focuses on observing the effects of knowledge transfer from a non-medical domain to a medical domain using the pre-trained network architectures of VGG-19, ResNet-50, MobileNet and Inception-V3. We also analyze the effects of fine-tuning on the performance of the network. Shallow tuning refers to fine-tuning only the last few layers of the deep network while deep tuning refers to fine-tuning all the layers of the deep convolutional network. We compare the performance of the state-of-the-art convolutional architectures pre-trained on the ImageNet database, for both shallow and deep tuning.**

Keywords—Convolutional Neural Network, Pulmonary Nodule Detection, Lung Cancer, ImageNet, Transfer Learning, Shallow Tuning, Deep Tuning

I. INTRODUCTION

Lung Cancer is the leading cause of cancer-related deaths worldwide and is the second most common form of cancer [1]. Early detection of the same increases the likelihood of survival. The first step in the diagnosis of lung cancer is to detect pulmonary nodules from CT Scans followed by the classification of the nodule as malignant or benign in the later stage of diagnosis [2]. An undetected nodule can prove to be fatal for the patient and a false positive sample can prove to be costly for both the patient and the doctor. In this paper, we address the problem of pulmonary nodule detection with the aim to design a robust model using deep neural networks with transfer learning and fine-tuning.

Computer-aided diagnosis was earlier guided by pixel- and texture-based manipulations [3, 4]. Now deep learning architectures have outperformed the conventional handcrafted features for computer vision tasks [5]. Deep learning with convolutional neural networks (CNN) has proven to be successful and effective in the field of medical image classification [6]. In [7], a 20-layer deep CNN is trained from scratch for detecting the lung nodules. However, training CNNs from scratch using random weight initialization requires large amounts of data from all classes, that is difficult to acquire in the medical domain. Synthetically augmented medical datasets are compiled by researchers to address the issue. Multiple streams of convolutional networks are used in [8] to capture different views of lung nodules, and a decision fusion combines the outputs of the multiple streams. Data augmentation, exclusively of the minority class, was performed in [9] to balance the class distribution prior to deep learning.

CNN models pre-trained on existing vast databases leverage knowledge learned from different domains and reduce the problem space for the given task [10]. Knowledge transfer from a non-medical domain such as the ImageNet dataset for medical diagnosis, is the subject of critical study and analysis by several experts [11, 12]. As per [6], there are two techniques by which this knowledge transfer could take place: (i) by using the pre-trained network for feature extraction (ii) by fine-tuning the pre-trained network on medical data. There are several proved cases of successful cross-domain transfer for medical applications. We discuss some of the techniques used for lung nodule and lung cancer detection below. The variation in appearance and size of the lung nodules poses a challenge to any detection algorithm. The ResNet pre-trained architecture was used successfully for detecting malignancy from lung nodule images in [13], and 50.66% positive cases were reported. Deep features were extracted from VGG-16 pre-trained network and trained on a support vector machine (SVM) classifier for the detection of lung nodules in [14]. Fine-tuning of the last few fully connected layers was performed prior to the deep feature extraction. The 4096-dimensional features extracted from the penultimate layer of the deep pre-trained network OverFeat were classified using SVM for lung nodule detection in [15]. Three ResNet-50 pre-trained models were trained in [16] on three different views of lung nodules to capture the variety in shape and size of the nodules in 3D CT scans.

The previous work of the authors introduced a class-weighted convolutional auto-encoder for distinguishing lung nodules from non-nodules in CT scan images [17]. The trained encoder part of the auto-encoder was then concatenated with fully connected layers and the end-to-end models was fine-tuned on the lung nodule images. As an extension of our previous work, in the current study, we assess how different deep convolutional architectures that are pre-trained on the ImageNet database compare with each

other for both shallow and deep tuning of network weights. We also compare our results with network architectures initialized with uniformly distributed random weights. In Section II, we discuss the architectures of the pre-trained networks followed by the training methodology. In Section III we explain the experimental setup; in section IV, we compare the results obtained after performing the experiments. The conclusions are summarized in Section V.

II. Shallow and Deep Tuning of Pre-trained Networks

The pre-trained networks are deep convolutional networks that have been extensively trained on a large-scale database like ImageNet. Certain works advocate using the pre-trained networks for cross-domain classification without any modification of weights [18]. However, the majority of researchers resort to fine-tuning of network weights to suit their target application. Fine-tuning involves copying the weights of a pre-trained network onto the new deep network, except for the last fully connected layer(s) whose weights are randomly initialized, and optimizing the weights of all the layers of the new network by training on the medical dataset. Either all the layers could be fine-tuned together [19] or the layers could be trained in a greedy manner starting with the last fully connected layer [20]. Alternatively, the original pre-trained network, with all layers intact, can be fine-tuned on the medical dataset [21]. A suggestive work that adaptively determines which pre-trained layers should be fine-tuned is that of Guo *et al.* in [22]. We discuss next the various aspects of the pre-trained networks and the fine-tuning procedure used in our experiments.

A. Pre-trained Networks

In our study, we focus on four state-of-the-art pre-trained deep architectures: - VGG-19 [23], Resnet-50 [24], MobileNet [25] and Inception-V3 [26]. All the networks are pre-trained on the large visual ImageNet database containing over a million images belonging to thousand classes. The architectures are CNNs having varying widths and a large number of trainable parameters. For uniformity, images of dimensions 75x75x3 are used as input to all models. A dense softmax layer having two neurons is added to the end of the network for classification of the images as "nodule" or "non-nodule".

a) VGG-19: The Visual Geometry Group (VGG) network architecture was introduced by Simonyan and Zisserman [23] in 2014. It contains 3x3 convolutional layers stacked together increasing the depth. Max pooling layers have been used in between to reduce the volume. There are two popular variants of the model that are currently in use: - VGG-16 and VGG-19, where 16 and 19 refer to the number of layers in the model. In our study, we use the deeper network, i.e. VGG-19.

b) ResNet-50: ResNet was first introduced by He *et al.* [24] in 2015. The residual network architecture is deeper than the VGG-19 network and consists of 50 convolutional layers. It incorporates skip connections to improve the learning accuracy of the model. Deep architectures often suffer from the vanishing gradient problem. The skip connections allow the gradients to back-propagate without attenuation. The model size is reduced due to the use of global averaging pooling layer rather than fully connected layers at the output.

c) MobileNet: MobileNet was introduced by Howard *et al.* [25] in 2017. The MobileNet architecture incorporates the concept of depth-wise separable convolutions which reduces model size and complexity. As per this procedure, a single filter is applied to each color channel. There is a separate layer for filtering and combining, which considerably reduces the cost of computation.

d) Inception-V3: The Inception architecture was introduced by Szegedy *et al*. [26] in 2014. The original architecture was called GoogleLeNet but subsequently the future versions were called Inception-Vn where n refers to the version number. The first version, Inception-V1 was improved by introducing batch normalization in Inception-V2. The network was further refined by implementing factorization methods in the third version, i.e. Inception-V3.

B. Training and Transfer Learning

A good introduction of deep convolutional networks used for computer aided detection, and their inherent characteristics, are presented in the study by Shin *et al.* in [10]. Each model must be assigned some initial weights for starting the training process. The initial weights determine the learning efficiency of the model. The weights of a neural network are normally optimized by gradient descent; the weights therefore tend to converge to local minima [27], though globally optimized network weights are feasible at some computational cost [28]. We examine the effects of using pre-trained weights to initialize the CNN, that are derived from a non-medical domain i.e. the ImageNet dataset. We compare the effects of shallow tuning versus deep tuning of the pre-trained deep convolutional neural networks for the task of lung nodule detection. Shallow tuning refers to tuning only the last few layers of the deep network while deep tuning refers to fine-tuning all the layers of the deep network [20].

a) Random initialization of weights: The CNN architectures are first trained from scratch by initializing the network with weights derived from a uniform distribution. Due to the random initialization, the model has no apriori knowledge and all features are learned by training on the LIDC/IDRI training dataset. The random nature of the initialization ensures that there is no predefined bias in the weights towards a particular class. This method of weight initialization requires a lot of training on a large target dataset to achieve optimal weight values.

b) Pre-trained ImageNet weights: The ImageNet [29] dataset is a large image dataset containing 1000 object classes and 1.2 million images. The ImageNet Large Scale Visual Recognition Challenge [30] (ILSVRC) evaluates different models for object and image classification.

As the architectures mentioned above (ResNet-50, MobileNet etc.) have several trainable parameters, random initialization of weights requires a large number of labelled medical images for learning the optimal weight values. Despite the difference between the ImageNet dataset and medical datasets, CNN models trained on a large-scale database may be helpful in transferring knowledge learned to make tasks in the medical domain more efficient.

c) Shallow tuning: The initial layers of the deep convolutional networks learn generic, low-level features that are common across all computer vision tasks. However, the higher-end layers learn high-level features specific to the dataset being trained upon. Shallow tuning refers to fine-tuning only the last few fully connected layers of the deep network that contribute to high-level features. This is equivalent to extracting features using a pre-trained network and training the features on another classifier like the support vector machine [31]. Sermanet *et al.* [32], who proposed OverFeat, fine-tuned only the last two fully connected layers in their deep network. On the same lines, we perform shallow tuning experiments by using the pre-trained features and training only the classification layer on the LIDC/IDRI dataset. Parameters in the convolutional layers of the architectures are frozen as shown in Fig. 1 and the existing classification layer is removed and replaced with a new layer initialized with random weights. The classification layer is fine-tuned for the classification of nodules and non-nodules. Therefore, using shallow tuning, the medical images are classified based on the features learned from the ImageNet dataset.

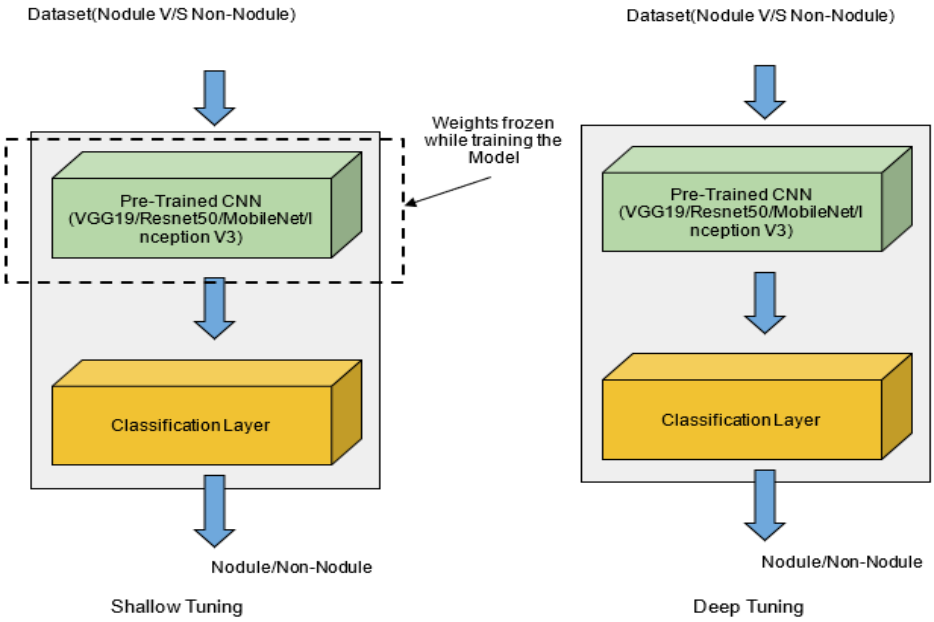


Fig. 1. Shallow Tuning versus Deep Tuning

d) Deep tuning: Deep tuning refers to fine-tuning all the layers of the deep convolutional network. In the deep tuning approach, we initialize the lower layers of the model with ImageNet weights and the last few fully connected layers of the network have randomly initialized weights. We do not freeze any weights, and fine-tune the entire model on the LIDC/IDRI pulmonary nodule dataset to leverage knowledge learned from both domains.

In general, the early layers of a CNN model learn the low-level features in images such as edges which may be applicable to all computer vision tasks, thus it is observed that in many cases freezing the first few layers is enough for feature extraction. However, considering the vast difference between the domains of the ImageNet dataset and the LIDC/IDRI dataset we unfreeze all layers and fine-tune the entire model to achieve a better performance.

Fig.1 illustrates the comparison of shallow tuning with deep tuning. We compare the results obtained from the two approaches to examine whether fine-tuning the end-to-end model performs better than fine-tuning only the final fully connected layers.

III. Experimental Setup

A. Data

The data has been derived from the publicly available Lung Image Database Consortium image collection (LIDC/IDRI) [10]. The reformatted version of the same which is provided by the LUNA16 challenge has been used. It is a set of 888 computed tomography (CT) Scans along with annotations describing the coordinates and the ground-level truth. Each CT Scan is a 512 x 512 x *n* image where *n* is the number of axial scans. The number of axial scans (*n*) per CT Scan on an average is 200. Since the amount of data is huge, the images have been cropped around the coordinates using the annotations provided by radiologists. The final dataset is a set of images with dimensions 75 x 75 x 3. Fig. 2 shows some samples of images from the dataset belonging to the classes - “nodule” and “non-nodule”.

The final distribution of the dataset is given in Table I. A total of 5187 samples have been used for training where 845 are nodules and 4342 are non-nodules. The testing dataset which is used to assess the performance of the model contains 282 nodule samples and 1340 non-nodule samples. The validation dataset is used while training to tune the

parameters of the model, and it contains 224 nodule samples and 1073 non-nodule samples.

TABLE I. OVERVIEW OF THE DATASET

Dataset	*Total Samples*	*nodule Samples*	*non-nodule Samples*
Training	5187	845	4342
Testing	1622	282	1340
Validation	1297	224	1073

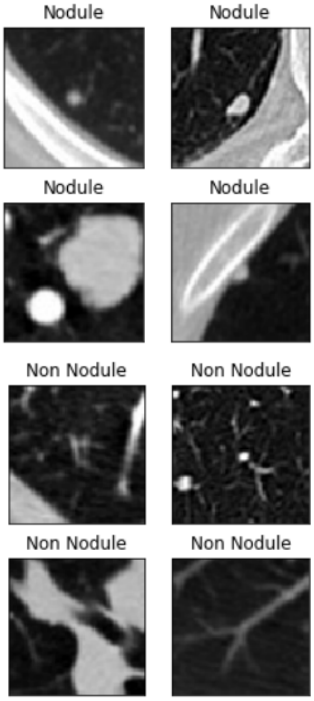


Fig. 2. Samples from the Dataset

B. *Software setup*

The original images have been preprocessed using Python and the OpenCV library. Each patch nodule is cropped into desired dimensions using the coordinates given and by ensuring that there is no intersection between the patches.

The Keras framework with TensorFlow as the backend is used for implementation of all the architectures. All the pre-trained models along with the pre-trained weights from ImageNet are provided by Keras. The Python code has been executed on the online Jupyter Notebook based service, Colaboratory by Google Research. The Tesla P4 GPU is used for faster processing which is provided by Colaboratory. All the networks are trained using the Adadelta Optimizer and the loss function employed is the Mean Squared Error. While training, the batch size is fixed at 128 and the number of epochs is set to 200.

IV. RESULTS

Table II summarizes the results obtained by using state-of-the-art deep convolutional models, i.e., VGG-19, Resnet-50, MobileNet and Inception-V3, and training by a) initializing the model with random weights, b) shallow tuning and c) deep tuning. In case of shallow and deep tuning, the initial low-level convolutional layers of the pre-trained networks are initialized with the pre-trained weights while the higher-level fully connected layers are randomly initialized. In random weight initialization, all the weights in the pre-trained architecture are randomly initialized from a uniform distribution.

The parameter used to measure the performance of all of the models is accuracy which is the percentage of correct classification. As observed from Table II, Inception-V3 with deep tuning achieves the maximum accuracy, i.e. 96.05%. Deep tuning of all layers has proved to be the superior weight optimization strategy for all the architectures.

TABLE II. PERFORMANCE OF PRE-TRAINED NETWORKS WITH SHALLOW AND DEEP TUNING AND RANDOM WEIGHT INITIALIZATION

Pre-trained network model	*Random Weight Initialization (in % Accuracy)*	*Shallow tuning (in % Accuracy)*	*Deep tuning (in % Accuracy)*
VGG-19	91.8	89.64	94.88
Resnet-50	82.6	83.1	95.35
MobileNet	91.9	81.3	95.56
Inception-V3	94.7	80.1	96.05

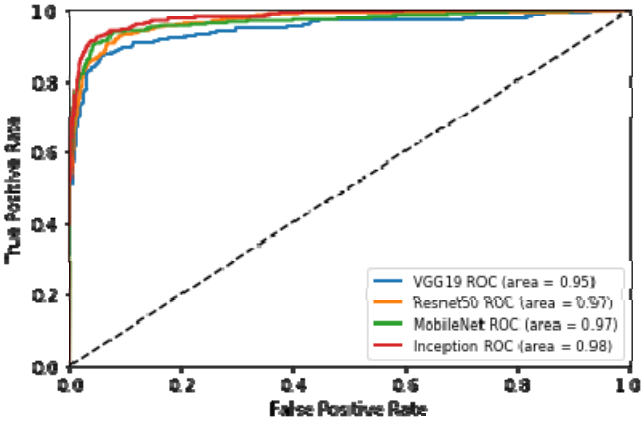


Fig. 3. ROC Curves for different pre-trained network architectures

In medical tasks we need the number of false positive cases to be very low, therefore we also need to examine the Receiver Operating Characteristic (ROC) curves to determine the ability of the models to accurately

discriminate between the two classes - nodule and non-nodule. Fig. 3 shows the ROC curves generated by the different pre-trained architectures, for the case of deep tuning which yields the best performance in Table II. The Inception-V3 model shows best results (=0.98) in terms of Area under Curve (AUC) readings among all models.

V. Conclusion

In this paper we have investigated different state-of-the-art deep pre-trained CNN architectures for pulmonary nodule detection, more specifically, VGG-19, ResNet-50, MobileNet and Inception-V3 for pulmonary nodule detection. We have examined the role of transfer learning to achieve a more robust and accurate detection system. Pre-trained weights were used as apriori knowledge to overcome the problem of non-availability of large amounts of training data. We compared the effects of shallow tuning with deep tuning for the four pre-trained architectures. Shallow tuning refers to tuning only the last few layers of the deep network while deep tuning refers to fine-tuning all the layers of the deep convolutional network. It is observed that the Inception-V3 model with deep tuning performs best out of all the models. Our results suggest that cross-domain transfer learning improves the accuracy of the pulmonary nodule detection, and it should be adopted more extensively for other applications in the medical domain. Our future work involves classifying the identified lung nodules as malignant or benign using pre-trained networks with deep tuning.

References

[1] Torre, Lindsey A., Rebecca L. Siegel, and Ahmedin Jemal. "Lung cancer statistics." In *Lung cancer and personalized medicine*, pp. 1-19. Springer, Cham, 2016.

[2] Dey, Raunak, Zhongjie Lu, and Yi Hong. "Diagnostic classification of lung nodules using 3D neural networks." In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 774-778. IEEE, 2018.

[3] Susan, Seba, M. Hanmandlu, Vamsi K. Madasu, and Brian C. Lovell. "Detection of skin lesions by fuzzy entropy based texel identification." In Image and Signal Processing and Analysis, 2009. ISPA 2009. Proceedings of 6th International Symposium on, pp. 244-249. IEEE, 2009.

[4] Hanmandlu, M., S. Susan, V. K. Madasu, and B. C. Lovell. "Fuzzy co-clustering of medical images using Bacterial Foraging." In *2008 23rd International Conference Image and Vision Computing New Zealand*, pp. 1-6. IEEE, 2008.

[5] Saini, Manisha, and Seba Susan. "Comparison of Deep Learning, Data Augmentation and Bag of-Visual-Words for Classification of Imbalanced Image Datasets." In *International Conference on Recent Trends in Image Processing and Pattern Recognition*, pp. 561-571. Springer, Singapore, 2018.

[6] Litjens, Geert, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I. Sánchez. "A survey on deep learning in medical image analysis." *Medical image analysis* 42 (2017): 60-88.

[7] Golan, Rotem, Christian Jacob, and Jörg Denzinger. "Lung nodule detection in CT images using deep convolutional neural networks." In *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 243-250. IEEE, 2016.

[8] Setio, Arnaud Arindra Adiyoso, Francesco Ciompi, Geert Litjens, Paul Gerke, Colin Jacobs, Sarah J. Van Riel, Mathilde Marie Winkler Wille, Matiullah Naqibullah, Clara I. Sánchez, and Bram van Ginneken. "Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks." *IEEE transactions on medical imaging* 35, no. 5 (2016): 1160-1169.

[9] Saini, Manisha, and Seba Susan. "Deep transfer with minority data augmentation for imbalanced breast cancer dataset." *Applied Soft Computing* 97 (2020): 106759.

[10] Shin, Hoo-Chang, Holger R. Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M. Summers. "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning." *IEEE transactions on medical imaging* 35, no. 5 (2016): 1285-1298.

[11] Cheplygina, Veronika, Marleen de Bruijne, and Josien PW Pluim. "Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis." *Medical image analysis* 54 (2019): 280-296.

[12] Day, Oscar, and Taghi M. Khoshgoftaar. "A survey on heterogeneous transfer learning." *Journal of Big Data* 4, no. 1 (2017): 29.

[13] Nibali, Aiden, Zhen He, and Dennis Wollersheim. "Pulmonary nodule classification with deep residual networks." *International journal of computer assisted radiology and surgery* 12, no. 10 (2017): 1799-1808.

[14] Shi, Zhenghao, Huan Hao, Minghua Zhao, Yaning Feng, Lifeng He, Yinghui Wang, and Kenji Suzuki. "A deep CNN based transfer learning method for false positive reduction." *Multimedia Tools and Applications* 78, no. 1 (2019): 1017-1033.

[15] Van Ginneken, Bram, Arnaud AA Setio, Colin Jacobs, and Francesco Ciompi. "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans." In *2015 IEEE 12th International symposium on biomedical imaging (ISBI)*, pp. 286-289. IEEE, 2015.

[16] Xie, Yutong, Yong Xia, Jianpeng Zhang, Yang Song, Dagan Feng, Michael Fulham, and Weidong Cai. "Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT." *IEEE transactions on medical imaging* 38, no. 4 (2018): 991-1004.

[17] Susan, Seba, Dhaarna Sethi, and Kriti Arora. "CW-CAE: pulmonary nodule detection from imbalanced dataset using class-weighted convolutional autoencoder." In *International Conference on innovative computing and communications*, pp. 825-833. Springer, Singapore, 2020.

[18] Reyes, Angie K., Juan C. Caicedo, and Jorge E. Camargo. "Fine-tuning Deep Convolutional Networks for Plant Recognition." *CLEF (Working Notes)* 1391 (2015): 467-475.

[19] Mohammadian, Saboora, Ali Karsaz, and Yaser M. Roshan. "Comparative study of fine-tuning of pre-trained convolutional neural networks for diabetic retinopathy screening." In *2017 24th National and 2nd International Iranian Conference on Biomedical Engineering (ICBME)*, pp. 1-6. IEEE, 2017.

[20] Tajbakhsh, Nima, Jae Y. Shin, Suryakanth R. Gurudu, R. Todd Hurst, Christopher B. Kendall, Michael B. Gotway, and Jianming Liang. "Convolutional neural networks for medical image analysis: Full training or fine-tuning?." *IEEE transactions on medical imaging* 35, no. 5 (2016): 1299-1312.

[21] He, Ge, and Zheng Wang. "A Combined Deep Learning and Semi-supervised Classification Algorithm for LS Area." In *E-Learning and Games: 12th International Conference, Edutainment 2018, Xi'an, China, June 28–30, 2018, Proceedings*, vol. 11462, p. 352. Springer, 2019.

[22] Guo, Yunhui, Honghui Shi, Abhishek Kumar, Kristen Grauman, Tajana Rosing, and Rogerio Feris. "Spottune: transfer learning through adaptive fine-tuning." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4805-4814. 2019.

[23] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).

[24] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778. 2016.

[25] Howard, Andrew G., Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam.

"Mobilenets: Efficient convolutional neural networks for mobile vision applications." *arXiv preprint arXiv:1704.04861* (2017).

[26] Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. "Rethinking the inception architecture for computer vision." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818-2826. 2016.

[27] Susan, Seba, Rohit Ranjan, Udyant Taluja, Shivang Rai, and Pranav Agarwal. "Neural net optimization by weight-entropy monitoring." In *Computational intelligence: theories, applications and future directions-volume II*, pp. 201-213. Springer, Singapore, 2019.

[28] Susan, Seba, Rohit Ranjan, Udyant Taluja, Shivang Rai, and Pranav Agarwal. "Global-best optimization of ANN trained by PSO using the non-extensive cross-entropy with Gaussian gain." *Soft Computing* (2020): 1-13.

[29] Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "ImageNet: A large-scale hierarchical image database." In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248-255. Ieee, 2009.

[30] Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang et al. "ImageNet large scale visual recognition challenge." *International journal of computer vision* 115, no. 3 (2015): 211-252.

[31] Sermanet, Pierre, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. "Overfeat: Integrated recognition, localization and detection using convolutional networks." *arXiv preprint arXiv:1312.6229* (2013).

[32] Bar, Yaniv, Idit Diamant, Lior Wolf, and Hayit Greenspan. "Deep learning with non-medical training used for chest pathology identification." In *Medical Imaging 2015: Computer-Aided Diagnosis*, vol. 9414, p. 94140V. International Society for Optics and Photonics, 2015.

[33] Armato III, Samuel G., Geoffrey McLennan, Luc Bidaut, Michael F. McNitt□Gray, Charles R. Meyer, Anthony P. Reeves, Binsheng Zhao et al. "The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans." *Medical physics* 38, no. 2 (2011): 915-931.

Proceedings of the Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)
IEEE Xplore Part Number:CFP20OSV-ART; ISBN: 978-1-7281-5464-0

Twitter data in Emotional Analysis – A study

Lokesh Singh
Computer Engineering
Delhi Technological University
New Delhi, India.
lokeshsingh.singh2@gmail.com

Prabhat Gupta
Computer Engineering
Delhi Technological University
New Delhi, India.
prabhatgupta490@gmail.com

Rahul Katarya
Computer Science Department
Delhi Technological University
New Delhi, India.
rahulkatarya@dtu.ac.in

Pragyat Jayvant
Computer Engineering
Delhi Technological University
New Delhi, India.
pragyat.jayvant9299@gmail.com

***Abstract*--As the innovation age encourages individuals to communicate their emotions via web-based media locales like Facebook, Instagram, Twitter, and so on. Numerous individuals share their musings and thoughts step by step in twitter similar to other tweets. This is considered as an exciting and attractive way to express ourselves as inspections are increasing bit by bit, due to which rundown of surveys are necessary for the job where summed up of text is required to give helpful data from the huge number of surveys. It is exceptionally hard for an individual to extricate helpful information or sum up it from the extremely enormous record. This paper focuses on comparison and analysis of different sentiment analysis techniques. It gives a far-reaching review about the recent and past examinations on sentiment analysis, challenges, and approaches for future angles.**

***Keywords*— Emotional analysis, Twitter data, sentiment analysis, social media, machine learning, tools.**

I. INTRODUCTION

The social platform is a novel and evolving computer model for the analysis and performance of social services that take place in various forums. It helps to produce smart and collaborative applications to achieve a better outcome. The wide range of social networking sites allow people to share their feelings or ideas about a particular event, product, or issue. It becomes very easy to express our feelings by writing about someone or a particular community. These texts can show immorality, bullying, trampling, bad words, and much more[1]. The extraction of such informal and separate data greatly helps to conclude various fields, although the poorly constructed format of the ideas found on the web makes the mining process a challenge.

The main challenge faced here is:

- Incompatible spellings- as people used highish, and they used different words to express something like the Hindi word मैं is mai is generally written as 'me' or 'mai'.
- Equivocal words- To solve the above challenge all words need to be informed separately.

Sensory analysis are also known as the digging of ideas. It is a type of text mining that separates text based on its type into different categories t

hus separating them to facilitate further processing. Post-processing ideas define unity and thus provide a clear idea of the concept/review phase. The information collected is analyzed and categorized namely, positive, negative, and neutral. These are the three basic classes of classification where each point is categorized according to the results obtained after consideration.

Most contact details are unorganized, making it difficult to analyze and extract important information[2]. The emotional analysis is an important method that helps to analyze and obtain the views and feelings of people on the social media platform[3]. The motivation behind this audit paper is to examine the idea of passionate examinations of twitter tweets. This paper is partitioned into different segments: Section II includes related work. Section III manages emotional analysis Section IV discusses emotional analysis techniques. Section V shows tools for performing emotional analysis. Section VI shows an emotional analysis of twitter. Section VII deals with a comparative study of emotional analysis techniques that work with Twitter data. SectionVIII concludes with the final reference.

II. RELATED WORK

In 2016, AkshayAmolik et al. [4], proposed an upgraded model for a passionate examination of twitter data about surveys on Movies. With the use of classifiers like SVM and Naïve-Bayes, these tweets are exactly separated. Naïve-Bayes has better exactness over SVM, however, lower precision and extra memory storing. SVM is more exact than the Naïve Bayes. Including Vector offers a superior enthusiastic investigation over the chose setting. The precision of the model will increment with the expansion in preparing information.

In 2019, S. Si et al. [5], the inspiration driving their work was to decide violence, disdain with the assistance of electronic media and unquestionably find and square utilizing modifiers, for instance, word vectors, incredible words (hand-made word reference), score sentiments, linguistic structures and emoji of procedural work. They have endeavored numerous AI approaches and commence to finish learning models and the outcomes show that the Vector Machine(SVM), Gradient Boosting Classifier (GBM), and XGBoost ClassifierSupport are most reasonable for this venture.

In 2018, Prabhsimran Singh et al. [6], audited the organization's procedure of making evil existences from a human perspective using the method to explore emotions and utilizing Twitter data, Tweets are gathered using the hashtag (#). Geo-based investigations (Collecting state-wise tweets).The enthusiastic investigations API used from the cloud and parceling the regions into six classes, upbeat,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 05:45:24 UTC from IEEE Xplore. Restrictions apply.

disastrous, pathetic, incredibly sprightly, unbiassed, and without data.

In 2019, M. Yadavetal. [7], utilized the information text document in the Devanagaritext, place away in the UTF-8 encoding framework. They have proposed 3 different ways to organize a mental investigation of Hindi area surveys. In Method-1, information classification is performed using the NN forecast utilizing pre-modified words. In Method-2, information order is achieved utilizing IIT-Bombay Hindi SentiWordNet(HSWN). In Method-3, information grouping is achieved utilizing NN expectations utilizing pre-personalized sentences, for example, named information. At last, they detailed reality in every possible way.

In 2016, P. Sharma et al. [8], have attempted to get ahead of political race results, using tweets in several languages. They extracted information (text) in 42,235 tweets collected longer than a month describing five philosophical groups in India, during the 2016 overall political race. They utilized Dictionary Based Naive Bayes and SVM calculation to make differentiations and set test information as impartial, negative, and sure. They take up the sentiments of Twitter users according to each philosophical group regarded as Indian.

In 2019, G. I. Ahmad et al. [9], When the assessment tried to give a last summation/synopsis of the Sentiment Analysis of Indian Languages with an un-normal enthusiasm on the cross-breed code of Indian Languages.

III. SENTIMENT ANALYSIS

A learning environment that interprets people's ideas, against a specific subject, about any occasion, and so forth unearthing of a text is known as the uncovering of thoughts or enthusiastic examinations/sentiment analysis[10]. Figure 1 shows the complete process of emotional analysis.

Analysis Levels:
Emotional analysis are divided into three different levels:

1. **Text Level Analysis:**
   This standard analysis that checks the whole data is what kind i.e. positive or negative. The text refers to the same topic.
2. **Judgment Level Analysis:**
   This standard inspects each line and checks whether each verdict speaks to a thought that is positive, negative, and neutral.

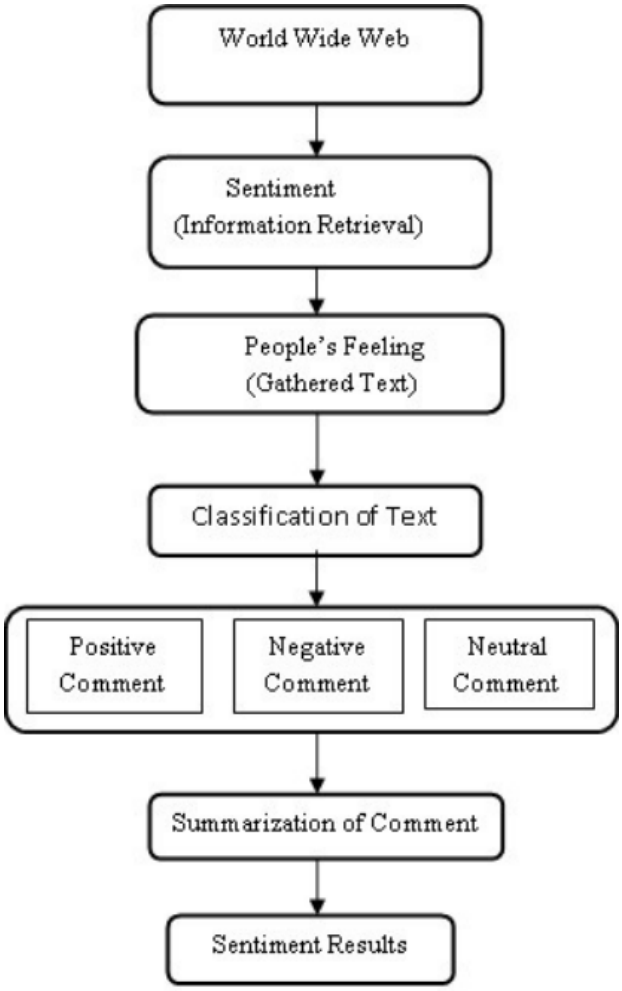


Figure 1: Emotional Analysis Process[11]

3. **body/portion Level Analysis:**
   The report and the sentence together - level investigations don't inspect the preferences of individuals. This level gives a total investigation Business/Aspect level was likewise called include level. It is based on the idea that the idea lies in the attitude and destination of the idea.

IV. SENTIMENT ANALYSIS METHODS

In the past few years, many ideas have been introduced based on erudition methods i.e. guided, supervised, or partially guided methods and using other basic-based approaches. Besides, the methods used are based on strategies identified by various domains. There is a hierarchical form of emotional analysis which is shown in Figure 2[12].

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 05:45:24 UTC from IEEE Xplore. Restrictions apply.

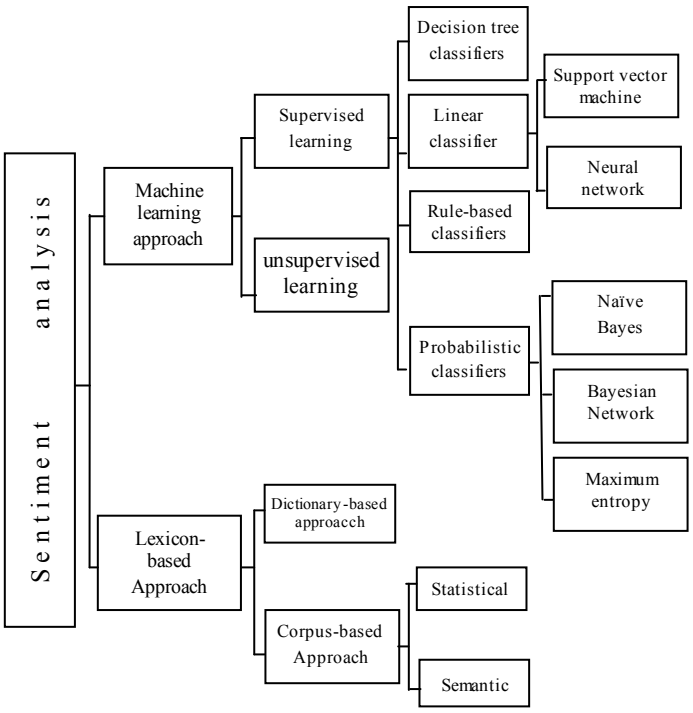


Figure 2: Separation of Emotional Analysis Methods[11]

1. **Supervised Reading**
   Research requires having a visual database and viewing labels/classes. A controlled learning approach is the way
   an effective solution in stages and provides very accurate results. Supervise classification techniques such as KNN, Naive Bayes, DT, SVM, and Regression are used[13].

2. **Unattended Reading**
   Research requires having a database with specific ideas without the need to find visual labels/classes. In the classification of texts, it is very complicated to generate labeled teaching manuals because they require human effort, which is why unregulated learning methods are very effective. Many research had been published in this area because of unsupervised learning depends on the domain and performs better[14].

3. **Less Supervised Reading:**
   Slightly monitored education uses labeled and unadulterated data. Improves supervised learning outcomes, where there are limited labeled data using less supervised reading[14].

V. TOOLS FOR SENTIMENT ANALYSIS

Open source text examination devices utilized for common language preparing (NLP, for example, extricating and division that can be utilized with web investigation:

1. **Python NLTK**
   The native language toolbox is an apparatus for text handling, bookkeeping, token making, suspension, marking, decryption, etc. Provides easy use of more than 50 corpora visual connectors and sources such as WordNet[15].

2. **WEKA**
   Weka is a tool that depends on machine learning innovations, JAVA coding language is utilized to utilize this device, and it has a Graphic user interface to show information. Numerous calculations and techniques utilized, for example, partitioning, compilation, refinement, queuing[16].

3. **Red Opals**
   Red Opals is a tool that empowers clients to decide the state of item thoughts dependent on their attributes. It offers focuses on every item dependent on an element taken from client audits.

4. **Stanford Parser**
   Utilized as a pos tagger and sentence isolation from the NLP gathering[11].

5. **Opinion Finder:**
   It underpins the distinguishing proof of individual sentences and makes various pieces of submission to the concept, including the subject header and the words include talks about good or bad communication Positions[17].

VI. EMOTIONAL ANALYSIS ON TWITTER

The function of the sensory analysis function is shown in Figure3. It contains four important modules: data module, data processing process, partition module, and output analysis.

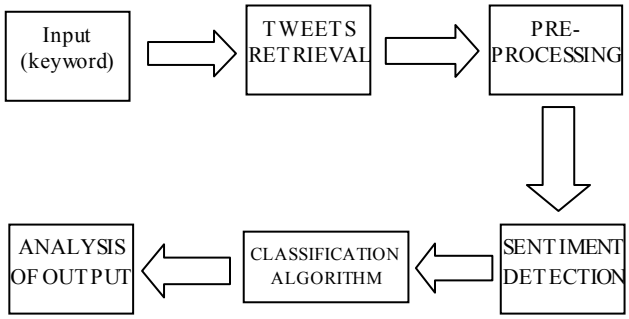


Figure 3: Sentiment Analysis Work Flow[3]

1. **Input (Keyword) module:**
   First thing is to do choose a subject, the subsequent advance is to gather the tweets by that input catchphrase and afterward play out the passionate investigations on those tweets.

2. **Tweets Recovery Module:**
   Returned Tweets can be a more sorted out, less organized, and unstructured sort. With sentiment analysis research, the tweets are gathered by utilizing R or Python or other programming languages.

3. **Pre-Processing module:**
   Information preparing/Data processing is only information separating cycle to eliminate

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 05:45:24 UTC from IEEE Xplore. Restrictions apply.

inconsistent, undesirable, and undesirable information.
Incorporates the following exercises:
•Removal of Retweets (if there should arise an occurrence of twitter dataset).
•Deleting URLs, uncommon characters, accentuation, Numbers, and so on.
• Deleting Stopwords
• Prevention
• Making tokens

4. Sensor Recovery Module:
Sensitive identification and classification made Status names can be categorized into Police, Illegal and Neutral names. A significant capacity in Sentiment investigations is to recognize the size of given tweets include. Polarities in three phases to be specific Positive, Negative, and Neutral. Extremity ID is finished utilizing different word references, for example, Bing Luis sentiment word reference, SentiWordNet and so on which assists with figuring enthusiastic force, passionate focuses, and so forth[3].

5. Phase Algorithm:

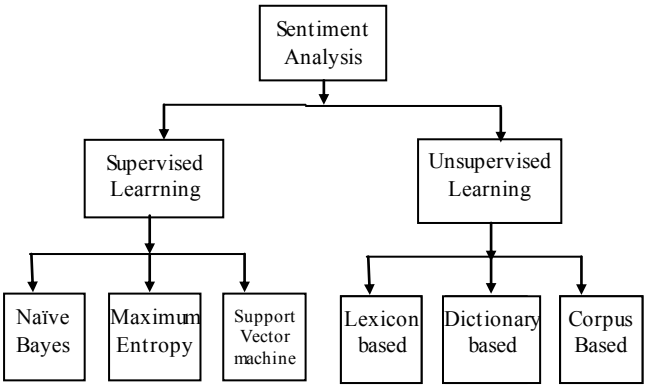


Figure 4: shows the classification of the various available algorithms[3]

6. Release module analysis:
After finishing the examinations, the outcomes are appeared on a diagram, for example, a pie outline, bar outline, and line charts.

VII. RELATIVE STUDY OF TECHNIQUES IN EMOTIONAL ANALYSIS ON TWITTER DATA

Table1: Past Research Articles on Twitter Dataset

Author & Year	Dataset	Techniques	Accuracy
Geetika Gautam [18]	Customer Review Twitter Dataset	Naive Bayes	88.2%
		Maximum Entropy	83.8%
		SVM	85.5%
		Semantic Analysis	89.9%
S. Si, A. Datta, S. Banerjee, and S. K. Naskar [5]	English dataset	Random forest	58.30
		LSTM	19.60
		Lexical and Semantic Features (WordNet)	58.75
		Majority voting(SVM, GBM, XGB)	68.13
Charu Nanda, MohitDua, and Garima Nanda [19]	Lexical resource, Annotated Nepali text, IMBD.com	Resource-based	47.2
		SVM	89.65
		NB	72.41
		Machine learning	23.6
Rincy Jose & Varghese S Chooralil [20]	Twitter feeds	SentiWordNet	21.05
		Naive Bayes	69.98
		HMM	64.06
		classifier using an ensemble approach	71.48

VIII. Discussion

This section contains the limitations of the existing work in detail.
In the wake of perusing and investigating all the past work in the field of estimation examination, it made sense of certain constraints that ought to be improved to show signs of improved precision and effectiveness in the proposed models.
Limitations:

1. Mixed language isn't an official language and thus there is no particular SentiWordNets present for it. Likewise, there is no predefined stopword list for the language.
2. For English stopwords handily eliminated however for informal dialects stopwords are taken out from the rundown and the rundown is not steady.
3. The official stemming calculations for blended language is absent thus port stemmer was utilized which diminishes the precision of the models.
4. Most of the work utilized the Twitter dataset just which limits the preparation space of the model, this influences the precision of the model.
5. Not many researches considered the emoticons in the dataset, which is a significant viewpoint in defining the polarity of a tweet.
6. data from other enormous destinations like Facebook can be utilized to build the space and size of preparing dataset, to have more exact outcomes

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 05:45:24 UTC from IEEE Xplore. Restrictions apply.

IX. CONCLUSION

The analysis of data is done in various ways with different techniques and approaches to extract sentiments and opinions. This paper illustrated the sentiment analysis and discussed different techniques and methods used in it in detail along with tools available to ease the process of analysis along with the comparison of existing approaches and techniques. As the data is increasing significantly it is not possible to process the whole data,but it is possible to extract important information from the data,which is feasible by doing sentiment analysis.

There are lots of unexplored areas in the field of sentiment analysis needed to explore to deal with present open challenges such as the failure of analytical methods to perform well in cross-country schemes, the low number of labeled data. The comparative study of various algorithms and exploration of previous work shows that exactness is improved when semantic examinations are trailed by machine learning strategies, for example, SVM, Maximum Entropy, and Naïve-Bayes. The exactness can likewise be expanded by up to 3-4% utilizing the hybrid technique. The comingadvancement in the field of sentiment analysis contains creating a process to create emotional isolation that can apply to any data outside the range.Text summarization and intelligent advanced modules can be made by applying machine learning techniques and algorithms. Emojis will be considered in the dataset as emojis are a significant perspective in conclusion examination and will absolutely build the exactness of the model.Deep-learning and progressed deep learning algorithms whenever applied on the datasets will help in achieving higher proficiency. Random forest,Regression and so forth are some extraordinary machine learning algorithms which can likewise work for characterization and forecast measure.
Also, linguistic diversity in communication data is an important issue that needs to be addressed in more recent stages. Therefore, this survey paper concludes that emotional/sentiment analysis is still an open domain for research and there is still a great scope of advanced and improved algorithms. This survey paper budding to be favorable and advantageous to those researchers who have been researching the field of sentiment/ emotional analysis.

REFERENCES

[1] N. Yadav, O. Kudale, S. Gupta, A. Rao, and A. Shitole, "Twitter Sentiment Analysis Using Machine Learning for Product Evaluation," in *Proceedings of the 5th International Conference on Inventive Computation Technologies, ICICT 2020*, Feb. 2020, pp. 181–185, doi: 10.1109/ICICT48043.2020.9112381.

[2] K. Ahmed, N. El Tazi, and A. H. Hossny, "Sentiment Analysis over Social Networks: An Overview," in *Proceedings - 2015 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2015*, Jan. 2016, pp. 2174–2179, doi: 10.1109/SMC.2015.380.

[3] R. Wagh and P. Punde, "Survey on Sentiment Analysis using Twitter Dataset," in *Proceedings of the 2nd International Conference on Electronics, Communication and Aerospace Technology, ICECA 2018*, Sep. 2018, pp. 208–211, doi: 10.1109/ICECA.2018.8474783.

[4] A. Amolik, N. Jivane, M. Bhandari, and M. Venkatesan, "Twitter Sentiment Analysis of Movie Reviews using Machine Learning Techniques."

[5] S. Si, A. Datta, S. Banerjee, and S. K. Naskar, "Aggression Detection on Multilingual Social Media Text," Jul. 2019, doi: 10.1109/ICCCNT45670.2019.8944868.

[6] P. Singh, R. S. Sawhney, and K. S. Kahlon, "Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government," *ICT Express*, vol. 4, no. 3, pp. 124–129, Sep. 2018, doi: 10.1016/j.icte.2017.03.001.

[7] M. Yadav and V. Bhojane, "Semi-supervised mix-Hindi sentiment analysis using neural network," in *Proceedings of the 9th International Conference On Cloud Computing, Data Science and Engineering, Confluence 2019*, Jan. 2019, pp. 309–314, doi: 10.1109/CONFLUENCE.2019.8776943.

[8] P. Sharma and T. S. Moh, "Prediction of Indian election using sentiment analysis on Hindi Twitter," in *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016*, 2016, pp. 1966–1971, doi: 10.1109/BigData.2016.7840818.

[9] G. I. Ahmad, J. Singla, and N. Nikita, "Review on Sentiment Analysis of Indian Languages with a Special Focus on Code Mixed Indian Languages," in *2019 International Conference on Automation, Computational and Technology Management, ICACTM 2019*, Apr. 2019, pp. 352–356, doi: 10.1109/ICACTM.2019.8776796.

[10] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–184, May 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.

[11] U. Kumari, D. Soni, and A. K. Sharma, "A Cognitive Study of Sentiment Analysis Techniques and Tools: A Survey 1," *Int. J. Comput. Sci. Technol.*, vol. 8.

[12] E. Tyagi and A. K. Sharma, "An intelligent framework for sentiment analysis of text and emotions - A review," in *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing, ICECDS 2017*, Jun. 2018, pp. 3297–3302, doi: 10.1109/ICECDS.2017.8390069.

[13] S. Brindha, K. Prabha, and S. Sukumaran, "A survey on classification techniques for text mining," Oct. 2016, doi: 10.1109/ICACCS.2016.7586371.

[14] E. Aydogan and M. Ali Akcayol, "A comprehensive survey for sentiment analysis tasks using machine learning techniques," Sep. 2016, doi: 10.1109/INISTA.2016.7571856.

[15] Weka3.7.5, "Data Mining with Open Source Machine Learning Software in Java," *Univ. Waikato*, 2011, Accessed: Sep. 10, 2020. [Online]. Available: https://www.cs.waikato.ac.nz/ml/weka/.

[16] P. Kumar Agarwal and A. K. Sharma, "Web Opinion Mining Techniques and Tools for Finding User's Opinion," 2016.

[17] "OpinionFinder 2 | MPQA." https://mpqa.cs.pitt.edu/opinionfinder/opinionfinder_2/ (accessed Sep. 11, 2020).

[18] G. Gautam and D. Yadav, "Sentiment analysis of twitter data using machine learning approaches and semantic analysis," in *2014 7th International Conference on Contemporary Computing, IC3 2014*, 2014, pp. 437–442, doi: 10.1109/IC3.2014.6897213.

[19] C. Nanda, M. Dua, and G. Nanda, "Sentiment Analysis of Movie Reviews in Hindi Language Using Machine Learning," in *Proceedings of the 2018 IEEE International Conference on Communication and Signal Processing, ICCSP 2018*, Nov. 2018, pp. 1069–1072, doi: 10.1109/ICCSP.2018.8524223.

[20] R. Jose and V. S. Chooralil, "Prediction of election result by enhanced sentiment analysis on Twitter data using Word Sense Disambiguation," in *2015 International Conference on Control, Communication and Computing India, ICCC 2015*, Mar. 2016, pp. 638–641, doi: 10.1109/ICCC.2015.7432974.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 05:45:24 UTC from IEEE Xplore. Restrictions apply.

VDDDA Based Multifunction Voltage-Mode MISO Filter

Dhruv Goel[1*], Deva Nand[1]

[1*]*Department of Electronics and Communication Engineering*
Delhi Technological University
New Delhi, India
dg210895@gmail.com[1*]

[1]*Department of Electronics and Communication Engineering*
Delhi Technological University
New Delhi, India
devkamboj07@gmail.com[1]

***Abstract* - This paper proposes a voltage-mode second-order multifunction filter that uses a voltage differencing differential difference amplifier (VDDDA) as an active building block. The proposed filter is sort of MISO multifunction filter having three input nodes and single output node. It comprises of one VDDDA, one resistor and two identical capacitors. In this proposal, the filter offers low-pass (LP), high-pass (HP), band-stop (BS), band-pass (BP), and all-pass (AP) responses for the same circuit configuration by selecting the appropriate input signals. The quality factor and natural frequency both can be electronically controllable and tunable with a dc bias current. Theoretical results have been conformed in SYMICA DE simulations using 0.18µm TSMCS CN018 CMOS process parameters with supply voltages of ±0.9V.**

Index Terms - VDDDA, analog filter, voltage-mode, MISO.

I. INTRODUCTION

Despite the significant development in digital signal processing and its real time applications, one can never replace the analog signal processing due to its particular advantages. Amplification, rectification, modulation, etc., processed on the message signals can only be performed by the analog processing. Various devices that can process the analog electrical signals have been designed and further modifications in circuits are done according to their characteristics. These modified devices could now further be used for applications such as filters, summers, clippers and clampers, amplifiers, oscillators, etc. [1].

The key application of analog signal processing is electronic filter. A filter can be designed by allowing the electrical signal of particular frequency range to pass through it depending upon their type and blocking the unwanted or undesired frequencies, which can be "low pass" , "high pass" , "band pass" , "band reject" and "all pass" filter. "Current–mode", "voltage-mode", "trans-conductance-mode" and "trans-resistance-mode filters" are various available modes of the filter operation [2]. Multifunction filter is the one that can perform two or more types of filtering functions whereas universal filter is the one that can provide all five types of filtering functions. Further according to number of inputs and outputs, filters are classified into four types which are "Single-Input-Single-Output" (SISO), "Single-Input-Multiple-Output" (SIMO), "Multiple-Input-Single-Output" (MISO) or "Multiple-Input-Multiple-Output" (MIMO) type. For the designing of MISO type voltage-mode universal filters in signal processing, designers are giving major importance due to following key features: (i) Realization of all standard five filter functions utilizing same circuit topology on the basis of input ports used, (ii) Minimum use of active and passive components in circuit design, and (iii) Design simplicity and versatility that bring cost reduction to IC manufacturer [3].

Voltage Differencing Differential Difference Amplifier (VDDDA) [4] is one such new versatile 'voltage differencing' active building block (ABB). It is a six-terminal universal block to realize various amplifiers with less transistor count and constitute dynamic properties required for any real-time applications. Device's internal structure consists of two active blocks, one is operational transconductance amplifier (OTA) [5] followed by differential difference amplifier (DDA) [6], [7]. This building block consists of three addition/subtraction input voltage terminals where the transconductance (g_m) of the block is electronically tuneable. In analog signal processing, the motivation behind this work and using VDDDA as an active building block is that it offers various advance features, for example, electronic controllability, avoiding the need of external voltage summing circuit in circuit design [3], quick circuit synthesization with least components count, etc. Especially, in the design of voltage-mode active filter, the voltage summing is required in filter thus; VDDDA is a good choice for designers. Moreover, the output impedance of VDDDA is low which allows the cascadability of devices without the need of voltage buffer. It can provide inverting as well as non-inverting voltage-mode outputs simultaneously from the same circuit topology.

This paper describes a three input and single output multifunction filter. The proposal comprises of single VDDDA, two capacitances and one resistance. All passive components are of floating type. It provides all five filter functions with three input voltage supplies which can be digitally selected. The natural frequency of a filter and its quality factor both can be electronically controllable and tuneable via a dc current source. It is cost efficient to IC manufacturer because of its simple and versatile design.

II. CIRCUIT DESIGN AND PROPOSED FILTER ANALYSIS

A. VDDDA Overview

Voltage Differencing Differential Difference Amplifier (VDDDDA) is a six-terminal universal active block. This building block involves three addition/subtraction input voltage nodes where the transconductance (g_m) of the block is electronically tuneable. The VDDDA circuit symbol and its equivalent circuit is represented in Fig. 1. V_+, V_-, Z, P and N are the five input voltage ports. The Z port is also characterized as the current output port and the W port is the voltage output port. Ideally, the input impedance at V_+, V_-, Z, P and N ports are high and the output impedance at W port is

978-1-7281-9287-1/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on January 05,2021 at 06:56:02 UTC from IEEE Xplore. Restrictions apply.

low. The difference in the input voltages, V_+ and V_- is transmitted as output current I_z via g_m which is tuneable too. The differential input voltages of V_z, V_n and V_p constitutes output voltage V_w.

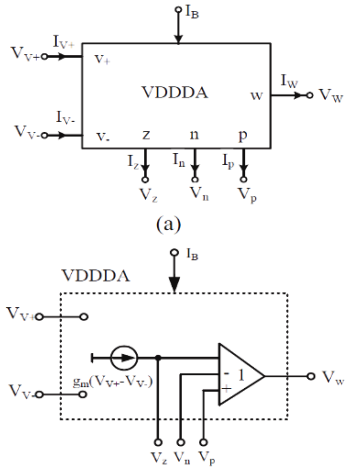


Fig. 1 (a) Symbolic representation of VDDDA, (b) Equivalent representation of VDDDA [4].

The port characteristic equation for an ideal VDDDA can be modelled in form of matrix equation which is as follows:

$$\begin{bmatrix} I_{v+} \\ I_{v-} \\ I_z \\ I_n \\ I_p \\ V_w \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ g_m & -g_m & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} V_{v+} \\ V_{v-} \\ V_z \\ V_n \\ V_p \\ I_w \end{bmatrix} \quad (1)$$

Following are the equations derived from matrix equation,

$$I_{v+} = I_{v-} = I_n = I_p \quad (2)$$

$$I_z = g_m \left(V_{v+} - V_{v-}\right) \quad (3)$$

$$V_w = V_z - V_n + V_p \quad (4)$$

where in ideal scenario, the currents at ports v_+, v_-, n and p are equal to zero from Eq. (2). In Eq. (3) and (4), g_m and V_w represent the transconductance and the output voltage of ideal VDDDA respectively.

B. Proposed MISO Filter Description

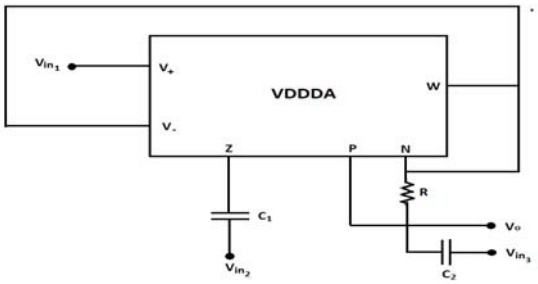


Fig. 2 VDDDA based voltage-mode MISO filter

The proposal of VM MISO filter is presented in Fig. 2. The filter comprises of single VDDDA, two capacitors and one floating resistor. In this proposal, the filter incorporates the least active and passive component count, which is a desirable feature for integration. It exhibits high input impedance for LP terminal. In future study, circuit can further be modified by replacing the resistor with CMOS implemented electronic resistor. In this filter design, there are three input voltage signals, namely V_{in1}, V_{in2}, V_{in3} and V_0 is the only output voltage. Assuming case of an ideal VDDDA block, mathematical analysis of the MISO filter offers the output voltage, which is given as,

$$V_0 = \frac{s^2 V_{in3} + s\frac{(C_1 V_{in2} + C_2 R g_m V_{in3})}{2C_1C_2R} + \frac{g_m V_{in1}}{2C_1C_2R}}{D(s)} \quad (5)$$

where,

$$D(s) = s^2 + s\frac{(C_1 + C_2 R g_m)}{2C_1C_2R} + \frac{g_m}{2C_1C_2R} \quad (6)$$

Eq. (5) concludes that the proposed MISO filter has unity gain and if we want to achieve desired gain from the active filter, designers need to add voltage amplifies to get additional gain. From Eq. (5), we can get all standard five output filter responses by selecting the input voltage terminals (V_{in}) accordingly and selection can be done digitally through below conditions,

- The second order LP response can be achieved if $V_{in1} = V_{in}$ and $V_{in2} = V_{in3} = 0$.
- The second order HP response can be achieved if $-V_{in2} = V_{in3} = V_{in}$ and $V_{in1} = 0$.
- The second order BP response can be achieved if $V_{in2} = V_{in}$ and $V_{in1} = V_{in3} = 0$.
- The second order BS response can be achieved if $V_{in1} = -V_{in2} = V_{in3} = V_{in}$.
- The second order AP response can be achieved if $V_{in1} = \frac{-V_{in2}}{3} = V_{in3} = V_{in}$.

For hardware implementation of HP and BS filter response, there is need of inverting unity gain amplifier and for AP filter implementation, inverting triple gain amplifier is required. From the characteristic equation i.e., Eq. (6), we can find the pole frequency and quality factor given by:

$$\omega_0 = \sqrt{\frac{g_m}{2C_1C_2R}} \quad (7)$$

$$Q_0 = \frac{\sqrt{2C_1C_2Rg_m}}{C_1 + C_2Rg_m} \quad (8)$$

Note from Eqs (7) and (8), the ω_0 and the Q_0 both can be electronically tuned through transconductance, g_m. However, it turns out that the proposed MISO filter requires a matching condition for its components ($C_1 = C_2Rg_m$) to achieve the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on January 05,2021 at 06:56:02 UTC from IEEE Xplore. Restrictions apply.

HP, BS and AP functions. The relative sensitivities of ω_0 and Q_0 for the proposed filter may be found as,

$$S_{g_m}^{\omega_0} = \frac{1}{2}; \quad S_{C_1}^{\omega_0} = S_{C_2}^{\omega_0} = S_R^{\omega_0} = -\frac{1}{2}; \tag{9}$$

$$S_{C_1}^{Q_0} = \frac{1}{2} - \frac{C_1}{C_1 + C_2 R g_m}; \; S_{g_m}^{Q_0} = S_{C_2}^{Q_0} = S_R^{Q_0} = \frac{1}{2} - \frac{C_2 R g_m}{C_1 + C_2 R g_m}$$

It is observed that the magnitude of relative sensitivities is either equal to one or less than one.

III. Analysis of Non-Ideal Behavior

In real-time scenarios, the overall performance and parameters of the MISO filter are altered by the influence of an error in voltage tracking of VDDDA block. In this section, we will study and analyse these non-ideal characteristics of VDDDA which can be represented as following:

$$V_w = \beta_z V_z - \beta_n V_n + \beta_p V_p \tag{10}$$

"where in Eq. (10), β_z, β_n and β_p represent the non-ideal voltage gain (or, voltage tracking error) of VDDDA from Z, N and P ports to W port, respectively and it is less than one in magnitude. For ideal case, is taken as unity gain." [4] The voltage filter response for the non-ideal case is represented by Eq (11):

$$V_0^* = \frac{s^2 V_{in3} + \frac{s\beta_z(C_1 V_{in2} + C_2 R g_m V_{in3})}{C_1 C_2 R(1+\beta_n)} + \frac{g_m \beta_z V_{in1}}{C_1 C_2 R(1+\beta_n)}}{D(s)} \tag{11}$$

where,

$$D^*(s) = s^2 + s\left(\frac{1}{C_2 R} + \frac{C_2 R g_m \beta_z - C_1 \beta_p}{C_1 C_2 R(1+\beta_n)}\right) + \frac{g_m \beta_z}{C_1 C_2 R(1+\beta_n)} \tag{12}$$

Note from Eq. (12) that the characteristic equation parameters (natural frequency and quality factor) are also affected by the voltage tracking error. Following are non-ideal estimations of $\omega_0$and Q_0:

$$\omega_0^* = \sqrt{\frac{g_m \beta_z}{C_1 C_2 R(1+\beta_n)}} \tag{13}$$

$$Q_0^* = \frac{1}{1 + \frac{C_2 R g_m \beta_z}{C_1(1+\beta_n)} - \frac{\beta_p}{(1+\beta_n)}} \sqrt{\frac{C_2 R g_m \beta_z}{C_1(1+\beta_n)}} \tag{14}$$

IV. Simulation Results

The functionality of CMOS implemented multifunction MISO filter is verified through SYMICA DE simulations with model library of 0.18 µm TSMC CMOS parameter file [8] at ±0.9V supply voltages. The CMOS structure of VDDDA is given in Fig. 3 and the transistor's sizes are illustrated in Table 1. Following are the values of the components used in proposed filter: $C_1 = C_2 = 30$ pF, R = 3.3 kΩ and the bias current I_B is set to 50 µA. Figs. (4-6) provide the gain functions of LP, HP and BP functions and Fig. 7 and 8 provide the gain and phase plots of BS and AP functions of the proposed MISO filter respectively. Due to the voltage tracking error or the non-ideal behavior of VDDDA, there is a deviation in the simulated natural frequency and the theoretical one. The simulated ω_0 = 1.106 MHz and the theoretical pole frequency from Eq. (7) is 1.131 MHz with the same value of passive components and bias current. Thus, it gives a deviation of 2.21% in natural frequency due to non-ideal behavior of VDDDA. The quality factor (Q_0) of filter is 0.707. The gain responses of BP function for different I_B are shown in Fig. 9, where $C_1 = C_2 = 30$ pF and I_B is equal to 50 µA, 100 µA & 150 µA and the corresponding values of ω_0 are 1.106 MHz, 1.272 MHz and 1.355 MHz respectively. Therefore, it is inferenced from Eqs. (7) and (8) that both ω_0 and Q_0 can be tuned electronically via g_m by changing I_B. Fig. 10 depicts the transient response for BP function. In this experiment, the proposed filter is fed with the input sinusoidal signal of 180 mV amplitude and frequency of 1 MHz.

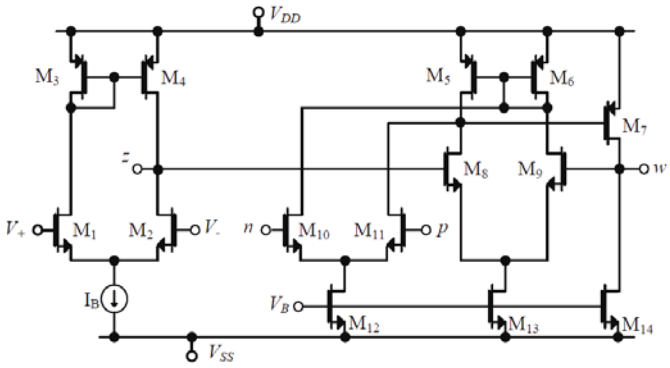


Fig. 3 VDDDA implemented by CMOS technology [4]

TABLE I
Aspect Ratios of Transistors [4]

MOS transistor	W/L (µm/µm)
M_1, M_2	9/1.08
M_3, M_4	3.96/1.08
M_5 - M_7	3.6/0.18
M_8 - M_{11}	0.72/1.08
M_{12} - M_{14}	2.16/1.08

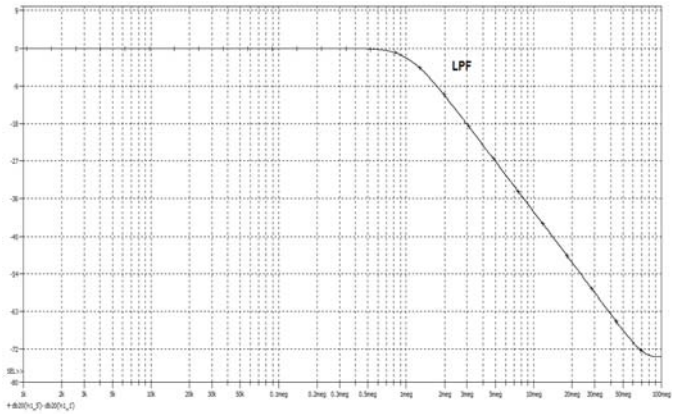


Fig. 4 Gain plot for LP

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on January 05,2021 at 06:56:02 UTC from IEEE Xplore. Restrictions apply.

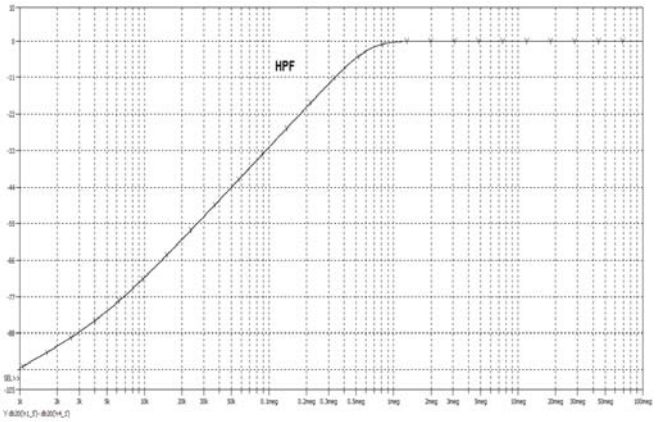


Fig. 5 Gain plot for HP

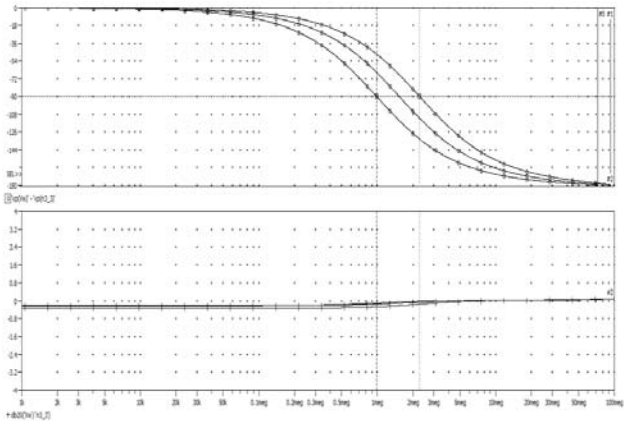


Fig. 8 Phase and gain plot for AP

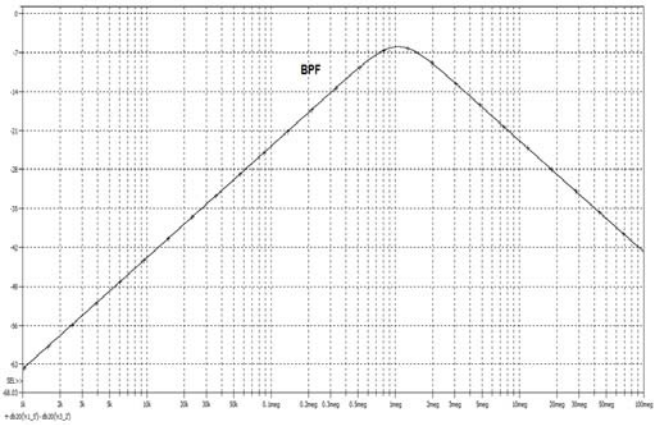


Fig. 6 Gain plot for BP

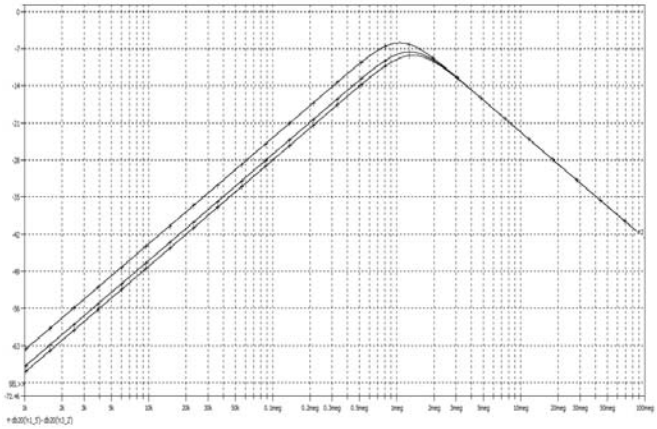


Fig. 9 BP plots for different values of I_B

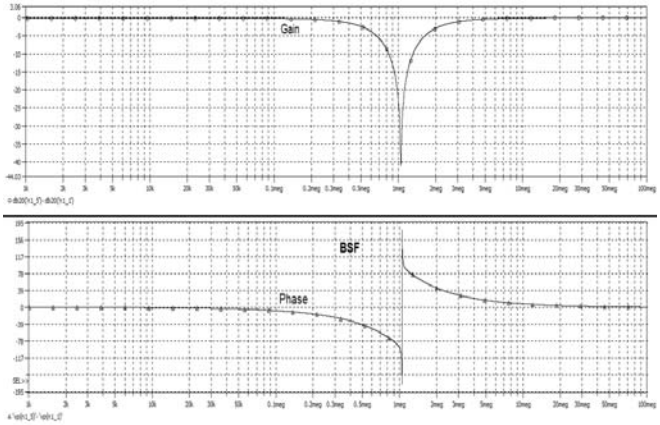


Fig. 7 Gain and phase plot for BS

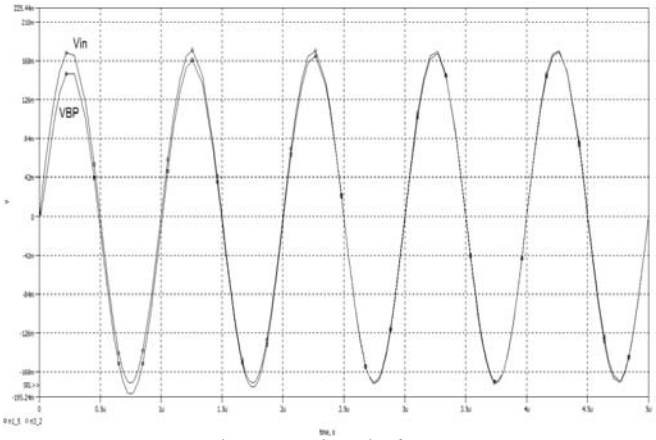


Fig. 10 Transient plot for BP

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on January 05,2021 at 06:56:02 UTC from IEEE Xplore. Restrictions apply.

V. Comparative Study

Many literature surveys have been conducted for voltage-mode multifunction filters using different active building blocks. A comparison has been done between the proposed MISO filter and the other referenced VM MISO filter as shown in Table 2. This table encapsulates the pros and cons of the previously designed filters with the proposed filter. From the table, for a second-order filtering function; we require at least two capacitors for quadratic equation. Filters in [9-11, 13, 15, 17, 18, 21-23] consist of floating passive elements in their circuit. Electronic tuning for the ω_0 and Q_0 can be achieved with filters [11, 13, 16-18, 20, 22-24]. All standard five filter responses can be realized in all surveyed filters. Those circuits that can't provide high input impedance are [10, 12, 13, 15, 17, 18, 20-23]. The filters that can provide low output impedance are [12, 14, 15, 21, and 24]. The mentioned voltage-mode MISO filters in [9, 10, 15, 18-20] require matching condition between passive elements. Also, the circuits in [9-11, 13, 18, and 23] need inverting input signals for filter responses.

TABLE II
Comparison of Different VM MISO Filters

Refs	ABB	No. of ABB	No. of R & C	Grounded passive elements	Electronic Tuning	Matching condition	High input impedance	Low output impedance	Inverting input
9	CCII	3	2+2	No	No	Yes	Yes	No	Yes
10	CFA	1	3+2	No	No	Yes	No	No	Yes
11	OTA & CCII	3	0+2	No	Yes	No	Yes	No	Yes
12	DDCC	2	2+2	Yes	No	No	No	Yes	No
13	OTA	2	0+2	No	Yes	No	No	No	Yes
14	DDCC	3	2+2	Yes	No	No	Yes	Yes	No
15	OTRA	1	4+4	No	No	Yes	No	Yes	No
16	OTA	6	0+2	Yes	Yes	No	Yes	No	No
17	CCTA	1	2+2	No	Yes	No	No	No	No
18	DVCC	1	2+2	No	Yes	Yes	No	No	Yes
19	DVCC	3	4+2	Yes	No	Yes	Yes	No	No
20	VDTA	1	0+2	Yes	Yes	Yes	No	No	No
21	CDBA	2	4+2	No	No	No	No	Yes	No
22	VD-DIBA	1	1+2	No	Yes	No	No	No	No
23	VDDDA	1	2+2	No	Yes	No	No	No	Yes
24	VDDDA	2	0+2	Yes	Yes	No	Yes	Yes	No
This work	VDDDA	1	1+2	No	Yes	Yes	LP	No	Yes

VI. Conclusions

In this proposal, voltage-mode biquad MISO filter has been proposed. The MISO filter proposes various advantages which are: First, the proposed filter can provide all five standard functions: Low-Pass (LP), High-Pass (HP), Band-Pass (BP), Band-Stop (BS) and All-Pass (AP) responses with digitally terminal switching of input ports, Secondly, the natural frequency and the quality factor both can be electronically controllable and tuneable. Thirdly, it consists of single VDDDA, two capacitors and one resistor which means it has least components count. Finally, it has high input impedance for LP response. Simulation results verified the theoretical concepts and validity of synthesis. However, in future with slight modification in the circuit, proposed filter can further be used in the implementation of voltage-mode quadrature oscillator, VDDDA based shadow filter and many other applications.

References

[1] A. S. Sedra and K. C. Smith, “Microelectronic circuits”, Florida: Holt, Rinehart and Winston, 1991.

[2] R. Raut and M. N. S. Swamy, “Modern Analog Filter Analysis and Design: A Practical Approach”, Copyright 2010 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

[3] M. E. Van Valkenburg, “Analog Filter Design”, Holt, Rinehart & Winston, 1982.

[4] N. Herensar, R. Sotner, B. Metin, J. Koton and K. Vrba, “VDDDA - New ‘voltage differencing’ device for analog signal processing,” Proceeding of the 8th International Conference on Electrical and Electronics Engineering, pp. 17-20, 2013.

[5] R. L. Geiger and E. Sanchez-Sinencio, “Active filter design using operational transconductance amplifiers: A tutorial”, IEEE Circuits Devices Mag., vol. 1, pp. 20–32, 1985.

[6] E. Sackinger and W. Guggenbiihl, “A versatile building block: CMOS differential difference amplifier”, IEEE J. Solid-State Circuits, vol. 22, pp. 287–294, 1987.

[7] W. Chiu, S.-I. Liu, H.-W. Tsao, and J.-J. Chen, “CMOS differential difference current conveyors and their applications”, IEE Proc. Circuits Devices Syst., vol. 143, pp. 91–96, 1996.

[8] TSMC 0.18 μm level-7 SCN018 CMOS technology parameters [online]: ftp://ftp.isi.edu/pub/mosis/ vendors/tsmc-018/t44e lo epi-params.txt.

[9] J. W. Horng. High-input impedance voltage-mode universal biquadratic filter using three plus-type CCIIs. IEEE Trans Circuits Syst II 2001; 48(10):996–7.

[10] J. W. Horng, C. K. Chang, J. M. Chu. Voltage-mode universal biquadratic filter using single current-feedback amplifier. IEICE Trans Fundam 2002; 85:1970–3.

[11] J. W. Horng. High input impedance voltage-mode universal biquadratic filter using two OTAs and one CCII. Int J Electron 2003; 90(3):153–7.

[12] C. M. Chang, H. P. Chen. Universal capacitor-grounded voltage-mode filter with three inputs and a single output. Int J Electron 2003; 90(6):401–6.

[13] J. W. Horng. Voltage-mode universal biquadratic filter using two OTAs and two capacitors. IEICE Trans Fundam 2003; 86A (2):411–3.

[14] J. W. Horng. High-input and low-output impedance voltage-mode universal biquadratic filter using DDCCs. IEEE Trans Circuits Syst II 2007; 54(8):649–52.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on January 05,2021 at 06:56:02 UTC from IEEE Xplore. Restrictions apply.

[15]S. Kılınc, A. Ü. Keskin, U. Cam. Cascadable voltage-mode multifunction biquad employing single OTRA. Frequenz 2007; 61:84–6.
[16]M. Kumngern, M. Somdunyakanok, P. Prommee. High-input impedance voltage-mode multifunction filter with three-Input single-output based on simple CMOS OTAs. Int Symp Commun Inf Technol 2008:426–31.
[17]N. Herencsar, J. Koton, K. Vrba. Single CCTA-based universal biquadratic filters employing minimum components. Int J Comput Electr Eng 2009;1(3):309–12.
[18]W. Tangsrirat, O. Channumsin. Voltage-mode multifunctional biquadratic filter using single DVCC and minimum number of passive elements. Indian J Pure Appl Phys 2011; 49:703–7.
[19]J. W. Horng, C. H. Hsu, C. Y. Tseng. High input impedance voltage-mode universal biquadratic filters with three inputs using three CCs and grounding capacitors. Radioengineering 2012; 21(1):290–6.
[20]J. Satansup, W. Tangsrirat. Single VDTA-based voltage-mode electronically tunable universal filter. In: The 27th International Technical Conference on Circuits Systems, Computers and Communications. 2012.
[21]J. K. Pathak, A. K. Singh, R. Senani. New voltage mode universal filters using only two CDBAs. ISRN Electron 2013 [Article ID 987867].
[22]K. L. Pushkar, D. R. Bhaskar, D. Prasad. A new MISO-type voltage-mode universal biquad using single VD-DIBA. ISRN Electron 2013 [Article ID 478213].
[23]A. Chaichana, S. Sangyaem, W. Jaikla. Multifunction voltage-mode filter using single voltage differencing differential difference amplifier. In: 3rd international conference on mechatronics and mechanical engineering; 2016.
[24]S. Sangyaem, S. Siripongdee, W. Jaikla, F. Khateb. Five-inputs single-output voltage mode universal filter with high input and low output impedance using VDDDAs. Optik 2017; 128:14–25.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on January 05,2021 at 06:56:02 UTC from IEEE Xplore. Restrictions apply.

Voltage and Frequency Control of Self Excited Induction Generator Integrated with PV System

Sombir
Department of Electrical Engineering
Delhi Technological University
Delhi, India
sskundu46@gmail.com

Madhusudan Singh
Department of Electrical Engineering
Delhi Technological University
Delhi, India
madhusudan@dce.ac.in

***Abstract*—this paper presents a VF Controller (VFC) for self excited induction generator (SEIG) integrated with PV system at DC link feeding linear/nonlinear load. It uses a capacitor bank to excite itself and provide SEIG with reactive power support. A proposed VF controller is consist of a 4-leg insulated gate bipolar transistor (IGBT)-based current controlled voltage source converter (CC-VSC), dump load (R_d) at its DC bus and a PV system support the frequency of the system by providing active power compensation for improving quality of voltage and current under changing load conditions. Despite changes in consumer load the VFC keeps the generated V and F constant. The CC-VSC with PV system fulfills the requirement of reactive and active power of integrated system. The entire system is designed in MATLAB and its performance is analyzed under different loading conditions.**

Keywords—seig, vfc, cc-vsc, pv array, mppt, boost convertor, ic method, pi controller

I. INTRODUCTION

The increasing use of non renewable fuel sources and their continued reduction over the experience of almost two decades about environmental pollution and global warming has led to uplift for renewable energy. Renewable Energy Sources (RES) can be used to produce electricity with less environment impact. The potential sources available for isolated locations include Pico hydro, wind and solar energy [1]. The wind energy conversion system (WECS) with self excited induction generator is becoming popular for electricity generation in isolated hilly areas, where conventional grid is not available. In operation of SEIG, the major problem is required reactive power compensation for self excitation and its voltage control. The self excitation of induction machine depends upon the suitable combination of speed, load and excitation capacitance [2-3]. The voltage regulation schemes based on shunt and series compensation, solid static switching devices and self commutating solid state switching convertors are extensively discussed in literature [4]. The shunt and series capacitors are used to improve voltage regulation under different loading conditions [5]. Due to their brushless construction with squirrel cage rotor, compact size, light weight, self short circuit safety and improved transient, etc., the SEIGs are gaining popularity in these applications compared with the other electric generators. But the key bottleneck in its operation is poor voltage and frequency control under changing load condition. The static compensator (STATCOM) fulfills the requirement of reactive power support, harmonic reduction and load leveling [6-10]. While improving voltage regulation of SEIG various configurations of VSC are discussed based on voltage regulation for a standalone power generation [11]. An Electronic Load controller (ELC) maintains the load power constant at SEIG terminal that constant the frequency of the system [12]. For remote rural locations, mostly loads are single phase type; hence star type SEIG system is more suitable as compared to the delta type distribution system. In such conditions a standalone SEIG integrated with PV system can feeds linear/dynamic loads. A SEIG system with suitable VFC provides neutral current compensation for single phase loads is an efficient and effective means for standalone applications. The VFC is used with CC-VSC to supply the reactive power required and to regulate voltage and frequency [13-15]. Various techniques for photovoltaic (PV) arrays with maximum power point tracking (MPPT) are also mentioned [16]. The VF regulation of the system allows the active power balance along with additional services such as reactive power support and reduction in voltage and current harmonics at PCC [18]. An integral controller is used to minimize the error signal generated by comparing instantaneous conductance to the incremental conductance [20]. In this paper, a VSC with PV system maintains the voltage and frequency of the SEIG by providing active and reactive power support and voltage and current power quality improves under linear/nonlinear load conditions in isolated applications.

II. SYSTEM TOPOLOGY

Fig.1 shows the components of the VFC integrated with PV system, with a capacitor bank, VF controller (consisting of CC-VSC, DC chopper and a PV array at DC link) and linear/nonlinear load at the consumer end. A star type capacitor bank of 3-phase is used for induction generators self-excitation purpose and for producing the no load rated voltage have to choose the excitation capacitor value. The DC bus chopper connected with dump load for the consumption of surplus power. PV system maintains constant active power of SEIG when the load demand increases. The PV system is to provide active power compensation when load demand increases. The generated voltage and frequency of SEIG will remain in limit during the change in load demand. The AC filter inductors are connected between the SEIG terminals and the output of the STATCOM. The VSC connected with DC linked capacitor for providing self supporting type DC bus filter voltage ripples. According to the changes in load demand, the surplus power is consumed by dump load.

Fig.2 represents regulation scheme of SEIG for controlling V and F integrated with PV system at DC link. The schemes control process is depend on reference source current generation has two components of AC voltage with in-phase and at quadrature axis. Controller in-phase unity templates have three phase sinusoidal functions are (u_a, u_b and u_c).

978-1-7281-5414-5/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 09:18:53 UTC from IEEE Xplore. Restrictions apply.

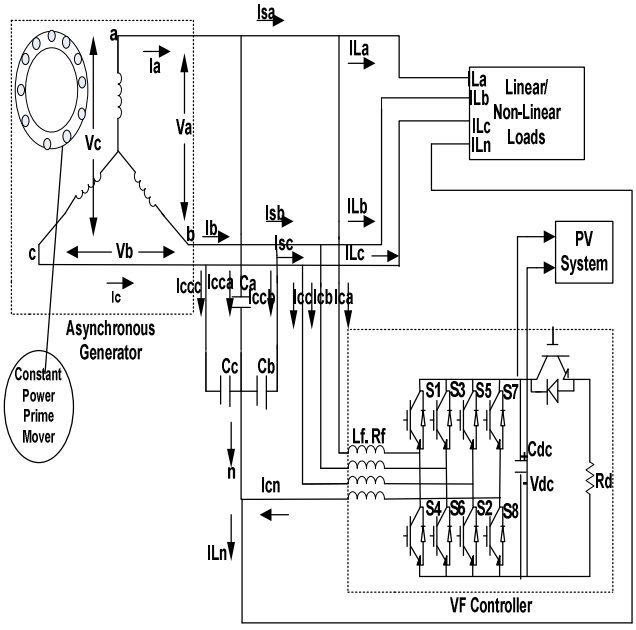


Fig.1 VFC with PV system for asynchronous generator feeding linear/nonlinear load

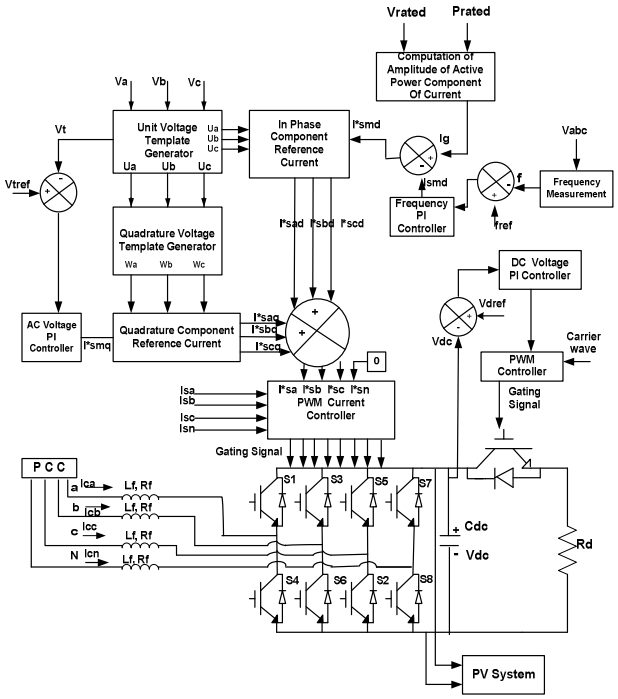


Fig. 2 Control diagram of VF controller integrated with PV system

The AC terminal voltage (V_t) of SEIG is sensed and compared with the reference voltage (V_{tref}). The voltage error is processed in the AC voltage PI controller. The output of the PI controller for AC voltage control loop decides the amplitude of reactive current (I^*_{smq}) which yields the quadrature component of the reference source currents (i^*_{saq}, i^*_{sbq} and i^*_{scq}). For constant power application, active power component of source currents are fixed at rated value, which is the amplitude of in phase component of source current (I^*_{smd}). Multiplication of in-phase unit templates (u_a, u_b and u_c) with in phase component of the reference source currents (I^*_{smd}) yields the in-phase component of the reference source currents (i^*_{sad}, i^*_{sbd} and i^*_{scd}). The sum of instantaneous quadrature and in-phase components of current is the reference source currents (i^*_{sa}, i^*_{sb} and i^*_{sc}), which are compared with the sensed source currents (i_{sa}, i_{sb} and i_{sc}). These current error signals are amplified and compared using PWM hysteresis controller for generating the PWM signal for switching of the devices of CC-VSC. Fourth leg of voltage source converter (VSC) is used to compensate the source neutral current (i_{sn}) and its reference value is taken zero (i^*_{sn}) through switching neutral connection. The excess generated power other than consumer load is dumped in the DC chopper dump load resistor (R_d) using DC bus voltage controller. The DC bus voltage of VSC of controller is compared with its reference voltage value. Then the error is processed in a PI Controller. The output of the PI controller is compared with triangular carrier wave to generate gating signal to DC chopper switch (IGBT) [13-15].

III. Mathematical Description of Control Scheme

The basic equations of proposed SEIG controller control scheme are given as:

A. CC-VSC Control scheme

The control scheme modeling for the 4-leg CC-VSC and DC chopper switch for VF control along with PV array is estimated as follows:

1) *Evaluation of in phase component of reference source current:* SEIG integrated with PV system will generate constant active power to feed either consumer load or dump load (R_d) for the constant active power generation, which is estimated as:

$$I^*_{smd} = \sqrt{2}\, P_{rated}/\sqrt{3}\, V_{rated} \quad (1)$$

Where P_{rated} is the rated power and V_{rated} is the rated voltage of SEIG and I^*_{smd} is peak value of rated active current of SEIG. At SEIG terminal the instantaneous line voltages (v_a, v_b and v_c) are assumed sinusoidal and corresponding maximum value is calculated as:

$$V_t^2 = \{(2/3)(v_a^2 + v_b^2 + v_c^2)\} \quad (2)$$

The unity in phase is derived from templates having instantaneous voltage as follows:

$$u_a = v_a/V_t\ ;\ u_b = v_b/V_t\ ;\ u_c = v_c/V_t \quad (3)$$

In-phase components in the form of Instantaneous values of reference source currents are calculated as:

$$i^*_{sad} = I^*_{smd} u_a\ ; i^*_{sbd} = I^*_{smd} u_b;\ i^*_{scd} = I^*_{smd} u_c \quad (4)$$

2) *Evaluation of quadrature component of reference source current:* The voltage error (v_{er}) at SEIG terminal at the n^{th} sampling instant calculated as:

$$v_{er(n)} = v_{tref(n)} - v_{t(n)} \quad (5)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 09:18:53 UTC from IEEE Xplore. Restrictions apply.

Where $v_{tref(n)}$ is the reference peak voltage and $v_{t(n)}$ is the peak value of the sensed 3-phase voltage at the SEIG terminals at the n[th] sampling instant is calculated as:

$$i^*_{smq(n)} = i^*_{smq(n-1)} + k_{pa}\{v_{er(n)} - v_{er(n-1)}\} + k_{ia(n)}v_{er(n)} \quad (6)$$

Where k_{pa} is the proportional gain and k_{ia} is the integral gain constants of the PI voltage controller. $v_{er(n)}$ is the voltage errors for n[th] instant and $v_{er(n-1)}$ is the voltage errors for (n-1)[th] instant and $I^*_{smq(n-1)}$ is the maximum value of quadrature component of the reference source current at (n-1)[th] instant. The quadrature components of the reference source currents for instantaneous values are calculated as:

$$i^*_{saq} = I^*_{smq}w_a;\ i^*_{sbq} = I^*_{smq}w_b;\ i^*_{scq} = I^*_{smq}w_c \quad (7)$$

Where (w_a, w_b and w_c) are the unit vectors having a 90^0 of phase change with respect to unit templates (u_a, u_b and u_c) are given below:

$$w_a = -u_b/\sqrt{3} + u_c/\sqrt{3} \quad (8)$$

$$w_b = \sqrt{3}\,u_a/2 + (u_b - u_c)/2\sqrt{3} \quad (9)$$

$$w_c = -\sqrt{3}\,u_a/2 + (u_b - u_c)/2\sqrt{3} \quad (10)$$

3) *Evaluation of total reference source current:* The total reference source currents are sum of the reference source current's in-phase and quadrature components calculated as:

$$i^*_{sa} = i^*_{saq} + i^*_{sad} \quad (11)$$

$$i^*_{sb} = i^*_{sbq} + i^*_{sbd} \quad (12)$$

$$i^*_{sc} = i^*_{scq} + i^*_{scd} \quad (13)$$

4) *Neutral current compensation:* VSC's Fourth leg is used to provide compensation for neutral current. The neutral current (i_{sn}) of the sensed source is compared to its reference current (i^*_{sn}) and the current error signal is amplified and compared using the hysteresis controller to produce the PWM signal to activate the fourth leg of VSC.

$$i^*_{sn} = 0 \quad (14)$$

$$i_{sa} + i_{sb} + i_{sc} = i_{ln} = (-)i_{cn} \quad (15)$$

$$i_{snerr} = i^*_{sn} - i_{sn} \quad (16)$$

5) *PWM current controller for 3-leg VSC:* Comparison of the reference source currents (i^*_{sa}, i^*_{sb} *and* i^*_{sc}) with the sensed source currents (i_{sa}, i_{sb} *and* i_{sc}). The PWM current controllers generate the ON / OFF switching patterns for the gate drive signals to the IGBTs. The current error is calculated as:

$$i^*_{saerr} = i^*_{sa} - i_{sa} \quad (17)$$

$$i^*_{sberr} = i^*_{sb} - i_{sb} \quad (18)$$

$$i^*_{scerr} = i^*_{sc} - i_{sc} \quad (19)$$

The amplified current error signals are compared with triangular carrier wave. If the amplified current error signal corresponding to phase A (i_{saerr}) is greater than the triangular wave signal switch S4 is ON and the switch S1 is OFF and the switch feature SA is set to 0. If the corresponding amplified current error signal (i_{saerr}) is equal to or less than the triangular wave signal switch S1 and switch S4 is OFF, and the value of SA is set to similar logic applied to other phases.

B. Control of DC Bus Chopper circuit

In order to maintain the DC bus voltage constant of the VSC constant, the error $v_{dcer(n)}$ at n[th] sampling instant estimated as:

$$v_{dcer(n)} = v_{dcref(n)} - v_{dc(n)} \quad (20)$$

Where $v_{dcref(n)}$ is reference DC voltage and $v_{dc(n)}$ is the sensed DC link voltage of the CC-VSC. The PI voltage controller output for maintaining constant DC bus voltage at n[th] sampling instant is estimated as:

$$v^*_{con(n)} = v^*_{con(n-1)} + k_{pp}\{v_{dcer(n)} - v_{dcer(n-1)}\} + k_{pi}v_{dcer(n)} \quad (21)$$

Where k_{pp} is the proportional gain constant and k_{pi} is the integral gain constant of the DC bus voltage PI controller.

$$\text{When } v^*_{con(n)} > v_{tri} \quad \text{SD=1 and}$$

$$\text{When } v^*_{con(n)} < v_{tri}, \quad \text{SD=0} \quad (22)$$

Where SD is the switching function which is used for generating the gating pulse of the IGBT based chopper switch [13-15].

C. Control of PV system

The PV system is connected at DC link for the implementation of maximum power point tracking (MPPT). The PV array block has two inputs that are used to varying sun irradiance (input 1 in W/m^2) and temperature (input 2 in deg. C). The switching duty cycle is optimized by an MPPT controller which uses the technique of incremental conductance and integral regulation. The integral regulator minimizes the error. To generate required voltage, MPPT system is used to vary the duty cycle automatically.

Fig.3 shows the Simulink design of PV system with an incremental conductance MPPT method. Incremental conductance was modeled on the basis of observing curve P-V characteristics. This algorithm was invented in 1993 [19]. The MPP can be expressed by using the relation between dI/dV and –I/V. The voltage depends on the photovoltaic current, then on the power P = I(V)×I.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 09:18:53 UTC from IEEE Xplore. Restrictions apply.

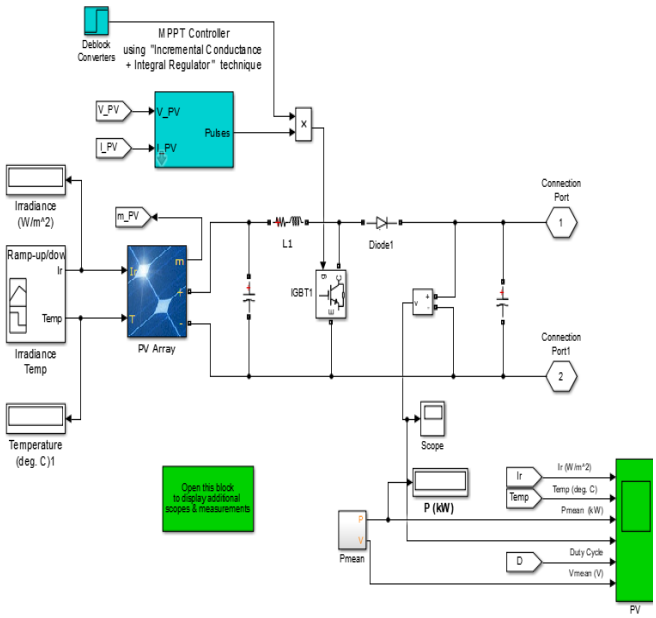


Fig.3. Simulink diagram of PV system with an IC MPPT technique

$$\text{Hence, } dP/dV = V \times (dI/dV) + I(V). \qquad (23)$$

Where, dP/dV represents the incremental conductance and -I/V represents instantaneous conductance.

When, dP/dV=0, then, dI/dV=-I(V)/V. Thus, the IC is equivalent to the negative of the instantaneous conductance. The dI/dV= -I(V)/V voltage is known as MPP voltage. This voltage is controlled by the controller until the solar irradiation adjusts and the process repeats. The error signal (e) is estimated as:

$$e = I/V + dI/dV \qquad (24)$$

The residual steady state error is reduced by an integral controller. A proportional integral used to decrease signal error [20].

IV. MATLAB Based Modeling and Simulation

Fig.4 shows the complete MATLAB based model of SEIG with CC-VSC, DC chopper and PV system for linear and non linear 3-phase 4-wire load feeding. The model of VSC with Universal Bridge and AC output voltage of VSC connected with SEIG with the help of filter inductors. The DC bus chopper is modeled using an IGBT switch. Chopper control is carried out by means of the DC bus voltage PI controller and compared with the saw-tooth wave generator; the PI controller output generate a chopper switching signal for a varying duty cycle. For the simulation a 3.7kW, 400V, 50Hz, 4-pole, star connected asynchronous machine is connected in MATLAB/Simulink. A PV system of 3.7 KW, 650V, using 6 cells in series and 2 arrays in parallel are placed at DC link. Simulation is carried in 16a version of MATLAB and toolboxes simulink in discrete mode at step size 5e-6s.

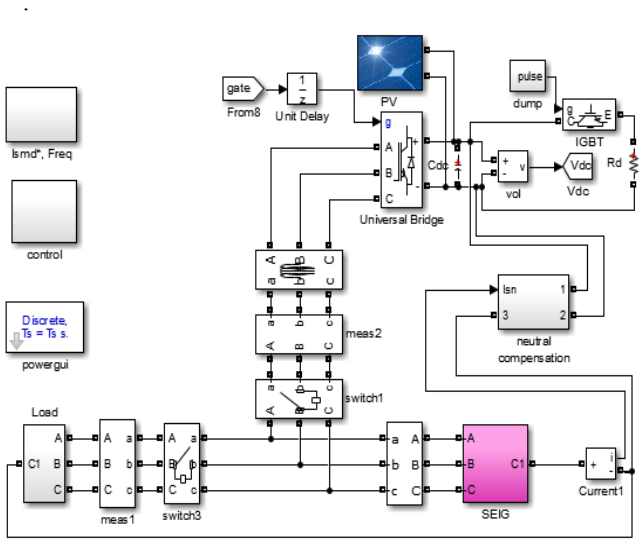


Fig.4. MATLAB/Simulink model of proposed electrical system

V. Results and Discussion

The model of proposed controller for SEIG with/without PV system feeding Linear and nonlinear load is designed and transient output of irradiance, temperature, voltage, power, duty cycle of boost convertor, source voltage (v_{abc}) and source current (i_{abc}), frequency (f), load voltage (v_L), load current (i_L), DC link voltage (v_{dc}), controller current (i_c), capacitive current (i_{cc}) and (I_{sn}) neutral current of the source are analyzed.

1) *Transient performance of SEIG at no load and on load:*

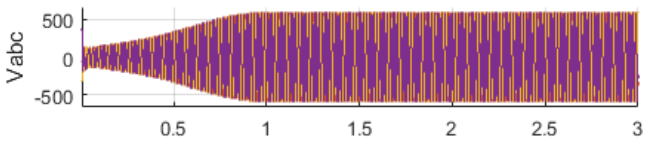


Fig.5 Voltage buildup of SEIG at no load

Fig.5 shows voltage buildup and the transient performance of SEIG at no load. For the voltage buildup of SEIG reactive power support is provided by the AC capacitor bank, which is connected across the stator terminal of SEIG. The voltage value of excitation capacitor bank is such that SEIG buildup its rated voltage 600V (peak) at no load.

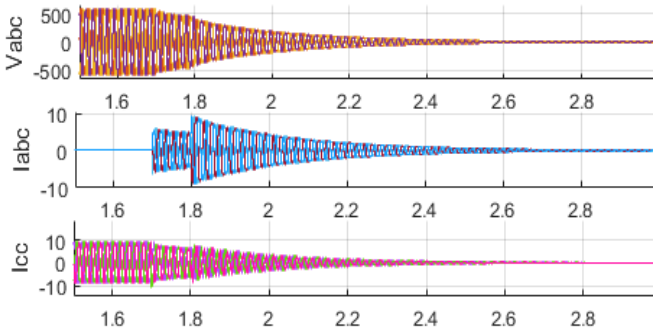


Fig.6 Transient performance of SEIG on load

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 09:18:53 UTC from IEEE Xplore. Restrictions apply.

Fig.6 above shows the transient output of the system on linear/nonlinear load without STATCOM. A linear load of 3KW resistive is applied at t=1.7 sec, then generator voltage (v_{abc}) and capacitive current (i_{cc}) decreases and a balanced nonlinear load of 2.5kW resistive and 6mH inductive at DC end of the 3-phase diode rectifier is applied at t=1.8 sec. Then the generator voltage (v_{abc}) and capacitive current (i_{cc}) collapse.

2) *Performance of SEIG with Linear/Nonlinear load with STATCOM:* Fig.7 demonstrates SEIG 's performance with linear/nonlinear load with STATCOM. A balanced linear load of 3kW resistive is applied at t= 1.7 sec. After this v_{abc}, v_L, I_c, v_{dc} and freq., are remains constant. At t=1.8 sec, a balanced nonlinear load of 2.5kW resistive and 6mH inductive at DC end of the 3-phase diode rectifier is applied. After sudden switching of nonlinear load at t=1.8 sec, v_{abc}, v_L, I_c, v_{dc} and freq., are reduced.

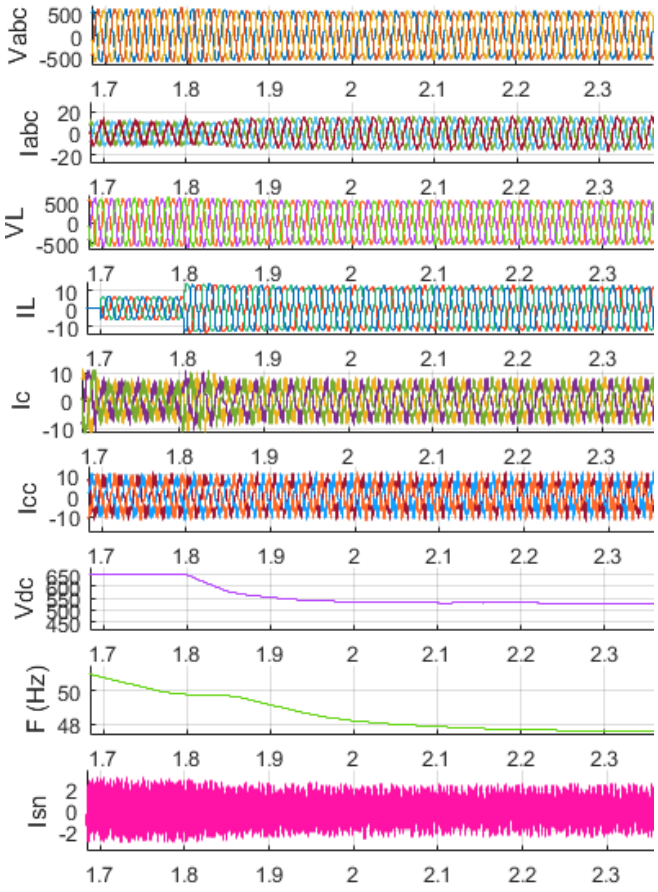


Fig.7. SEIG output with linear/nonlinear load with STATCOM

3) *Performance of SEIG with Linear/Non-linear load at fixed irradiance*: fig.8 shows the output of the SEIG with linear/nonlinear load with fixed sun irradiance of PV system. At t= 1.7 sec, a balanced linear load of 3kW resistive is applied. At t=1.8 sec, a balanced nonlinear load of 2.5kW resistive and 6mH inductive at DC end of the 3-phase diode rectifier is applied. The variation in irradiance (1000 W/m^2) and temperature (25 deg. C) depends upon the PV system inputs. The result shows the performance of PV system and SEIG with fixed solar irradiance. PV system provide active power compensation to SEIG when load demand increase. SEIG with PV system maintains v_{abc}, v_L, freq., v_{dc} constant. The simulated result shows that voltage and frequency remains within the limit during the load changes.

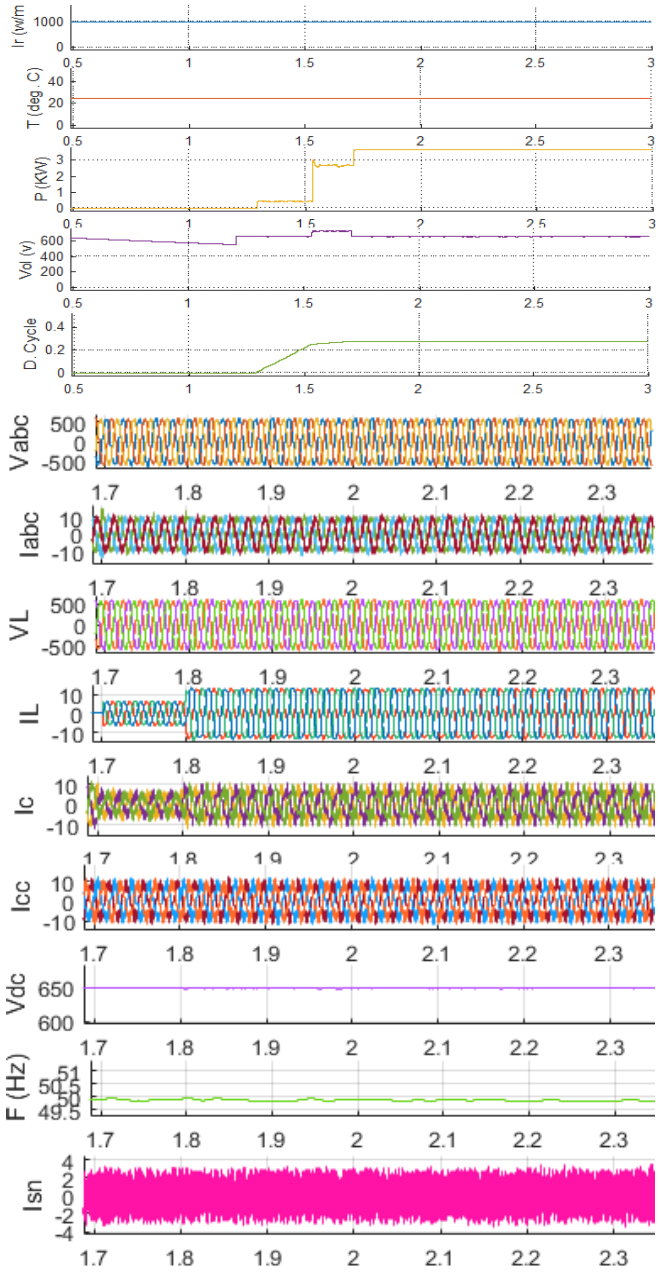


Fig.8 SEIG output with linear/nonlinear load at fixed solar irradiance

4) *Performance of SEIG with Linear/Non-linear load under varying irradiance:* in fig.9 the performance of SEIG under change in solar irradiance from 1000 w/m^2 to 700 w/m^2 is shown. In case of decrease in the irradiance, the output power of PV system is reduced. At t= 1.7 sec, a balanced linear load of 3kW resistive is applied. At t=1.8 sec, a balanced nonlinear load of 2.5kW resistive and 6mH inductive at DC end of the 3-phase diode rectifier is applied. The reduction in output power of PV system affect the performance of SEIG. The simulated result shows that v_{abc}, v_L, v_{dc} and frequency are reduced.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 09:18:53 UTC from IEEE Xplore. Restrictions apply.

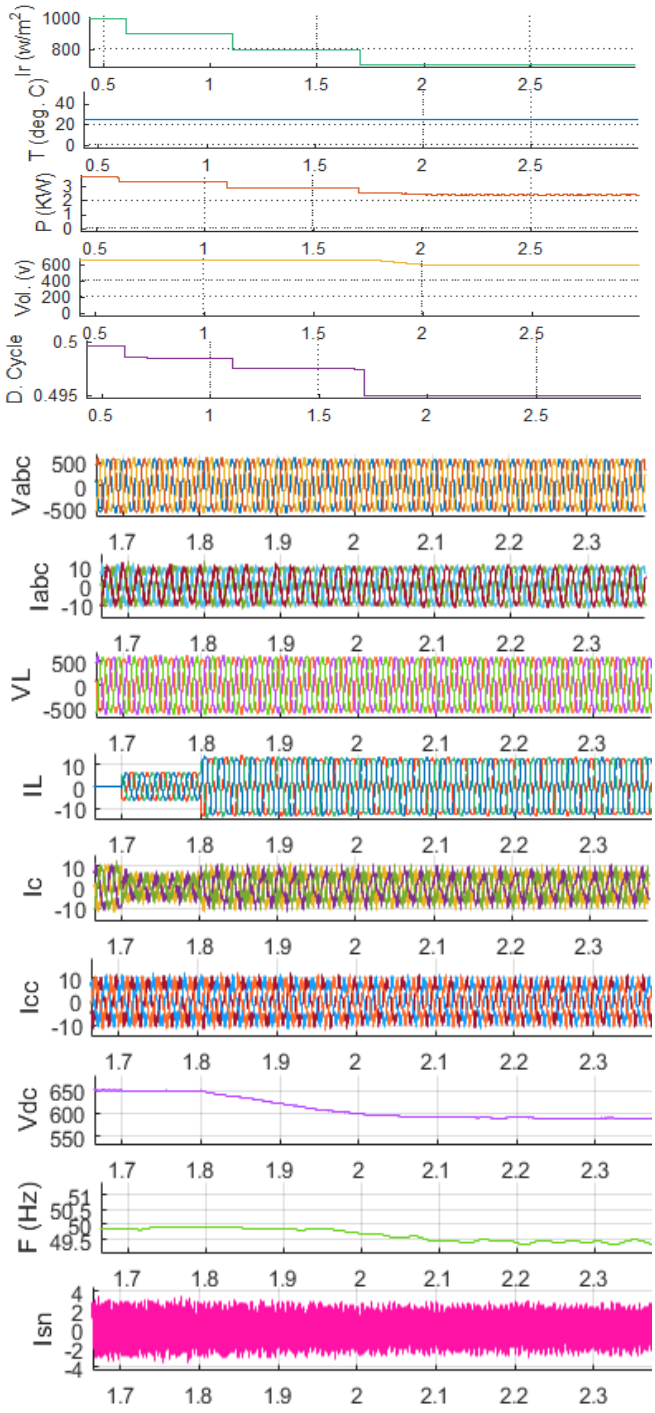


Fig.9. Performance of SEIG with linear/nonlinear load under change in solar irradiance

VI. CONCLUSION

An efficient VF control scheme is developed and designed in MATLAB. The various results above shows that a VF controller for SEIG integrated with PV system are able to regulate voltage and frequency. The voltage and frequency of the SEIG is regulated using CC-VSC and PV system by means of the reactive and real power support. The simulated result shows the effectiveness of the controller in maintaining the voltage and frequency of SEIG under different loading conditions. The introduced hybrid energy source system can operate in standalone or grid connected to increase the flexibility of the energy sources. The VFC should design along with MPPT to provide continuous support of reactive and real power for its constant power operation.

REFERENCES

[1] R.C. Bansal, "Three phase self excited induction generators: an overview" IEEE Trans. On Energy Conversion, vol 20, no. 2, pp. 292-299, June 2005

[2] E.D. Bassett and F.M. Potter, "Capacitive Excitation for Induction Generators" ,AIEE Transactions on Electrical Engineering, vol. 54, pp. 540-545, May 235

[3] C. Chakraborty, S. N Bhadra & A. K. Chattopadhyay, "Excitation requirements for stand alone three-phase induction generator" IEEE Transactions on Energy Conversion, 13(4), 358–365. https://doi.org/10.1109/60.736321

[4] Yogesh K. Chauhan,"A Prospective on Voltage Regulation of Self Excited Induction Generators for Industry Applications" IEEE trans. On industry applications, vol. 46, no. 2, March/April 2010.

[5] Bhim Singh, Madhusudan, Vishal Verma and A.K. Tandon " Rating Reduction of Static Compensator for Voltage Control of Three-Phase Self-Excited Induction Generator" IEEE ISIE 2006, July 9-12, 2006, Montreal, Quebec, Canada

[6] B. Singh, S.S. Murthy and Sushma Gupta "Modelling and Analysis of STATCOM Based Voltage Regulator for Self Excited Induction Generator with Unbalanced Loads" IEEE-2003, 0-7803-7651-x/03

[7] Woei-Luen Chen, Yung-Hsiang Lin, Hrong-Sheng Gau, and Chia-Hung Yu "Statcom Controls for a Self Excited Induction Generator Feeding Random Loads" IEEE Transactions on power delivery, vol. 23, no. 4, October 2008

[8] B. Singh, S.S. Murthy and Sushma Gupta "STATCOM Based Voltage Regulator for Self Excited Induction Generator Feeding Non-Linear Loads" 0-7803-7906-3/03 2003 IEEE

[9] Gaurav Kumar Kasal and Bhim Singh "Harmonic Elimination, Voltage control and Load Balancing in an Isolated Power Generation" Euro. Trans. Electr. Power 2010 published online 22 june 2009 in Wiley Online Library

[10] S.C. Kuo and L.Wang "Analysis of Voltage control for a Self Excited Induction Generator Using a Current Controlled Voltage source Inverter (CC-VSI)" IEEE Proc.-Gener. Trans. Distrib., vol. 148. No. 5, September 2001

[11] G.K. Kasal and Bhim Singh "Design and Control of Voltage Regulators for a Standalone Power Generation" IETE Technical Review, 26:2, 115-136

[12] Bhim Singh and G.K. Kasal "Analysis and Design of Voltage and Frequency Controllers for Isolated Asynchronous Generators in Constant Power Applications" IEEE-2006, 0-7803-9771-1-06

[13] B. Singh, & G. K. Kasal, "Voltage and frequency controller for isolated asynchronous generators feeding 3-phase 4-wire loads" Proceedings of the IEEE International Conference on Industrial Technology, 2773–2778.

[14] Bhim Singh, Gaurav Kumar Kasal, Ambrish Chandra and Kamal-Al-Haddad "A Frequency Based Electronic Load Controller for an Isolated Asynchronous Generator Feeding 3-Phase 4-Wire Loads" IEEE 2008, 978-1-4244-1666-0/08

[15] Bhim Singh, G.K. Kasal, Ambrish Chandra, Kamal-AL-Haddad "Voltage and Frequency Controller for an Autonomous Micro Hydro Generating System" IEEE-2008

[16] Trishan Esram, Patrick L. Chapman "Comparison of Photovoltaic Array Maximum Power Point Tracking Techniques" IEEE Trans. On Energy Conservation, Vol. 22, No. 2 June 2007

[17] Seema and bhim Singh "Analysis and Control of a standalone PV-BESS-DG Based Microgrid" IEEE-2016, 978-1-5090-4530-3/16

[18] Seema Kewat, Bhim Singh and Ikhlaq Hussain "Power management in PV-battery-hydro based standalone microgrid" IET Tenew. Power Gener.,2018, Vol. 12, Iss. 4, pp. 391-398

[19] M.Rifa and Ratnalka.P, "Pemodelandan Analisis Panel Photovoltaik",Conference Informatic, Telecommunications Electrical Engineering.UGM Yogyakarta, 2012.

[20] Md Hasan Anowar and provashish roy, "A modified incremental conductance based photovoltaic charge controller" IEEE 2019, 978-1-5386-9111-3/19

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 04,2020 at 09:18:53 UTC from IEEE Xplore. Restrictions apply.

VolTAGE: Volatility Forecasting via Text Audio Fusion with Graph Convolution Networks for Earnings Calls

Ramit Sawhney
Netaji Subhas Institute of Technology
ramits.co@nsit.net.in

Piyush Khanna
Delhi Technological University
piyushkhanna_bt2k17@dtu.ac.in

Arshiya Aggarwal
MIDAS, IIIT Delhi
arshiya.dtu@gmail.com

Taru Jain
MIDAS, IIIT Delhi
jaintaru@ieee.org

Puneet Mathur
University of Maryland, College Park
puneetm@cs.umd.edu

Rajiv Ratn Shah
IIIT Delhi
rajivratn@iiitd.ac.in

Abstract

Natural language processing has recently made stock movement forecasting and volatility forecasting advances, leading to improved financial forecasting. Transcripts of companies' earnings calls are well studied for risk modeling, offering unique investment insight into stock performance. However, vocal cues in the speech of company executives present an underexplored rich source of natural language data for estimating financial risk. Additionally, most existing approaches ignore the correlations between stocks. Building on existing work, we introduce a neural model for stock volatility prediction that accounts for stock interdependence via graph convolutions while fusing verbal, vocal, and financial features in a semi-supervised multi-task risk forecasting formulation. Our proposed model, VolTAGE, outperforms existing methods demonstrating the effectiveness of multimodal learning for volatility prediction.

1 Introduction

Motivation Financial risk modeling is of great interest to capital market participants for making sound investment decisions. Stock volatility is a vital indicator of a company's risk profile (Poon and Granger, 2003; Yang et al., 2020). The stock market presents various opportunities that increasingly attract investors, who utilize the market's potential to generate profits, wherein stock volatility is a vital risk modeling factor. One underexplored, yet crucial event that leads to significant fluctuations in stock volatility, is the earnings conference call. These calls are held periodically by publicly traded companies' executives to summarize and prognosticate company's performance (Qin and Yang, 2019). Harnessing the interplay between the multimodal verbal and vocal cues in earnings calls can help better analyze the impact these calls may have on financial markets and forecast stock volatility (Dichev and Tang, 2009; Yang et al., 2020).

Challenges While stock trading presents unparalleled investment opportunities, accurately predicting the rise and fall of stock prices has numerous challenges (Campbell et al., 1997). Conventional research in finance revolves around using historical stock data to develop statistical models and recurrent neural networks (RNNs) capable of forecasting price trends (Kristjanpoller et al., 2014; Zheng et al., 2019). They are influenced by many factors ranging from public opinion to the movements of other related stocks (Malkiel, 2003). Recent advances in deep learning present a promising prospect in multimodal stock forecasting by analyzing online news (Hu et al., 2018), and social media (Guo et al., 2018) to learn latent patterns affecting stock prices (Jiang, 2020). However, the challenging aspect in stock forecasting is that most existing work treats stock movements to be independent of each other, contrary to true market function (Diebold and Yılmaz, 2014). Additionally, existing research has not leveraged the rich audio signals in company executives' speech, which could indicate the emotional and affective state of the speakers, and provide insights into company performance. More recently, the use of audio processing for earnings calls has gained an interest in both financial and linguistic research (Burgoon et al., 2015; Jiang and Pell, 2017).

Multimodal approaches can extract complementary information from multiple modalities to improve financial modeling, MDRM (Qin and Yang, 2019), and HTML (Yang et al., 2020) validate the premise of such approaches for volatility forecasting. Additionally, advances in graph-based deep learning (Kipf and Welling, 2017) have led to the rise of graph neural networks (GNNs) that can model the relationships between related stocks (Feng et al., 2019). Publicly available online company information can be used to identify connections between stocks that might influence each other, such as those having the same CEO or belonging to the same industry. Financial tasks are often correlated, thus making multi-task learning a promising choice for financial forecasting.

Contributions Building on advances in the intersection of financial research, graph neural networks, and natural language processing, we present VolTAGE: **Vol**atility forecasting via **T**ext-**A**udio fusion with **G**raph convolution networks for **E**arnings calls. VolTAGE comprises a set of neural components to capture cross-modal signals from earnings calls transcripts, CEO speech, inter stock dependence graphs, and numerical financial features. First, VolTAGE combines the verbal-vocal coherence between earnings calls transcripts and speech via an inter-modal multi-utterance attention mechanism. The fused features are then fed to a graph convolution network (GCN) to simultaneously solve two homogeneous stock volatility tasks - average volatility (main task) and single-day volatility prediction (auxiliary task), in a semi-supervised fashion. Through a set of comparative, qualitative, and ablation experiments on real-world S&P 500 index data, we show VolTAGE's utility of augmenting vocal and verbal cues with graph-based features in a multi-task setup.

Ethical Considerations and Limitations Examining a CEO's speech and tone in earnings calls is a well-studied phenomenon in financial literature (Crawford Camiciottoli, 2011; Qin and Yang, 2019). Our work focuses only on calls for which companies publicly release transcripts and audio recordings. The data used in our study corresponds to earnings calls of S&P 500 companies. We acknowledge the presence of gender bias in our study, given the imbalance in the gender ratio of CEOs of S&P 500 companies. We also acknowledge the demographic bias in our study, as the S&P 500 companies are organizations listed in the US, and may not generalize directly to non-native speakers.

2 Background

Extensive studies have shown the utility of employing historical financial data (Jones, 2017; Dichev and Tang, 2009) for volatility prediction, yet financial forecasting using multiple modalities remains an underexplored avenue. While newer work focuses on data across multiple modalities, there exist drawbacks and understudied approaches to improve current methods, which we describe next.

Volatility Forecasting Forecasting stock volatility is a crucial pillar across multiple financial domains and has focused on numerous academic studies. Volatility is a key indicator of uncertainty and is a decisive variable to many investment decisions and portfolio creations. Previous work in this domain has mainly relied on numerical features (Liu and Chen, 2019; Nikou et al., 2019), such as macroeconomic indicators (Hoseinzade et al., 2019). This includes discrete (GARCH (Duan, 1995), rolling regression (Peng et al., 2018)), continuous (Andersen, 2007), and neural approaches (Kogan et al., 2009). This comprehensive work illustrates the significance of volatility in investment, security valuation, and risk management.

Natural Language Processing and Finance Extensive studies incorporating related text information have proven successful in financial forecasting tasks. Mohan et al. (2019); Tan et al. (2019) utilized financial news articles to improve the accuracy of stock price predictions. Hu et al. (2018) propose a hybrid attention network to predict the stock trend based on the related sequential news articles. Researchers have also observed the influence of textual data in online media on stock markets (Bollen et al., 2011; Mittermayer and Knolmayer, 2006). Si et al. (2014) showed sentiment analysis based on social media is predictive of each stock's market. However, utilizing multimodal sources of information remains an underexplored avenue in financial forecasting.

Speech Processing and Finance Newer studies (Qin and Yang, 2019; Yang et al., 2020) illustrate the gains obtained using vocal cues from the CEO's earnings conference calls for volatility prediction. Yet, the majority of the current work does not utilize speech based data. The audio features add

greater context and provide psycho-linguistic signaling about the speaker's emotional state (Jiang and Pell, 2017). Qin and Yang (2019) illustrated that late fusion of audio and text features from earnings calls could be used to forecast stock volatility following the earnings call. The verbose quarterly earnings calls (Wang and Hua, 2014) act as a medium of voluntary disclosure (Tasker, 1998), thereby resulting in significant stock movements (Ding et al., 2015), yet the majority of existing approaches do not focus on such highly volatile macro activities, where the market microstructure is highly uncertain (Rogers et al., 2009). During these macro events, the stock returns' predictability can be improved since the disclosure of informed investors influences volatility spreads (Atilgan, 2014). Although multiple sources of information are crucial, not all modalities contribute equally (Akhtar et al., 2019). Noise in one modality can be detrimental in such multimodal frameworks (Morris-Drake et al., 2016).

Multimodality and Finance The Efficient Market Hypothesis (Malkiel, 2003) illustrates the success of multimodal data sources for predictive financial tasks. The more recent multimodal HTML (Yang et al., 2020) is a transformer-based model that uses BERT (Devlin et al., 2019) for textual modeling, and the same hand-crafted audio features as MDRM, in an early fusion formulation. Both MDRM and HTML assume stocks' independence and do not exploit these relations between stock movements. Relations like the same industrial base and co-ownership also result in related stock movements (Feng et al., 2019). Recent works exploit stock relations through graph neural networks (Kipf and Welling, 2017; Veličković et al., 2018) for stock movement prediction (Kim et al., 2019; Sawhney et al., 2020).

Building on these gaps in existing literature, we propose VolTAGE for volatility prediction.

3 Forecasting Stock Volatility

Following Kogan et al. (2009) and Qin and Yang (2019) we define stock volatility as a regression task. For a given stock, with a close price of p_i on trading day i, we calculate the average log volatility over n days following the earnings call as:

$$v_{[0,n]} = \ln\left(\sqrt{\frac{\sum_{i=1}^{n}(r_i - \bar{r})^2}{n}}\right) \quad (1)$$

where, the return price r_i is defined as $\frac{p_i}{p_{i-1}} - 1$ and $\bar{r}$ is the mean of the return price over the period from day-0 to day-n. Additionally, for our auxiliary task we define the single-day log volatility using the daily log absolute returns as follows:

$$v_n = \ln\left(\left|\frac{p_n - p_{n-1}}{p_{n-1}}\right|\right) \quad (2)$$

Problem Statement Given an earning call e, comprising of an audio A, and aligned text T, and stock prices $p_{[0,n]}$, we aim to learn a predictive regression function $f(e_{\{T,A\}}) \rightarrow v_{[0,n]}$.

4 VolTAGE: Architecture and Learning

Below, we describe both the individual components and joint optimization of VolTAGE, and present an overview of the architecture in Figure 1.

4.1 Verbal Cues: Transcript Encoding

We use FinBERT[1] (Araci, 2019) as a sentence encoder, which is a pre-trained language model based on BERT, for language modeling specific to the financial domain. Recent works (Araci, 2019; Keith and Stent, 2019) in this domain indicate the benefits of using a language model pre-trained on financial corpora and retrofitting pre-computed embeddings, achieving considerable performance gains; thereby giving us a strong ground to incorporate the same. FinBERT has been pre-trained on 46,000 documents of financial news articles and has shown state-of-the-art performance on FiQA[2] and Financial PhraseBank benchmarks (Malo et al., 2013).

Formally, we represent the transcript utterances of each call as $(t_1, t_2, ..., t_K)$, where t_i is the i^{th} text utterance and K is the number of sentences, which are encoded as follows:

$$s_i = \text{FinBERT}(t_i) \quad (3)$$

We then pass the sequence of these sentence representations to a BiLSTM as:

$$\overrightarrow{T_t^{(f)}} = \text{BiLSTM}^{(f)}(s_t, T_{t-1}^{(f)}) \quad (4)$$

$$\overleftarrow{T_t^{(b)}} = \text{BiLSTM}^{(b)}(s_t, T_{t+1}^{(b)}) \quad (5)$$

$$T_t = [\overrightarrow{T_t^{(f)}}, \overleftarrow{T_{T-t}^{(b)}}] \quad (6)$$

[1] https://github.com/abhijeet3922/finbert_embedding

[2] https://sites.google.com/view/fiqa

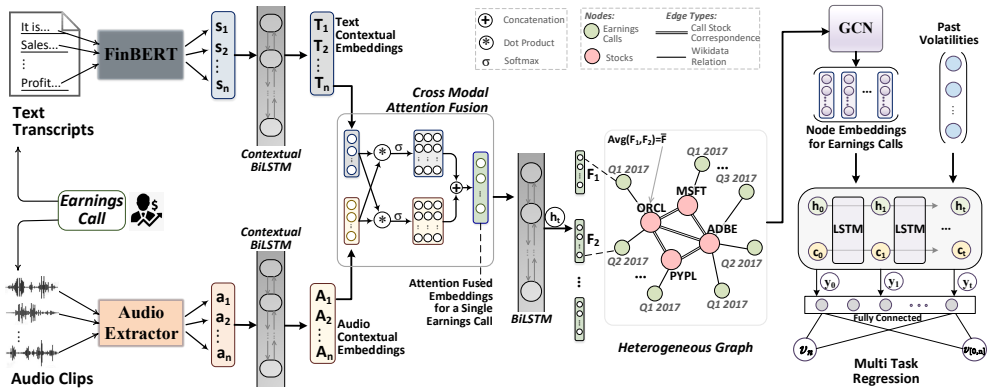


Figure 1: VolTAGE architecture overview: feature extraction, semi-supervised learning and multi-task regression.

4.2 Vocal Cues: Audio Call Encoding

Audio-based features provide prosodic cues related to the affective state of speakers (Montacié and Caraty, 2018). Capturing the emotional valence of the CEO can alter the understanding of the underlying linguistic utterances in an earnings call (Schröder et al., 2001). We extract a set of 26 acoustic features from each aligned audio clip at a sampling rate of $10ms$ for each sentence. These feature time series were then summarized by statistical functions such as mean, median, min, and max to yield a fixed dimensional representation for each sentence. We extend the feature sets of previous works (Qin and Yang, 2019; Yang et al., 2020). These features have shown to be correlated to the speaker's affective states such as stress and anxiety (APQ 11 Shimmer, DDA Shimmer) (Li et al., 2007; Mongia and Sharma, 2014), vocal pace reflecting inconsistencies in vocal cues (ratio of voiced to unvoiced frames in audio) (Přibil and Přibilová, 2009; Viswanathan et al., 2012) and deception (pitch) (Burgoon et al., 2015). We extracted these 26 features from each audio utterance using Praat (Boersma and Van Heuven, 2001).

Text-Audio Alignment Following Qin and Yang (2019), we use the pre-aligned dataset for earnings calls, where the audio is segmented and aligned with each corresponding utterance of the transcript using the Iterative Forced Alignment (IFA) algorithm. IFA is the process of determining the time interval (in the audio file) containing the spoken text for each fragment of the transcript. Qin and Yang (2019) implemented IFA using Aeneas[3] as the fundamental forced alignment method. Formally, we represent the segmented audio clips as $(a_1, a_2, ..., a_K)$ where a_i is the i^{th} audio clip and K being the number of clips of an earning call, with each clip being represented by 26 acoustic features. Similar to the processing of verbal utterances, we employ a BiLSTM layer to sequentially encodes these features, and obtain an audio encoding A_t as:

$$\overrightarrow{A_t^{(f)}} = \text{BiLSTM}^{(f)}(a_t, A_{t-1}^{(f)}) \quad (7)$$

$$\overleftarrow{A_t^{(b)}} = \text{BiLSTM}^{(b)}(a_t, A_{t+1}^{(b)}) \quad (8)$$

$$A_t = [\overrightarrow{A_t^{(f)}}, \overleftarrow{A_{T-t}^{(b)}}] \quad (9)$$

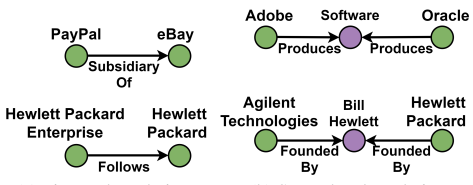
4.3 Verbal-Vocal Attention

The acoustic features provide context and structure to the verbal cues. To capture the associations between verbal and vocal cues, we employ a Cross-Modal Gated Attention Fusion (CM Attn) mechanism that simultaneously learns the alignment weights between audio features and text sentence sequences. Thus, we employ this mechanism to highlight the contributing features by giving more attention to the respective utterance and neighboring utterances. Motivated by Akhtar et al. (2019); Dhingra et al. (2016), we employ the multiplicative gated attention mechanism to generate modality-specific attentive representations.

Formally, a multiplicative gating mechanism is used to attend the important components of text and audio sequences to get the final attentive feature embeddings F_t, F_a which are then combined as:

$$W_a = softmax(T \cdot A^T), W_t = softmax(A \cdot T^T) \quad (10)$$

[3] https://github.com/readbeyond/aeneas



(a) First-order relation (b) Second-order relation

Figure 2: Wiki-company based relations

$$G_a = W_a \cdot T,\ G_t = W_t \cdot A \quad (11)$$

$$H = F_a \oplus F_t = (G_a \odot A) \oplus (G_t \odot T) \quad (12)$$

$\cdot$ is the dot product, $\odot$ represents element-wise multiplication, and $\oplus$ represents concatenation. The fused verbal-vocal feature vector per earnings call is then fed to a GCN, as described next.

4.4 Graph-based Semi Supervised Learning

Mining Stock Relations First, we construct a company graph, inspired by the relations defined by Feng et al. (2019). We mine connections between companies from Wikidata (Vrandečić and Krötzsch, 2014). Wikidata represents relations in the form of statements like (subject; predicate ;object), such as (Facebook; founded by; Mark Zuckerberg).[4] We say that Company A has a first-order relation with company B if there is a statement with A as the subject and B as the object. Similarly, there exists a second-order relation between them if they are related by an intermediate entity. This Wiki-Company graph $G_{WC} = (V, E_{WC})$ is a homogeneous graph, where each node represents a company, and two nodes are connected by an edge representing either a first or second-order relation. We present examples of first and second-order relations in Figure 2. Since the companies are related and not the earnings calls, we extend the graph G_{WC} by incorporating nodes corresponding to earnings calls. Each call is connected to the company it corresponds to through an edge. This extended graph $G(V, E)$ is heterogeneous with two types of nodes (companies and earnings calls).

Graph Convolution Network We frame the task as a graph-based semi-supervised learning problem since we have labels (volatility values) available for a subset of nodes (i.e., earnings call nodes) (Kipf and Welling, 2017). Our intuition behind applying GCNs is to allow the model to distribute gradient information from the supervised loss on the labeled earnings call nodes. As shown in Figure 1, we feed the fused verbal-vocal features H as node features for each earnings call node to the GCN. As for the stock nodes, since a stock may have multiple earnings calls, we consider the mean of feature vectors of all calls pertaining to a stock as its feature vector, to incorporate features across all earnings calls corresponding to that stock. Formally, let $F \in \mathbb{R}^{n \times m}$ represent the input feature matrix comprising these feature vectors of length m for the nodes in G, and D represent the diagonal degree matrix defined as $D_{ii} = \sum_j A_{ij}$. The update rule at layer l of the GCN is then:

$$O^{(l)} = ReLU(\widetilde{A}O^{(l-1)}W^{(l)}) \quad (13)$$

where the first layer is represented as:

$$O^{(1)} = ReLU(\widetilde{A}FW^{(1)}) \quad (14)$$

We experiment with a single layer and a 2-layer GCN, and find better results with the latter. We formulate the exact computation our GCN performs to yield estimated volatility values as follows:

$$O = linear(\widetilde{A}ReLU(\widetilde{A}FW^{(1)})W^{(2)}) \quad (15)$$

Using the earnings call node labels, we train the GCN on the MSE loss using the semi-supervised learning mechanism. This mechanism generates feature representations for both the company nodes and the earnings call nodes, of which we use the latter. Subsequently, these earnings call node features, denoted by O_e are fed along with the financial features to a multimodal LSTM network in a multi-task learning setup as described next.

4.5 Multimodal LSTM for Risk Forecasting

Prior work (Figlewski, 1994) in the financial domain has shown the benefits of using past data for future volatility forecasting. However, fusing the sequential historical volatility data with non-temporal GCN embeddings poses a challenge. To overcome this disparity, we employ multimodal "conditioned" LSTM networks (Karpathy and Fei-Fei, 2015). In our case, we add GCN node embeddings from the first layer with the ReLU non-linearity to the hidden state of the LSTM model at the first time-step to integrate temporally diverse

[4] https://www.mediawiki.org/wiki/Wikibase/DataModel/JSON

modalities. Further, the past data introduces historical context in cases where calls may not have major announcements that would lead to large fluctuations in stock volatility.

Network Optimization To incorporate financial data, we extract past n-day average volatilities prior to the earning call, where $n \in [2, 30]$. Formally, training the LSTM model takes the sequence input vectors $(x_1, .., x_T)$ representing the past financial data along with the earnings call node embeddings O_e, obtained using GCN. The model computes a series of hidden states $(h_1, .., h_T)$ and a sequence of outputs $(y_1, ..y_T)$, by repeating the following recurrence relation from time $t = 1$ to T:

$$h_t = f(W_{hx}x_t + W_{hh}h_{t-1} + O_e + b_h) \quad (16)$$

$$y_t = softmax(W_{oh}h_t + b_o) \quad (17)$$

Here, $W_{hx}, W_{hh}, W_{oh}, x_i, b_h, b_o$ are learnable parameters and x_t is the average t-day past volatility. Following Karpathy and Fei-Fei (2015), we feed the GCN embeddings to the LSTM only at the first iteration. We use the output y_T from the last LSTM unit for the final multi-output prediction.

Network Optimization We finally train VolTAGE by optimizing a multi-task loss as:

$$\mathcal{L} = \frac{1}{2n}\left(\mu \sum_i (\hat{y_i} - y_i)^2 + (1-\mu)\sum_j (\hat{y_j} - y_j)^2)\right) \quad (18)$$

Here, $\hat{y_i}$, $\hat{y_j}$ are predicted volatilities and y_i , y_j are true volatilities for the main and auxiliary tasks, respectively. μ is a parameter that controls the relative weight of the loss between the two tasks.

5 Experimental Setup

5.1 Data

We used the S&P 500 2017 Earnings Conference Calls dataset (Qin and Yang, 2019).[5] The dataset consists of 559 earnings call audio recordings and their transcripts for 277 public companies in the S&P 500 index. Each call is segmented into a sequence of audio clips aligned with their corresponding text sentences, as spoken by the Chief Executive Officer (CEO) during the call. We temporally divide the data into train, validation, and test sets in a ratio of 70 : 10 : 20 respectively to ensure future data is not used for forecasting. We extract stock prices for each company using Yahoo Finance[6] from 1 January'17 till 31 December'17. The stock data for 11 earnings calls was not available on Yahoo Finance; hence we excluded these calls from our dataset. Following Qin and Yang (2019); Yang et al. (2020), we experiment with $n \in \{3, 7, 15, 30\}$ days to analyze the performance over both short and long term periods.

5.2 Baselines

We compare VolTAGE with the following methods:

- $\mathbf{V_{past}}$: Following Qin and Yang (2019), we use V_{past}, the average log volatility of the past d days to predict the future d days' average log volatility.
- **bc-LSTM:** We also compare against bc-LSTM (Poria et al., 2017) which extracts the uni-modal features using separate contextual Bi-LSTMs and fuses them.
- **MDRM:** Qin and Yang (2019) extract pre-trained GloVe embeddings and hand-crafted acoustic features that are fed to separate BiLSTMs to get their uni-modal contextual embeddings, which are then fused and fed to a two-layer dense network.
- **HTML:** Yang et al. (2020) is the state-of-the-art model using WWM-BERT to encode text tokens. HTML makes use of the same audio features as MDRM. The unimodal features are fused and fed to a sentence-level transformer to get the multimodal representations for each call.

5.3 Training Setup

We tune the hyperparameters on the validation mean square error (MSE) to get: dropout $\delta \in [0, 0.8]$, learning rate $\lambda \in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$, batch size $b \in \{8, 16, 32, 64\}$ and epochs (< 100).

We use FinBERT with default pre-training parameters, which outputs a 768-dimensional embedding for each sentence. The maximum number of audio clips in any call is 520. Hence, we zero-pad the calls that have less than 520 clips for efficient batching. The number of neurons in the time distributed dense layer following the audio and text BiLSTMs is 100. The heterogeneous graph

[5] We were unable to map price data for 11 data points, which were subsequently dropped

[6] https://finance.yahoo.com/

Model	$\overline{\text{MSE}}$	MSE_3	MSE_7	MSE_{15}	MSE_{30}	R^2_3	R^2_7	R^2_{15}	R^2_{30}
Vpast	1.12	2.99	0.83	0.42	0.23				
LSTM	0.75	1.97	0.46	0.32	0.24	0.34	0.44	0.24	-0.02
HAN (Glove)	0.60	1.43	0.46	0.31	0.20	0.52	0.44	0.27	0.14
MDRM (Audio)	0.60	1.41	0.44	0.32	0.22	0.53	0.47	0.25	0.03
MDRM (Text+Audio)	0.58	1.37	0.42	0.30	0.22	0.54	0.49	0.29	0.06
HTML (Text)	0.46	1.18	0.37	**0.15**	**0.13**	0.61	0.55	**0.64**	**0.42**
HTML (Text+Audio)	0.40	0.85	0.35	0.25	0.16	0.72	0.58	0.40	0.32
VolTAGE	**0.31**	**0.63**	**0.29**	0.17	0.14	**0.79**	**0.65**	0.60	0.39

Table 1: n-day volatility MSE and coefficient of determination R^2 for all models. **Bold** represents the best results.

contains 559 nodes for earning calls connected to the 277 interrelated company nodes. The GCN is trained using Pytorch Geometric (Fey and Lenssen, 2019).[7] We use two GCN layers having 200 and 100 units respectively, inter-spaced by ReLU and followed by a single dense layer. The 200 dimensional feature vectors from the first layer of the GCN after the ReLU activation are fed to a 200-unit conditioned LSTM model for multi-task volatility prediction. We optimize VolTAGE using the Adam (Kingma and Ba, 2014) optimizer.

6 Results and Analysis

6.1 Comparative Analysis

We present the volatility prediction performance of VolTAGE and the baselines in Table 1. We report the MSE averaged across 10 different runs for all models for the main task (n-day average prediction). Our choice of using MSE as a comparative metric is motivated by prior work (Qin and Yang, 2019; Yang et al., 2020). Additionally, we also report the coefficient of determination $R^2 = 1 - \frac{MSE}{MSE_{V_{past}}}$, to illustrate the improvements with V_{past}. We observe gains across the multimodal HTML that leverages both text and audio modalities. We ascribe this improvement to the cross-modal attention fusion mechanism, which uses associations between audio and text modalities over each contextual utterance instead of concatenation used in HTML. Moreover, a key limitation of the baselines is the assumption of independence of inter-stock movements. VolTAGE captures the correlations between price movements of related stocks through the GCN, and hence, volatility, amplifying performance. Similar to prior work (Qin and Yang, 2019), Table 1 illustrates that forecasting volatility in the short-term is a more intricate task than long-term. Based on Post Earnings Announcement Drift (PEAD) (Bernard and Thomas, 1989), a documented financial phenomenon, we note that the price fluctuations around earning calls tend to stabilize over long periods. We observe that VolTAGE outperforms the baselines by a large margin in short-term prediction ($n = 3, 7$); however the margin diminishes over longer durations ($n = 30$).

[7] We extract features for nodes using the last layer of verbal-vocal fusion tuned only for average n-day volatility prediction. The verbal-vocal attention fusion was not trained on multi-task loss, and VolTAGE is not trained end-to-end.

Model	$\overline{\text{MSE}}$	MSE_3	MSE_7	MSE_{15}	MSE_{30}
Glove	0.68	0.99	0.67	0.55	0.49
BERT	0.52	0.85	0.50	0.37	0.35
FinBERT	0.49	0.81	0.50	0.35	0.31
Audio	0.53	0.85	0.52	0.41	0.33
Audio+FinBERT (CM Attn)	0.45	0.77	0.47	0.31	0.24
Audio+FinBERT (CM Attn)+GCN	0.37	0.66	0.39	0.23	0.22
VolTAGE	0.31	0.63	0.29	0.17	0.14

Table 2: Ablation Results over model components

6.2 Ablation Study

We observe an improvement across the text modality (T), when compared to the HTML (Text) model (Yang et al., 2020) in Table 1 and Table 2. This performance gain can be attributed to FinBERT, which is trained to handle language tasks in the financial domain, while the sentence-level transformer employed in the HTML (Text) model is a generalized implementation of BERT (Devlin et al., 2019). We also note that representations learned by FinBERT outperform both GloVe (Pennington et al., 2014) and BERT embeddings, reiterating the effectiveness of domain-specific pre-training. Further, we observe from Table 2, that the Audio+FinBERT (CM Attn) model outperform unimodal components, demonstrating the utility of multimodal verbal-vocal cues for volatility prediction. On adding the GCN, we observe a gain of 17.7%, likely due to the GCN's ability to learn correlations between price movements of related stocks that are captured by the company relations. Finally, on introducing the financial modality via

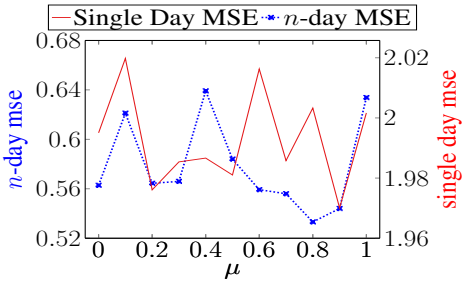
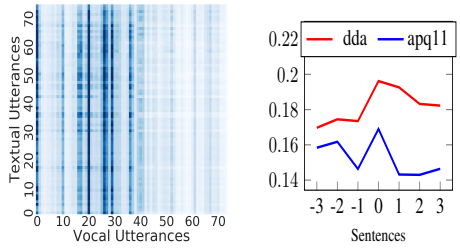


Figure 3: 3-Day Validation MSE vs μ



(a) Audio Aware Text Attention (b) DG: Shimmer Analysis

Figure 4: Verbal and Vocal features from the earnings call for Dollar General from Q3 2017.

the conditioned LSTM network helps in counteracting the impact of PEAD by introducing earnings call independent information into the model; this can be observed in Table 2. We note that VolTAGE outperforms all its ablative components, demonstrating how each of its multimodal components complement each other.

6.3 On Multi-task Learning

Training a network for multiple tasks jointly has shown to improve performance on tasks that share a conceptual similarity (Caruana, 1997). In our case, we optimize VolTAGE on both n-day average and single-day volatility prediction tasks a multi-task formulation. In Figure 3, we analyze the variation of the weight parameter μ with the 3-day validation MSE of n-day average, and single-day predicted volatility. As both tasks share a weighted loss function, by tuning μ, we trade-off between the two tasks. We observe from Figure 3, that at the extreme values of the weight parameter $\mu = 0$ and $\mu = 1$, that represent single task learning on the single-day and n-day average prediction tasks respectively, VolTAGE does not obtain optimal performance. Empirically, we find the optimal $\mu = 0.8$ for 3-day volatility forecasting on the main task, thus validating our hypothesis that multi-task learning across both average and single day spans of volatility prediction improve predictive power.

7 Qualitative Analysis

We analyze the Q3-2017 earnings call for DG (Dollar General), an American variety store company. The stock's price became highly volatile for a few days following the earnings call. Figure 4a shows the audio-aware text attention heatmap for the duration of the earning call. The heatmap represents cross-modal attention weights assigned to textual utterances using corresponding vocal cues. Here each cell (i, j) represents the weight of j^{th} vocal utterance on the i^{th} textual utterance. It is observed that the highest attention is towards the middle of the call, suggesting that the verbal cues of this portion have the highest impact on the text contextual embeddings for most of the sentences in the call. Earning calls of companies are often structured such that the beginning of the call involves introductory disclaimer and greetings, while the CEO starts presenting financial results for the reporting quarter along with future goals of the company towards the middle of the call, which indicates why we see such influential utterances in this portion.

Figure 4b shows the disparity between CEO's vocal and verbal cues around the utterance. While textual content seems positive, a sudden spike in shimmer features in CEO's voice while speaking this sentence suggests disharmony between verbal and vocal cues. Past research in acoustics (Li et al., 2007) suggests an elevated shimmer could be indicative of underlying stress in speech. After the earning call, it was noted that the gross margin of the company slipped by 0.4%, due to the increased transportation costs due to hurricane Irma in 2017. On analyzing the graph, we observe that DG has edge connections with WMT (Wallmart) and TGT (Target Corp.), both of which are retail variety stores, like DG. Analysts had estimated a negative impact of about $2.8 Billion on the retail sector due to the hurricane Irma. This examination is also reflected in the high volatilities recorded for WMT and TGT during the same quarter. A unimodal model may miss these subtle disparities between text and audio. Therefore, VolTAGE, by leveraging cross-modal attention fusion and correlation graphs, accurately forecasts the volatility of DG, three days post the earnings call.

8 Conclusion and Future Work

Volatility, measured as a deviation in returns, is a reliable indicator of market risk linked with a stock. A rich source of company information is earnings calls that provide high risk-reward opportunities given their uniqueness and critical information disclosure. Although evidence shows that enriching models with speech and inter-stock correlations can improve volatility forecasting, this area is underexplored. We propose VolTAGE, a neural architecture that jointly exploits coherence over speech, text, and inter-stock correlations for volatility forecasting following earnings calls. Through experiments on S&P 500 index data, we show the merit of cross-modal gated attention fusion, graph-based learning, and multi-task prediction for volatility forecasting.

There are several promising directions of future work that we wish to explore. First, we want to improve upon the audio feature extraction. To model the speech of CEOs in earnings calls, using semitones rather than raw frequency for pitch-related features. Experimenting with other sets of commonly used acoustic features such as MFCC coefficients, OpenSMILE features and auDeep features for representing audio utterances also form a future direction for audio feature extraction. Second, we want to expand the analysis presented in this paper beyond the S&P 500 index and US-based companies. Existing research (Qin and Yang, 2019; Yang et al., 2020) and this work at the intersection of natural language processing and earnings calls are limited to a small set of companies and earnings calls. Analyzing the demographic, cultural, and gender bias in research pertaining to financial disclosures, particularly earnings calls, forms a future direction of research. We would also want to work on studying a wider set of earnings calls and companies spanning multiple languages, demographics, speakers and gender.

References

Md Shad Akhtar, Dushyant Chauhan, Deepanway Ghosal, Soujanya Poria, Asif Ekbal, and Pushpak Bhattacharyya. 2019. Multi-task learning for multi-modal emotion recognition and sentiment analysis. 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 370–379.

Leif BG Andersen. 2007. Efficient simulation of the heston stochastic volatility model. *Available at SSRN 946405*.

Dogu Araci. 2019. Finbert: Financial sentiment analysis with pre-trained language models.

Yigit Atilgan. 2014. Volatility spreads and earnings announcement returns. *Journal of Banking & Finance*, 38:205–215.

Victor L. Bernard and Jacob K. Thomas. 1989. Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research*, 27:1–36.

Paul Boersma and Vincent Van Heuven. 2001. Speak and unspeak with praat. *Glot Int*, 5:341–347.

Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. *Journal of computational science*, 2(1):1–8.

Judee Burgoon, W. Mayew, Justin Giboney, Aaron Elkins, Kevin Moffitt, Bradley Dorn, Michael Byrd, and Lee Spitzley. 2015. Which spoken language markers identify deception in high-stakes settings? evidence from earnings conference calls. *Journal of Language and Social Psychology*, 35.

John Y Campbell, John J Champbell, John W Campbell, Andrew W Lo, Andrew W Lo, and A Craig MacKinlay. 1997. *The econometrics of financial markets*. princeton University press.

Rich Caruana. 1997. Multitask learning. *Machine learning*, 28(1):41–75.

Belinda Crawford Camiciottoli. 2011. Ethics and ethos in financial reporting: Analyzing persuasive language in earnings calls. *Business Communication Quarterly*, 74(3):298–312.

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.

Bhuwan Dhingra, Hanxiao Liu, Zhilin Yang, William W Cohen, and Ruslan Salakhutdinov. 2016. Gated-attention readers for text comprehension. *arXiv preprint arXiv:1606.01549*.

Ilia D Dichev and Vicki Wei Tang. 2009. Earnings volatility and earnings predictability. *Journal of accounting and Economics*, 47(1-2):160–181.

Francis X Diebold and Kamil Yılmaz. 2014. On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1):119–134.

Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2015. Deep learning for event-driven stock prediction. In *Twenty-fourth international joint conference on artificial intelligence*.

Jin-Chuan Duan. 1995. The garch option pricing model. *Mathematical finance*, 5(1):13–32.

Fuli Feng, Xiangnan He, Xiang Wang, Cheng Luo, Yiqun Liu, and Tat-Seng Chua. 2019. Temporal relational ranking for stock prediction. *ACM Transactions on Information Systems*, 37(2):1–30.

Matthias Fey and Jan E. Lenssen. 2019. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*.

Stephen Figlewski. 1994. Forecasting volatility using historical data.

Xue Guo, Hu Zhang, and Tianhai Tian. 2018. Development of stock correlation networks using mutual information and financial big data. *PloS one*, 13:e0195941.

Ehsan Hoseinzade, Saman Haratizadeh, and Arash Khoeini. 2019. U-cnnpred: A universal cnn-based predictor for stock markets.

Ziniu Hu, Weiqing Liu, Jiang Bian, Xuanzhe Liu, and Tie-Yan Liu. 2018. Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. In *Proceedings of the eleventh ACM international conference on web search and data mining*.

Weiwei Jiang. 2020. Applications of deep learning in stock market prediction: recent progress. *arXiv preprint arXiv:2003.01859*.

Xiaoming Jiang and Marc D Pell. 2017. The sound of confidence and doubt. *Speech Communication*, 88:106–126.

C Kenneth Jones. 2017. Modern portfolio theory, digital portfolio theory and intertemporal portfolio choice. *American Journal of Industrial and Business Management*, 7:833–854.

Andrej Karpathy and Li Fei-Fei. 2015. Deep visual-semantic alignments for generating image descriptions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3128–3137.

Katherine Keith and Amanda Stent. 2019. Modeling financial analysts' decision making via the pragmatics and semantics of earnings calls. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 493–503, Florence, Italy. Association for Computational Linguistics.

Raehyun Kim, Chan Ho So, Minbyul Jeong, Sanghoon Lee, Jinkyu Kim, and Jaewoo Kang. 2019. Hats: A hierarchical graph attention network for stock movement prediction. *arXiv preprint arXiv:1908.07999*.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization.

Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*.

Shimon Kogan, Dimitry Levin, Bryan R. Routledge, Jacob S. Sagi, and Noah A. Smith. 2009. Predicting risk from financial reports with regression. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 272–280, Boulder, Colorado. Association for Computational Linguistics.

Werner Kristjanpoller, Anton Fadic, and Marcel C Minutolo. 2014. Volatility forecast using hybrid neural network models. *Expert Systems with Applications*, 41(5):2437–2442.

X. Li, J. Tao, M. T. Johnson, J. Soltis, A. Savage, K. M. Leong, and J. D. Newman. 2007. Stress and emotion classification using jitter and shimmer features. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, volume 4, pages IV–1081–IV–1084.

Jiexi Liu and Songcan Chen. 2019. *Non-stationary Multivariate Time Series Prediction with Selective Recurrent Neural Networks*, pages 636–649.

Burton G Malkiel. 2003. The efficient market hypothesis and its critics. *Journal of economic perspectives*, 17(1):59–82.

Pekka Malo, Ankur Sinha, Pyry Takala, Pekka Korhonen, and Jyrki Wallenius. 2013. Financialphrasebank-v1.0.

Marc-Andre Mittermayer and Gerhard F Knolmayer. 2006. Newscats: A news categorization and trading system. In *Sixth International Conference on Data Mining (ICDM'06)*, pages 1002–1007. Ieee.

Saloni Mohan, Sahitya Mullapudi, Sudheer Sammeta, Parag Vijayvergia, and David C Anastasiu. 2019. Stock price prediction using news sentiment analysis. In *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*, pages 205–208. IEEE.

Puneet Mongia and R.K. Sharma. 2014. Estimation and statistical analysis of human voice parameters to investigate the influence of psychological stress and to determine the vocal tract transfer function of an individual. *Journal of Computer Networks and Communications*, 2014.

Claude Montacié and Marie-José Caraty. 2018. Vocalic, lexical and prosodic cues for the interspeech 2018 self-assessed affect challenge. In *Interspeech*, pages 541–545.

Amy Morris-Drake, Julie M Kern, and Andrew N Radford. 2016. Cross-modal impacts of anthropogenic noise on information use. *Current Biology*, 26(20):R911–R912.

Mahla Nikou, Gholamreza Mansourfar, and Jamshid Bagherzadeh. 2019. Stock price prediction using deep learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems in Accounting, Finance and Management*, 26.

Yaohao Peng, Pedro Henrique Melo Albuquerque, Jader Martins Camboim de Sá, Ana Julia Akaishi Padula, and Mariana Rosa Montenegro. 2018. The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with support vector regression. *Expert Systems with Applications*, 97:177–192.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Ser-Huang Poon and Clive WJ Granger. 2003. Forecasting volatility in financial markets: A review. *Journal of economic literature*, 41(2):478–539.

Soujanya Poria, Erik Cambria, Devamanyu Hazarika, Navonil Majumder, Amir Zadeh, and Louis-Philippe Morency. 2017. Context-dependent sentiment analysis in user-generated videos. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 873–883, Vancouver, Canada. Association for Computational Linguistics.

Jiří Přibil and Anna Přibilová. 2009. Spectral flatness analysis for emotional speech synthesis and transformation. *Lecture Notes in Computer Science*, 5641:106–115.

Yu Qin and Yi Yang. 2019. What you say and how you say it matters: Predicting stock volatility using verbal and vocal cues. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 390–401, Florence, Italy. Association for Computational Linguistics.

Jonathan L Rogers, Douglas J Skinner, and Andrew Van Buskirk. 2009. Earnings guidance and market uncertainty. *Journal of Accounting and Economics*, 48(1):90–109.

Ramit Sawhney, Shivam Agarwal, Arnav Wadhwa, and Rajiv Ratn Shah. 2020. Spatiotemporal hypergraph convolution network for stock forecasting. In *2020 IEEE International Conference on Data Mining (ICDM)*.

Marc Schröder, Roddy Cowie, Ellen Douglas-Cowie, Machiel Westerdijk, and Stan Gielen. 2001. Acoustic correlates of emotion dimensions in view of speech synthesis. volume 1, pages 87–90.

Jianfeng Si, Arjun Mukherjee, Bing Liu, Sinno Jialin Pan, Qing Li, and Huayi Li. 2014. Exploiting social relations and sentiment for stock prediction. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1139–1145.

Jinghua Tan, Jun Wang, Denisa Rinprasertmeechai, Rong Xing, and Qing Li. 2019. A tensor-based elstm model to predict stock price using financial news. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

Sarah C Tasker. 1998. Bridging the information gap: Quarterly conference calls as a medium for voluntary disclosure. *Review of Accounting Studies*, 3(1-2):137–167.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph attention networks. In *International Conference on Learning Representations*.

Murlikrishna Viswanathan, Zhen-Xing Zhang, Xue-Wei Tian, and Joon S. Lim. 2012. Emotional-speech recognition using the neuro-fuzzy network. In *Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication*, ICUIMC '12, New York, NY, USA. Association for Computing Machinery.

Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: A free collaborative knowledgebase. *Communications of the ACM*, 57:78–85.

William Yang Wang and Zhenhao Hua. 2014. A semiparametric Gaussian copula regression model for predicting financial risks from earnings calls. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1155–1165, Baltimore, Maryland. Association for Computational Linguistics.

Linyi Yang, Tin Lok James Ng, Barry Smyth, and Riuhai Dong. 2020. Html: Hierarchical transformer-based multi-task learning for volatility prediction. In *Proceedings of The Web Conference 2020*, pages 441–451.

Jie Zheng, Andi Xia, Lin Shao, Tao Wan, and Zengchang Qin. 2019. Stock volatility prediction based on self-attention networks with social information. In *2019 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, pages 1–7. IEEE.

A Appendices

A.1 Pitch Analysis

We extract the following audio features corresponding to pitch:

1. **Minimum pitch:** The minimum pitch frequency of the frames within a specified time duration.

2. **Maximum pitch:** The maximum pitch frequency of the frames within a specified time duration.

3. **meanF0:** The mean of the fundamental frequency (f0) within a specified time duration.

4. **stdevF0:** The standard deviation of the fundamental frequency (f0) within a specified time duration.

5. **Number of pulses:** The number of pulses within a given time window.

6. **Number of periods**: The number of periods/cycles within a given time window.

7. **Degree of voice breaks:** This is the total duration of the breaks between the voiced parts of the signal, divided by the total duration of the analysed part of the signal.

8. **voiced_frames:** The number of voiced frames. A frame is regarded as locally unvoiced if it has a voicing strength below the voicing threshold (whose standard value is 0.45), or a local peak below the silence threshold (whose standard value is 0.03).

9. **Voiced to total ratio:** The number of voiced frames in a window divided by the total number of frames.

10. **Voiced_to_unvoiced_ratio:** The number of voiced frames in a window divided by the number of unvoiced frames (unvoiced frames are given by total frames-voiced frames).

A.2 Voice Analysis

Under voice analysis, we extract different features quantifying *jitter* and *shimmer* in the earnings call audio. *Jitter* is the relative average vocal perturbation while *Shimmer* is the moment-to-moment amplitude variation. We now describe the various features extracted in this category:

1. **Jitter (local):**This is the fraction of average absolute difference between consecutive periods by the average period.

2. **Jitter (local, absolute):** This is the average absolute difference between consecutive periods in seconds.

3. **Jitter (RAP):** This is the Relative Average Perturbation, the average absolute difference between a period and the average of it and its two neighbours, divided by the average period.

4. **Jitter (ppq5):** This is the five-point Period Perturbation Quotient, the average absolute difference between a period and the average of it and its four closest neighbours, divided by the average period.

5. **Jitter (ddp):** This is the average absolute difference in jitter between consecutive periods divided by the average period. This is Praat's original Get jitter and is proportional to three times RAP.

6. **Shimmer (local) :** This is the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude.

7. **Shimmer (local, dB):** This is the average absolute base-10 logarithm of the difference between the amplitudes of consecutive periods, multiplied by 20.

8. **Shimmer (apq3):** This is the three-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of its neighbours, divided by the average amplitude.

9. **Shimmer (apq5):** This is the five-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its four closest neighbours, divided by the average amplitude.

10. **Shimmer (apq11):** This is the 11-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its ten closest neighbours, divided by the average amplitude.

11. **Shimmer (ddp):** This is the average absolute difference of shimmer between the amplitudes of consecutive periods. This is Praat's original

Get shimmer and its value is proportional to three times APQ3.

A.3 Intensity Analysis

We extract the following intensity features:

1. **Mean Intensity:** The mean (in dB) of the intensity values of the frames within a specified time duration.

2. **Minimum intensity:** The minimum (in dB) of the intensity values of the frames within a specified time duration.

3. **Maximum intensity:** The maximum (in dB) of the intensity values of the frames within a specified time duration.

4. **SD_energy:** Standard deviation of energy in the frames within a specified time duration.

A.4 Harmonicity Analysis

We extract the *Harmonics-to-Noise Ratio (HNR)* of the earnings calls audio which has shown to be a measure of the ”hoarseness of a speaker”.

1. **Harmonics-to-Noise Ratio (HNR):** It represents the degree of acoustic periodicity. It is expressed in decibels. It can be used as a measure for the signal-to-noise ratio of periodic voice signals.